12th Canadian Symposium on Remote Sensing
12e Symposium Canadien sur Teledetection

1989 INTERNATIONAL GEOSCIENCE AND REMOTE SENSING SYMPOSIUM (IGARSS '89)

and URSI Commission F Meeting

IGARSS'89 12TH CANADIAN SYMPOSIUM ON REMOTE SENSING 12e SYMPOSIUM CANADIEN SUR LA TELEDETECTION

Quantitative Remote Sensing:
An Economic Tool for the Nineties

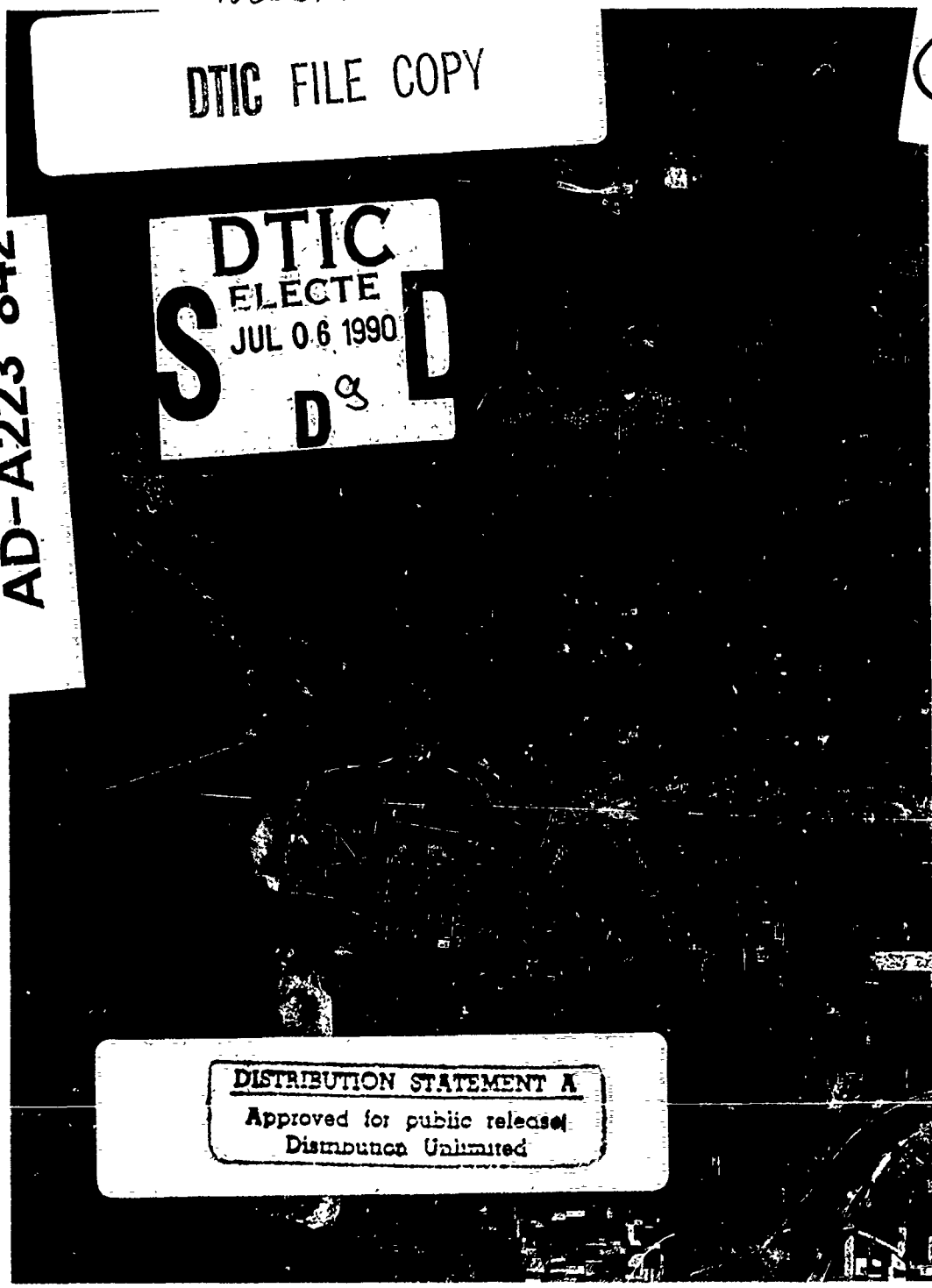
N00014-89-T-1012

DTIC FILE COPY

1

DTIC
ELECTE
JUL 06 1990
S D

AD-A223 842



DISTRIBUTION STATEMENT A
Approved for public release
Distribution Unlimited

IEEE

U.R.S.I.

VANCOUVER, CANADA JULY 10-14, 1989

Volume 1. Monday July 10

90 07 3 230

Printed in Canada

LIBRARY OF CONGRESS #89-84217
IEEE #89CH2768-0

VOLUME 1

IEEE No. 89CH2768-0
LIBRARY OF CONGRESS No. 89-84217

IGARSS '89

12th CANADIAN SYMPOSIUM ON REMOTE SENSING
12e SYMPOSIUM CANADIEN SUR LA TELEDETECTION

REMOTE SENSING:

AN ECONOMIC TOOL FOR THE NINETIES

JULY 10 - 14 th 1989 VANCOUVER, CANADA



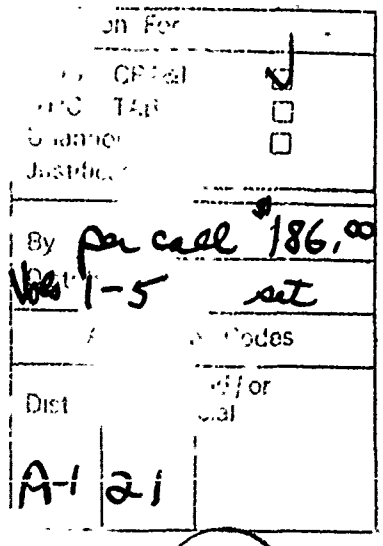






SPONSORS:
IEEE, CRSS, URSI

AVAILABLE FOR $186 FOR SET per Institute of Electrical and Electronics Engineers, Inc. 345 East 47th St., New York, NY 10017
TELECON 7/5/90 VG





IGARSS '89
REMOTE SENSING: AN ECONOMIC TOOL FOR THE NINETIES
VANCOUVER, CANADA
JULY 10-14, 1989

COPYRIGHT - 1989 by IEEE

Proceedings published by:
IGARSS '89 12th Canadian Symposium on Remote Sensing

Printed in Canada

IGARSS '89 - 12th CANADIAN SYMPOSIUM ON REMOTE SENSING
IGARSS '89 - 12th SYMPOSIUM CANADIEN SUR LA TELEDETECTION

IGARSS '89 - 12th CANADIAN SYMPOSIUM ON REMOTE SENSING SPONSORS
SPONSORS COMMANDITAIRES

IGARSS ' 89 - 12th Canadian Symposium on Remote Sensing is sponsored by the Canadian Remote Sensing Society, the IEEE Geoscience and Remote Sensing Society, the International Union of Radio Science, and the Canada Centre for Remote Sensing.

It is co-sponsored by:

B.C. Ministry of Environment and Parks
B.C. Ministry of Forests, Inventory Branch
B.C. Research Corporation
Binary Image Corporation
Canadian Forest Service
Cominco Ltd.
Fisheries and Oceans Canada
MacDonald, Dettwiler and Associates Ltd.
National Research Council Canada
Pamap Graphics Ltd.
University of British Columbia

IGARSS '89 - 12th CANADIAN SYMPOSIUM ON REMOTE SENSING ORGANIZING COMMITTEE

Dr. John S. MacDonald, General Chairman

Dr. Ed Jull, Vice Chairman, URSI
Dr. Peter Murtha, Vice Chairman, CRSS
Dr. Keith Raney, Vice Chairman, IEEE/GRS-S

Dr. Jim Gower, Technical Program Chairman

Dr. Josef Cihlar, Technical Co-Chairman, CRSS
Dr. David Goodenough, Technical Co-Chairman, IEEE/GRS-S

M. Jean-Marie Dubois, French Transcription
Mr. Frank Hegyi, Social Program and Tours
Ms. Nedenia Holm, Secretary
Dr. Y. Jim Lee, Exhibits
Mr. Peter Louie, Finance
Dr. Pam Sallaway
Dr. Hans Schreier, UBC Facilities/Registration
Dr. Mark Sondheim, Publicity
Dr. Mit Tilkov
Dr. Geoff Tomlins, Publication

CRSS Technical Committee

Chair: J. Cihlar, Canada Centre for Remote Sensing, Canada
F.J. Ahern, Canada Centre for Remote Sensing, Canada
Kohei Arai, National Space Development Agency of Japan, Japan
F. Bonn, Universitie de Sherbrooke, Canada
G.A. Borstad, G.A. Borstad Associates Ltd., Canada
J. Brown, Canada Centre for Remote Sensing, Canada
R. Currie, Pacific Geoscience Centre, Canada
G. Duchossois, European Space Agency, Canada
A. Fabbri, Canada Centre for Remote Sensing, Canada
B.E. Goodison, Atmospheric Environment Service, Canada
B. Guidon, Canada Centre for Remote Sensing, Canada
F.G. Hall, NASA Goddard Space Flight Centre, USA
D. Halpern, Jet Propulsion Laboratories, USA
F. Hegyi, Ministry of Forests, Canada
J. Iisaka, Canada Centre for Remote Sensing, Canada
J. Kosalos, Honeywell Marine Systems, USA
L. McNutt, Canada Centre for Remote Sensing, Canada
J.B. Mercer, Intera Technologies Ltd., Canada
R.E. Murphy, NASA/Headquarters, USA
R.A. O'Neil, Canada Centre for Remote Sensing, Canada
P. Sallaway, PAMAP Graphics Ltd., Canada
L. Sayn-Wittgenstein, Canada Centre for Remote Sensing, Canada
R. Slaney, Radarsat Project Office, Canada
W.M. Strome, PCI Inc., Canada
T. Toutin, Canada Centre for Remote Sensing, Canada
B. Woodham, University of British Columbia, Canada
V. Zsilinszky, Ontario Centre for Remote Sensing, Canada

IEEE/GRSS Technical Committee

Chair: David G. Goodenough, Canada Centre for Remote Sensing, Canada
R.C. Beal, The John Hopkins University, USA
G. Brachet, SPOT Image, France
F. Carsey, Jet Propulsion Laboratories, USA
K.R. Carver, University of Massachusetts, USA
R. Cheney, National Ocean Service NOAA, USA
J. Cimino, Jet Propulsion Laboratories, USA
M. Ehlers, University of Maine, USA
A.F.H. Goetz, University of Colorado, USA
K.I. Itten, University of Zurich-Irchel, Switzerland
E. Ledrew, University of Waterloow, Canada
C.A. Luther, Office of Naval Research, USA
J.P. Muller, University College London, UK
J.A. MacDonald, MacDonald Dettwiler Associates, Canada
J. Megier, Ispra Establishment, Italy
R.K. Raney, Radarsat Project Office, Canada
P.N. Slater, University of Arizona, USA
C.T. Swift, University of Massachusetts, USA
Mikio Takagi, University of Tokyo, Japan
G.F. West, University of Toronto, Canada

URSI Technical Committee

Chair: J.F.R. Gower, Institute of Ocean Sciences, Canada
J.R. Apel, The John Hopkins University, USA
R.K. Crane, Darmouth College, USA
J.L. Green, NOAA Aeronomy Laboratory, USA
A. Guissard, Universitie Catholilque de Louvain, Belgium
D.C. Hogg, University of Colorado, USA
R. Humphreys, MacDonald Dettwiler Associates, Canada
E. Jull, University of British Columbia, Canada
R.K. Moore, University of Kansas Centre for Research Inc., USA
D.W. Oldenburg, University of British Columbia, Canada
R. Olsen, Communications Research Centre, Canada
B. Segal, Communications Research Centre, Canada
O.H. Shemdin, Ocean Research and Engineering, USA
J.R. Wait, University of Arizona, USA
J. Walsh, Memorial University of Newfoundland, Canada

Book 1
Table of Contents

Partial contents:

	Page
Determining intrinsic surface reflectance in rugged terrain and changing illumination R.J. Woodham	1
Identification to go beyond discrimination/The status of mineral and lithological identification with high resolution spectrometer data; Examples and challenges Jonathan F. Huntington	6
Towards quantitative observation from Synthetic Aperture Radar R.K. Raney, Radarsat Project Off., Ottawa, Ont., Canada	12
Simple models for complex natural surfaces: a strategy for the hyperspectra era of remote sensing John B. Adams	16
Investigation of the relationship of SAR HH and VV backscatter to surface roughness and dielectric constant S.D. Wall, J.J. van Zyl, NASA/JPL, Pasadena, CA, USA	22
Radar scattering classification maps from multifrequency imaging radar polarimetric data H.A. Zebker, J.J. van Zyl, T.G. Farr, NASA/JPL, Pasadena, CA, USA	24
Applications of a SAR image simulator to the study of the polarization signature of man-made targets J.M. Nasr, Aerospatiale DSS, STS/M, Les Mureaux Cedex, France	25
Classification of earth terrain cover using multifrequency imaging radar polarization data J.J. van Zyl, C.F. Burnette, NASA/JPL, Pasadena, CA, USA	29
Segmentation of multifrequency polarimetric radar images to facilitate the inference of geophysical parameters C.F. Burnette, P.C. Dubois, J.J. van Zyl, NASA/JPL, Pasadena, CA, USA	30
Comparison of sea ice type classification using polarimetric and non-polarimetric synthetic aperture radar D.P. Winebrenner, D.A. Rothrock, H. Stern, Univ. of Washington, Seattle, WA, USA	34
Detection of dielectric cylinders on a rough surface using a multi-frequency polarimetric radar D. Kahny, S. Riegger, W. Wiesbeck, Univ. Karlsruhe, Karlsruhe, FR Germany	37
Polarimetric matched filter for coherent imaging A.B. Kostinski, B.D. James, W.-M. Boerner, Univ. of Illinois, Chicago, IL, USA	41
Data management in the FIFE information system D.E. Strebel, VERSAR, Inc., Columbia, MD, USA, J.A. Newcomer, ST Systems Corp., Lanham, MD, USA, J.P. Ormsby, F.G. Hall, NASA/GSFC, Greenbelt, MD, USA, P.J. Sellers, Univ. of MD, College Park, MD, USA	42

Page

Incorporation of polarimetric radar images into multisensor data sets 46
D.L. Evans, J.J. van Zyl, M. Kobrick, C.F. Burnette, S.D. Wall, NASA/JPL, Pasadena, CA, USA

Fuzzy logic and knowledge-based GIS, a prospectus 47
Leung Yee, Chinese Univ. of Hong Kong, Shatin, Hong Kong

Consideration of geographical information systems as tools for classification 51
K.O. Niemann, Univ. of Victoria, Victoria, BC Canada, L. Nikiforuk, Alberta Res. Council, Edmonton, Alta Canada, A. Brierley, Agriculture Canada, Edmonton, AB, Canada

Segmentaton d'image SPOT integree a l'information cartoghapique en vue de l'etablissement de la carte d'utilisation de sol au Maroc 56
M. Ait Belaid, K.P.B. Thomson, G. Edwards, J.M. Bealieu, Univ. Laval, Ste-Foy, Que., Canada

Integration of digital elevation model parameters with Landsat and SPOT imagery for mapping mountain glaciers: Building a glaciological information system 60
D.J. Gratton, P.J. Howarth, Univ. of Waterloo, Waterloo, Ont., Canada

Remote sensing and Geographic Information Systems: Towards integrated spatial information processing 63
M. Ehlers, Univ. of Maine, Orono, ME, USA

Interactive query of Canada - wide mineral deposits data, using the SPANS Geographic Information System 67
T. Webster, TYDAC Tech. Inc., Ottawa, Ont., Canada, G.F. Bonham-Carter, D.F. Garson, R.M. Laramee, Geological Survey of Canada, Ottawa, Canada

Synergism of remotely sensed and contextual data to monitor changes in land use 69
H. Xu, J.A.T. Young, Univ. of Edinburgh, Edinburgh, UK

Gestion des ressources agricoles a l'aide d'un systeme d'information geographique: application a l'evaluation quantitative de l'erosion hydrique 73
D. Baril, S. Perras, Univ. de Sherbrooke, Sherbrooke, Que., Canada, A. Pesant, Stat. de rech. de Lennoxville, Qc, Canada, F. Bonn, Univ. de Sherbrooke, Sherbrooke, Que., Canada

Integrated image-map products for township level land information 77
M.J. Manore, CCRS, Ottawa, Ont., Canada, R. Protz, B.A. Proud, Univ. of Guelph, Guelph, Ont., Canada

The use of microcomputer geographical information system for land evaluation 81
G. Li, Inst. of Geography, Chinese Academy of Sciences, Beijing China

Automatic detection of lineaments from Landsat data 85
A.K. Joshi, Syracuse Univ., Syracuse, NY, USA

Three-dimensional computer analysis and modelling of remote sensing-structural geologic problems 89
R.L. Thiessen, J.R. Eliason, E.R. Rieken, Washington State Univ., Pullman, WA, USA

Edge following as graph searching and Hough transform algorithms for lineament detection 93
J. Wang, P.J. Howarth, Univ. of Waterloo, Waterloo, Ont., Canada

Geomorphic patterns produced by the last Canadian ice sheet: Matching the scales of remote sensing with the frequency of natural variation 97
C.D. Clark, G.S. Boulton, Edinburgh Univ., Edinburgh, UK

Page

Methode assistee par ordinateur d'identification et d'analyse des lineaments geologiques a l'aide d'une image ROS (Resultats Preliminaires) 101
J. Daoust, Q.H.J. Gwyn, Univ. de Sherbrooke, Sherbrooke, Que., Canada, C. Roux, Ecole Nat. Super. de Telecomm., Bret., France

Lineament identification and the Edwards Aquifer of Central Texas 105
T. Kleeman, Univ. of Texas, Austin, TX, USA, C.M. Woodruff Jr., Consulting Geologist, Austin, TX, USA

Remote sensing of fracture indicators near the Moodus, Connecticut, seismic area 108
C.A. Shuman, Pennsylvania State Univ., Univ. Park, PA, USA

Origin and significance of Hemavathi-Tirthahalli mega-lineament - a concept 112
K.G. Raj, andian Sp. Res. Organ., Bangalore, India

Application of GIS and remote sensing for lineament mapping of structural features in the Rouyn-Beauchastel property, Quebec 116
S. Gravelle, M.S. Akhavi, Nova Scotia Coll. of Geog. Sci., Lawrencetown, NS, Canada

Application of remote sensing in the evalution of selected earthquake-prone areas in Vietnam 119
D.N. Luu, Inst. for Geophysics, Hanoi, Viet Nam

Digital mosaicing technology for synthetic aperture radar image products 120
F. Schultz, J.B. Mercer, P. Button, INTERA Tech. Ltd., Calgary, Alta., Canada

Satellite image processing workstations for the EOS era 124
W. Markwitz, G. Schreier, H. Rister, DFVLR, Oberpfaffenhofen, FR Germany

Satellite Image Processing System (SIPS) 128
T.C. Sarma, B. Lakshmi, D.S. Jain, B. Jaipal Reddy, K.M.M. Rao, M.L. Mittal, K. Jairam Hebbar, Nat. Rem. Sens. Agency, Hyderabad, India

Digital photogrammetric workstations 130
P. Lohmann, Univ. of Hannover, Hannover, FR Germany

Aspects of digital photogrammetric workstations for GIS data aquisition and processing 134
J. Weidenhammer, Univ. of Hanover, Hannover, FR Germany

Image processing workstations and data bases for quality control of geocoded satellite images 138
G. Schreier, DFVLR, Oberpfaffhofen, FR Germany

Reduction of ground control requirements for SPOT imagery by pass processing 142
B. Sharpe, K. Wiebe, MacDonald Dettwiler, Richmond, BC, Canada

Simulating mapping, sensing and feature extracion in a PC 146
T.N. Truske, BDM Int. Inc., Albuquerque, NM, USA

Image analysis on a Macintosh II 150
D.G. Barber, J.D. Dunlop, J.M. Piwowar, E. LeDrew, Univ. of Waterloo, Waterloo, Ont., Canada

Modeling the effects of rain on communications and remote sensing systems 154
R.K. Crane, Dartmouth College, Hanover, NH, USA

Electromagnetic model for propagation through clouds 155
S. Selim Seker, Bogazici Univ., Istanbul, Turkey

Page

Phase matrices of radiative transfer equation for strongly fluctuated continuous random media 159
Ya-Qiu Jin, Fudan Univ., Shanghai, PR China

Radiobrightness of periodically heated, two-phase media 163
A.W. England, Univ. of Michigan, Ann Arbor, MI, USA

Texture analysis in matching 3D tree models to remotely sensed data 165
D. Jayasinghe, R. Fournier, J. Miller, York Univ., North York, Ont., Canada

A set of surface emissivity simulation for microwave remote sensing from satellite-borne sensors 166
Ya-Qiu Jin, Fudan Univ., Shanghai, PR China

Reduction of interfering signals in swept-frequency radar measurements using complex Weiner filtering 170
D.E. Weissman, Hofstra Univ., Hempstead, NY, USA, B. Jersak, Univ. of Texas, Arlington, TX, USA, L.D. Staton, NASA/Langley Res. Cen., Hampton, VA, USA

A high resolution fully polarimetric C-Band radar 171
A. Stjernman, T. Sturk, G. Wannberg, Swedish Inst. of Sp. Physics, Kiruna, Sweden

Role of space observations in the Canadian Global Change Program 173
J. Cihlar, CCRS, Ottawa, Ont., Canada, E. LeDrew, Univ. of Waterloo, Waterloo, Ont., Canada, H. Edel, Fish. and Oceans Canada, Ottawa, Ont., Canada, W. Evans, G. McKay, Atmos. Env. Serv., Env. Canada, Canada, L. McNutt, CCRS, Ottawa, Ont., Canada, A. Royer, Univ. de Sherbrooke, Sherbrooke, Que., Canada

Solar surface albedo estimation using remotely sensed spectral data 177
D.L. Toll, NASA/GSFC, Greenbelt, MD, USA

Climate change effects on the snowmelt hydrology of western North American mountain basins 181
A. Rango, V. van Katwijk, USDA Hydro. Lab., Beltsville, MD, USA

Using remotely sensed data to monitor land surface climatology variations in a semi-arid grassland 184
L.F. Johnson, N.A. Bryant, NASA/JPL, Pasadena, CA, USA, A.J. Brazel, Arizona State Univ., Tempe, AZ, USA, C.F. Hutchinson, Univ. of Arizona, Tuscon, AZ, USA, R.C. Balling, Arizona State Univ., Tempe, AZ, USA

Adaptation of the ISCCP cloud detection algorithm to combined AVHRR and SMMR Arctic data 188
J. Key, R.G. Barry, Univ. of Colorado, Boulder, CO, USA

Tropical deforestation in Para, Brazil: analysis with Landsat and Shuttle Imaging Radar-A 192
T.A. Stone, G.M. Woodwell, R.A. Houghton, The Woods Hole Res. Cen., Woods Hole, MA, USA

A comparison of AVHRR GAC and HRPT data for regional environmental monitoring 196
A.S. Belward, J-P. Malingreau, Institute for Remote Sensing Applications, Ispra, Italy

Radarsat: Canada's microwave satellite 197
E. Shaw, CCRS, Ottawa, Ont., Canada, E. Langham, Radarsat Proj. Off., Ottawa, Ont., Canada

Page

Radarsat: Review of pertinent economic and commercial issues 200
M. St-Pierre, CSA, Ottawa, Ont., Canada

Radarsat International Inc.: commercial distribution of radarsat products 204
D.R. Inkster, INTERA Tech. Ltd., Calgary, Alta., Canada

Land applications of Radarsat 208
R.J. Brown, CCRS, Ottawa, Ont., Canada

Radarsat: Ice surveillance in an Ice Information System context 211
L. McNutt, CCRS, Ottawa, Ont., Canada, J. Falkingham, Env. Can., Ottawa, Ont., Canada

The Radarsat system 213
S. Ahmed, H.R. Warren, Radarsat Proj. Off., Ottawa, Ont., Canada, M. Symonds, R. Cox, Spar Aerospace Ltd., Ste.-Anne, Qc, Canada

The Radarsat Synthetic Aperture Radar System 218
A.P. Luscombe, Spar Aerospace Ltd., Ste-Anne-de-Bellevue, Que., Canada

User access and production control of Canada's operational SAR data processing facility 222
N. Evans, MacDonald Dettwiler, Richmond, BC, Canada, N. Denyer, CCRS, Ottawa, Ont., Canada

Telemetry antenna for Radarsat 226
A. Kumar, AK Electromagnetic Inc., Dollard-des-Ormeaux, Que., Canada

The need for SAR calibration 230
A. Freeman, NASA/JPL, Pasadena, CA, USA

Practical SAR radiometric calibration measurements and experiments 234
E.S. Kasischke, Env. Res. Inst. of Mich., Ann Arbor, MI, USA

Calibration for airborne SAR; → (over) 238
R.K. Hawkins, T.I. Lukowski, A.L. Gray, C.E. Livingstone, CCRS, Ottawa, Ont., Canada

Approach to derivation of SIR-C science requirements for calibration 243
P. Dubois, D. Evans, A. Freeman, J.J. van Zyl, JPL, Pasadena, CA, USA

SIR-C calibration procedures 247
A. Freeman, J.C. Curlander, NASA/JPL, Pasadena, CA, USA

Progress in calibration: SAR-580 to ERS-1 248
A.M. Smith, D.G. Corr, I. Ward, SD-SCICON, London, UK, J.-P. Guignard, ESTEC, Noordwijk, Netherlands, M. Doherty, ESRIN, Frascati, Italy

Results of the 1988 NASA/JPL airborne SAR calibration campaign 249
A. Freeman, C. Werner, J.D. Klein, NASA/JPL, Pasadena, CA, USA

The Saskatoon SAR calibration experiment 254
T.I. Lukowski, R.K. Hawkins, B. Brisco, R. Brown, R. Ford, CCRS, Ottawa, Ont., Canada, P. Daleman, Intera Technologies Ltd., Ottawa, Ont., Canada

Possibility of the use of a transponder as an active SAR calibration target 258
N. Kodaira, Rem. Sens. Tech. Cen. of Japan, Tokyo, Japan

Effect of calibration errors on measuring the radar cross section of diffuse targets with SAR 261
C.C. Wackerman, E.S. Kasischke, Env. Res. Inst. of Mich., Ann Arbor, MI, USA

Page

Initial results of a synthetic aperture microwave radiometer 264
D.M. Le Vine, M. Kao, NASA/GSFC, Greenbelt, MD, USA, A. Tanner, C.T. Swift, A. Griffis, Univ. of Massachusetts, MA, USA

Beam compressed microwave scanning radiometer 268
K. Arai, NASDA, Saitama, Japan, T. Suzuki, Univ. of Electro-Comm., Tokyo, Japan

Topex microwave radiometer system calibration; refining the SMMR heritage 271
C.S. Ruf, M.A. Manssen, S.E. Keihm, NASA/JPL, Pasadena, CA, USA

Advanced Microwave Scanning Radiometer (AMSR): preliminary study on calibration approach 273
Y. Ito, K. Tachi, Y. Sato, K. Arai, NASDA, Ibaraki, Japan

Millimeter-wave electronically scanned imaging radiometer system for shipborne ice navigation application 277
G. Healy, COM DEV Ltd., Cambridge, Ont., Canada, C. Grant, York Univ., North York, Ont., Canada

Dual frequency microwave radiometer 281
P. Xu, P. Jin, Z. Yao, Z. Fang, M. Li, Z. Wang, J. Zheng, Shanghai Univ., Shanghai, PR China

Soil moisture estimation with L-Band radiometric measurements 285
J.R. Wang, NASA/GSFC, Greenbelt, MD, USA, J.C. Shiue, NASA/GSFC, Greenbelt, MD, USA, E.T. Engman, USDA, Beltsville, MD,

An imaging passive microwave radiometer for precipitaion retrieval and mesoscale storm system studies 286
J.A. Galliano, R.H. Platt, Georgia Inst. of Tech., Atlanta, GA, USA

An improved design for a microwave radiometric profiling system 290
J.R. Jordan, J.B. Snider, T.A. Stermitz, NOAA/ERL/WPL, Boulder, CO, USA

Converting digital passive microwave radiances to kelvin units of brightness temperature 295
L.D. Farmer, D.T. Eppler, A.W. Lohanick, NORDA Polar Ocean. Branch Off., Hanover, NH, USA

Advanced technology processes for satellite instruments 297
J.P. Hackett, J.B. Dorey, R.S.H. Tong, A. El Shakweer, COM DEV Ltd., Cambridge, Ont., Canada

Experimental investigation of imaging radar techniques 303
R.F. Schindel, A.J. Blanchard, Univ. of Texas at Arlington, Arlington, TX, USA

A coaxial-circular waveguide for dielectric measurement 307
M.R. Taherian, Schlumberger-Doll Res., Ridgefield, CO, USA, D.G. Yuen, Dept. of E.E. and Computer Science, MIT USA, T.M. Habashy, Schlumberger-Doll Research USA

Magnetotelluric inversion using approximate inverse mappings 308
D.W. Oldenburg, R.G. Ellis, Univ. of British Columbia, Vancouver, BC, Canada

Fuzzy pattern recognition in geophysics 309
B. Lashgari, Unocal Corp., Brea, CA, USA

Diagnostic electromagnetic seam wave imaging of anomalous geologic disturbance zones in layered formations 310
L.G. Stolarczyk, Stolat, Inc., Raton, NM, USA

Page

Interpretation of the depolariz ects in vector (polarization) diffraction tomography 311
N.A. Soliman, W.-M. Boerner, Univ. of Illinois, Chicago, IL, USA

Recent advances in physical optics inverse scattering theory for non-symmetric, conducting, closed shapes. The depolarizing effects for the monostatic and bistatic cases 312
B.-Y. Foo, W.-M. Boerner, Univ. of Illinois, Chicago, IL, USA

Spherical approximation theory of inverse scattering 313
T. Sengor, Yildiz Univ., Istanbul, Turkey

An automated method for estimating oceanic flow fields from satellite imagery 315
S.D. O'Keefe, K.M. Crawford, Univ. of Texas at Austin, Austin, TX, USA

AVHRR observations of the horizontal structure of the surface layer of the ocean under low wind conditions 318
P. Flament, Univ. of Hawaii, Manoa, HI, USA

Satellite observed spatial and temporal variability of the mixed layer depth in the Sargasso Sea 323
Xiao-Hai Yan, M.J. Bowman, J.R. Schubel, State Univ. of New York, Stony Brook, NY, USA

Spectral analysis of the AVHRR sea surface temperature variability off the west coast of Vancouver Island 327
R. Burgert, Univ. of British Columbia, Vancouver, BC, Canada, W.W. Hsieh, Univ. of British Columbia, Vancouver, Canada

Sea surface thermal signature and estuarine flushing events in the Straits of Georgia and Juan de Fuca 331
P.H. LeBlond, D.A. Griffin, D. Laplante, Univ. of British Columbia, Vancouver, BC, Canada

Water quality mapping of Augusta Bay, Italy from Landsat-TM data 335
S. Khorram, H. Cheshire, N. Carolina State Univ., Raleigh, NC, USA, A.L. Geraci, G. La Rose, Univ. di Catania, Catania, Italy

Rivers in the sea: Can we quantify pigments in the Amazon and the Orinoco river plumes from space? 339
F.E. Muller-Karger, J.J. Walsh, K.L. Carder, Univ. of South Florida, St. Petersburg, FL, USA, R.G. Zika, Univ. of Miami, Miami, FL, USA

The radar image of the turbulent wake generated by a moving ship 343
J.K.E. Tunaley, J.R. Dubois, J.B.A. Mitchell, London Res. and Dev., London, Ont., Canada

Detection of low contrast features in SAR images 347
K. Nicolas, Naval Res. Lab., Washington, DC, USA

Effect of the Kelvin wake transverse wave component on SAR imagery of surface wave wakes 348
B.A. Hughes, Def. Res. Est. Pacific, Victoria, BC, Canada

The use of the Radon transform for wake detection in Seasat images 354
M.T. Rey, DREO, Ottawa, Ont., Canada, J.T. Folinsbee, R.D. Ball, Imago Mfg. Ltd., Ottawa, Ont., Canada, J.T. Tunaley, London Res. &. Dev., London, Ont., Canada, M.R. Vant, DREO, Ottawa, Ont., Canada

Principles and performance of an automated ship detection system for SAR images 358
K. Eldhuset, Norwegian Defence Res. Estab. (NDRE), Kjeller, Norway

Page

The effect of ship's screws on the ship wake and its implication for the radar image of the wake 362
E.H. Buller, J.K.E. Tunaley, London Res. and Dev., London, Ont., Canada

SAR imaging of ships and ship wakes during NORCSEX '88 366
T. Wahl, A. Skoelv, Norwegian Def. Res. Estab., Kjeller, Norway

Active microwave classification of sea ice 369
R.G. Onstott, Env. Res. Inst. of Mich., Ann Arbor, MI, USA

Discrimination between new sea ice and open water - comparison of C-Band theory with measurements 375
D.P. Winebrenner, Univ. of Washington, Seattle, WA, USA, P. Gogineni, J. Bredow, Univ. of Kansas, Lawrence, KS, USA, L. Tsang, Univ. of Washington, Seattle, WA, USA

Combined active/passive microwave classification of sea ice 376
C.E. Livingstone, CCRS, Ottawa, Ont., Canada

The discrimination of sea ice types using SAR backscatter statistics. (EDC) 381
R.A. Shuchman, C.C. Wackerman, A.L. Maffett, R.G. Onstatt, L.L. Sutherland, Env. Res. Inst. of Michigan, Ann Arbor, MI, USA

Detection and characterization of ice ridges in the Baltic Sea using CV-580 SAR imagery 386
R. Johansson, Chalmers Univ. of Tech., Gothenburg, Sweden

LIMEX'87 ice surface characteristics and their effect upon C-Band SAR signatures 390
M.R. Drinkwater, NASA/JPL, Pasadena, CA, USA

Extraction of sea ice information from SAR imagery 394
T.I. Olausen, O.M. Johannessen, R. Karpus, Nansen Remote Sensing Centre, Bergen, Norway

Determining Intrinsic Surface Reflectance in Rugged Terrain and Changing Illumination

Robert J. Woodham*

Laboratory for Computational Vision
University of British Columbia
Vancouver, BC, Canada

Abstract

Remote sensing measurement is a many-to-one mapping because multiple scene factors, including topography, ground cover, illumination and atmosphere, interact in each measurement. Remote sensing measurements cannot be interpreted directly as scene properties because the inversion problem, formally posed, is underconstrained. In rugged terrain, one must separate changes in measured brightness owing to shape from changes owing to surface material under spatially and spectrally varying conditions of illumination and atmospheric attenuation. Typically, trade-offs arise that cannot be uniquely resolved.

One idea is to use additional scene knowledge in the form of a digital terrain model (DTM) and, where available, existing ground cover maps. This allows an image to be synthesized for any given date and time. Synthesis is based on an image irradiance equation that combines the bidirectional reflectance distribution function (BRDF) of the surface material, the spatial and spectral distribution of light sources and a simple model of atmospheric attenuation and path radiance. The unknown parameters of the model are physical parameters that can, in principle, be supplied externally. Otherwise, they are estimated from the real image. The parameters are used in radiometric correction to estimate an intrinsic reflectance factor related to ground cover.

Keywords: radiometric correction, reflectance, digital terrain model, topography, illumination, atmosphere

Introduction

Consider a hypothetical nadir viewing satellite, in perfect geosynchronous orbit, that repeatedly views the identical portion of the earth's surface. As a first guess, one might suppose that such a satellite would produce copies of what is essentially the same image. But, this would not be so. The direction of dominant illumination changes with the time of day as the sun follows its trajectory from sunrise to sunset. Similarly, sky radiance and atmospheric transmission change with weather and other factors. The measurements obtained by the hypothetical satellite would change, and these changes would be more acute in areas of rugged terrain.

Consider a second hypothetical satellite, also in perfect geosynchronous orbit, that repeatedly views the same portion of the earth's surface but from an off-nadir viewpoint. As a first guess, one might suppose that both satellites would record identical measurements when viewing identical points on the earth's surface. Again, this would not be so. In general, the amount of light reflected by a surface element depends on the relative geometry of surface and viewer. The measurements obtained by the two hypothetical satellites would differ, and these differences would be more acute in areas of rugged terrain and when atmospheric attenuation is noticeable.

Thus, radiometric correction is not simply a process to make multiple images appear to have been measured by a single standard sensor. Radiometric correction also is required to account for differences in illumination, atmosphere and viewpoint, even if measurements are acquired by a single sensor.

For ten years now, I have considered the problem of radiometric correction of multispectral scanner data posed as the problem of determining an intrinsic reflectance factor characteristic of the surface material being imaged that is invariant to topography, position of the sun, atmosphere and position of the viewer. Four papers describe the development of the key analytic tools and related experimental results (Woodham, 1980a), (Woodham *et al.*, 1985), (Woodham and Lee, 1985) and (Woodham and Gray, 1987). In this paper, I summarize the relevant technical considerations and experimental results. I also suggest, based on my experience, implications that this work has to the general task of determining intrinsic surface reflectance in rugged terrain and changing illumination.

Technical Considerations

Bidirectional Reflectance Distribution Function

The intrinsic reflectance of a surface material is specified by its bidirectional reflectance distribution function (BRDF). The BRDF was introduced by (Nicodemus *et al.*, 1977) as a unified notation for the specification of reflectance in terms of both the incident and reflected ray geometry. The BRDF, denoted by the symbol f_r, is the ratio of the reflected radiance, dL_r, in the direction toward the viewer to the irradiance, dE_i, in the direction from a portion of the source. That is,

---

* Fellow of the Canadian Institute for Advanced Research.

$$f_r(\theta_i,\phi_i;\theta_r,\phi_r) = \frac{dL_r(\theta_i,\phi_i;\theta_r,\phi_r;E_i)}{dE_i(\theta_i,\phi_i)} \quad [\mathrm{sr}^{-1}] \qquad (1)$$

Directions are given in spherical coordinates (θ,ϕ). Subscript i denotes quantities associated with the incident radiant flux and subscript r denotes quantities associated with the reflected radiant flux.

A Lambertian surface has BRDF

$$f_r = \frac{\rho}{\pi} \qquad (2)$$

where ρ is called the bidirectional reflectance factor $(0 \le \rho \le 1)$. When $\rho = 1$ the Lambertian surface is perfectly diffuse (lossless). Since Equation (2) doesn't depend on (θ_i,ϕ_i) or (θ_r,ϕ_r), the parameter ρ is invariant to conditions of illumination and viewing. For surfaces that are not Lambertian, the bidirectional reflectance factor, ρ, is defined to be the ratio of the radiant flux reflected by a sample surface to that which would be reflected into the same reflected-beam geometry by an ideal (lossless) Lambertian standard surface irradiated in exactly the same way as the sample. In general, this ratio depends on (θ_i,ϕ_i) and (θ_r,ϕ_r). Thus, for the general case, the bidirectional reflectance factor, ρ, is not a scalar but, like the BRDF itself, it is a function of (θ_i,ϕ_i) and (θ_r,ϕ_r).

The BRDF allows one to derive scene radiance, L_r, for any light source distribution and viewer geometry by integrating over the specified solid angles. A systematic approach to this problem is described in (Horn and Sjoberg, 1979), including results for the Lambertian case given below. Some simple situations, like the Lambertian case, lead to closed form solutions of the associated integral. But, most general situations are more difficult and closed form solutions are rare.

Scene radiance for a Lambertian surface illuminated by a collimated source with irradiance E_0 measured perpendicular to the beam of light arriving from direction (θ_0,ϕ_0) is given by

$$L_r = \frac{E_0}{\pi}\,\rho\,\cos(i) \qquad (3)$$

Equation (3) is often considered the defining equation of a Lambertian surface. In fact, it is a derived equation for one particular illumination. By comparison, when illuminated by a hemispherical uniform "sky" with radiance L_0 over the visible hemisphere, one obtains

$$L_r = L_0\,\rho\,\frac{1+\cos(s)}{2} \qquad (4)$$

where s is the slope angle of the surface measured with respect to the horizontal. The dependence on s arises because differing surface elements see differing amounts of sky depending on surface slope. As slope increases, more of the sky is obscured. (Note, this is purely a local effect and does not take occlusion of the sky by adjacent terrain into account.) The essential point is that Equations (3) and (4) differ only due to the difference in illumination. Both correspond to a Lambertian surface.

Spectral Properties

The BRDF also depends on the wavelength λ of the radiation in question. To make this dependence explicit, let $f_r(\theta_i,\phi_i;\theta_r,\phi_r;\lambda)$ be the spectral bidirectional reflectance distribution function (SBRDF). Selective reflection can alter the spectral distribution of the reflected beam. If there is interaction between spectral and geometric factors, as can be the case for materials with significant internal scattering, then the geometric distribution also is affected. On the other hand, if there is no interaction between wavelength and the geometric dependence of reflection then

$$f_r(\theta_i,\phi_i;\theta_r,\phi_r;\lambda) = f_r(\theta_i,\phi_i;\theta_r,\phi_r)\,f_r(\lambda) \qquad (5)$$

where $f_r(\lambda)$ is a weighting function that determines relative reflection as a function of λ. If equation (5) holds, the SBRDF is said to be separable. Otherwise, there will necessarily be a change in the spectral distribution of scene radiance as a function of the surface and viewer geometry. The SBRDF of a Lambertian surface is separable with reflectance factor, $\rho = \rho(\lambda)$, a function of λ. On the other hand, some materials noticeably change color when viewed from different directions, as nonseparability requires. (One example is the neck feathers of certain waterfowl that change color with movement due to the presence of significant internal scattering by wax particles.)

Modeling and Measuring Reflectance

The question, "How does one measure reflectance?" is similar to the question (Mandelbrot, 1982), "How long is the coast of Britain?" because the answer inevitably depends upon the scale at which the measurement occurs. The BRDF, defined by Equation (1), is a derivative with instantaneous values that can never be measured directly. Any real measurement involves incident and reflected beams and hence can yield only average values of f_r over the finite solid angles subtended by the light source and viewer. In particular, an imaging device necessarily measures radiance reflected from a surface element with finite area and hence can yield only average values of f_r over the area subtended by the instantaneous field of view (IFOV) of the sensor. Most natural surfaces are not optically smooth across all scales. Consequently, average values over the IFOV of the sensor are highly dependent on spatial scale and are difficult to relate to the underlying BRDF's and microstructure of the surface material.

Nevertheless, relating image brightness to models of surface microstructure is an important tool in several fields, including: lunar astronomy (Minnaert, 1961), (Hapke, 1971); reflectance spectroscopy (Wendlandt and Hecht, 1966); robot vision (Horn, 1986); and computer graphics (Cook and Torrance, 1982). These results have been increasingly applied to remote sensing, especially to model vegetation canopies (Kimes, 1983), (Goel and Deering, 1985), (Li and Strahler, 1985). One goal is to invert the canopy reflectance model to determine agrophysical parameters, given measured reflectances. Agrophysical parameters include: leaf reflectance, leaf transmittance, leaf area index (LAI), leaf angle distribution (LAD), planting density and direction, biomass, as well as reflectances and transmittances of underlying structures such as stems and soil. A second goal is to predict how a given canopy will appear under different conditions of illumination and viewing.

Composite materials with significant surface roughness become more difficult to model as geometric and radiometric factors interact. For example, a simple thought experiment suggests that open forest cover is not spectrally separable at typical remote sensing measurement scales because there can be a differing amount of green (canopy) versus brown (bare soil) seen per pixel as the sensor moves from directly overhead to an oblique view. Some canopy reflectance models assign different spectral characteristics to each constituent component to account for this phenomenon.

Basic theoretical modeling also is increasingly supported by field measurement. Field measurement is difficult, compared to laboratory measurement, in part, because it is not possible to control the illumination and, in part, because it is difficult to match the scale of measurement in the field to the scale of measurement of the satellite images to which the model will be applied. Thus, it is difficult to use measurements acquired under one condition of illumination and viewing to predict reflectance for another condition of illumination and viewing. To do that requires knowledge of the BRDF and, as noted above, the BRDF itself is not directly measurable.

An Image Irradiance Equation

Complete modeling of all components of the image irradiance equation is not now feasible. Let $z = f(x,y)$ be the elevation of target point (x,y) and let $E(x,y)$ be the measured image irradiance. (Sjoberg and Horn, 1983), (Woodham and Lee, 1985) and (Woodham and Gray, 1987) develop the image irradiance equation

$$E(x,y) = T_u \frac{\rho}{\pi} \left(E_0 T_d \cos(i) + E_S \frac{(1 + \cos(s))}{2} \right) + L_P \quad (6)$$

where

T_u is the upward transmission through the atmosphere,
E_0 is the solar irradiance at the top of the atmosphere,
T_d is the downward transmission through the atmosphere,
i is the solar incident angle at the target,
E_S is sky irradiance, integrated radiance over the hemisphere of the sky,
s is the surface slope,
L_P is path radiance and
ρ is the target's bidirectional reflectance factor.

Expressions for T_u and T_d are estimated as functions of elevation using optical thickness, $\tau(z)$, as an auxiliary quantity where

$$\tau(z) = \tau_0 e^{-z/H_\tau} \quad (7)$$

Parameter $\tau_0 = \tau(0)$ is the optical thickness at sea-level and parameter H_τ is the scale height. The upward and downward transmissions then become

$$T_u(z) = e^{-\tau(z)} \quad (8)$$

$$T_d(z) = e^{-\tau(z)/\cos(\theta)} \quad (9)$$

where θ is the solar zenith angle. Similarly, E_S and L_P are estimated as functions of elevation.

$$E_S(z) = E_{S0} e^{-z/H_S} \quad (10)$$

$$L_P(z) = L_{P0} e^{-z/H_P} \quad (11)$$

where parameters $E_{S0} = E_S(0)$ and $L_{P0} = L_P(0)$ are the sky irradiance and path radiance at sea-level and parameters H_S and H_P are the respective scale heights. The image irradiance equation (6) is expressed in terms of six atmospheric parameters τ_0, H_τ, E_{S0}, H_S, L_{P0} and H_P. These all are physical parameters that can be determined independently or estimated directly from remote sensing data. If these parameters are known and if z, $\cos(i)$ and $\cos(s)$ are determined from the DTM, then Equation (6) can be solved for ρ, the intrinsic reflectance factor of the surface material.

Experimental Results

The main study site is a 21.6 km by 30.4 km area surrounding St. Mary Lake in southeastern British Columbia, Canada. The area has rugged terrain with elevations varying from 944 m to 2684 m above sea-level. Topographic effects dominate Landsat MSS images in areas of rugged terrain. This led to the development of automatic methods for geometric rectification. These methods use a DTM and the known position of the sun to predict image features that can be reliably and accurately located (Little, 1982). Once an image is geocoded, the next step was to try to remove variations in brightness due to slope and aspect in order to better delineate changes due to ground cover. Experimental work consistently demonstrated an additional dependence of brightness on elevation (Woodham, 1980a). The minimum recorded brightness in any band is a decreasing function of elevation. The decrease is especially noticeable in the shorter wavelength bands. New methods to estimate path radiance, L_P, were based on this observation (Woodham and Lee, 1985).

Sky radiance also was shown to be significant. (Woodham, 1980a) noted that bright targets (eg., snow) in shadow were often brighter than dark targets (eg., conifer forest) in direct sunlight. One attempt to measure sky irradiance used cast shadow boundaries (Woodham and Lee, 1985). The location of cast shadow boundaries varies with the position of the sun. One can therefore assume that the ground cover under a cast shadow boundary remains locally constant, provided that one excludes cast shadows that happen also to coincide with terrain breaks. Further, one can assume that sky irradiance and optical thickness also are locally constant at cast shadow boundaries. Estimates were then obtained for τ_0, H_τ, E_{S0} and H_S.

Results, based on cast shadow boundaries, were not as successful as expected. There are several possible explanations. In hindsight, points near cast shadow boundaries are poor candidates to use to estimate overall sky irradiance because they correspond to points for which a significant fraction of the sky is occluded by adjacent terrain.

Greater success was achieved using snow as a target of known reflectance (Woodham and Lee, 1985). Radiometric correction was demonstrated using a very low sun angle (elevation 13.8°) January 8, 1979, Landsat MSS image of St. Mary Lake with 43% of the study site in shadow.

Subsequently, we acquired machine readable forest cover data for the St. Mary Lake area from the B.C. Ministry of Forests. The forest cover map data was used to determine areas of homogeneous ground cover. We looked for forest types that occurred over a wide range of elevation, slope and aspect. For homogeneity, we also looked for uniform age and height and, preferably, a closed canopy. Unfortunately, very

few natural vegetation types occur over a wide range of topographic positions. After much search, a target forest type was selected with the following attributes: Lodgepole pine ($\geq$ 80% pure); 80-100 years old; 10-20 m high; and 60-100% crown closure. Over the study site, the target set had the following range of topographic attributes: elevation 1205-2102 m; incident solar angle i 21.7-78.0 degrees; and slope angle s 7.2-35.7 degrees. The final target set consisted of 329 pixels.

Multiple linear regression was applied to estimate the parameters of Equation (6). Of course, the model equation itself is not linear so that some reformulation was required. If the terms T_u, T_d, E_S and L_P each are approximated by a function of the form $a\,z + b$ then the resulting model equation is linear in the eight terms $\cos(i)$, $\cos(s)$, z, $z\cos(i)$, $z\cos(s)$, z^2, $z^2\cos(i)$ and $z^2\cos(s)$. Regression analysis for the target set showed that not all eight terms were correlated, at the 99% significance level, to measured brightness. Various subsets of the eight terms were then considered before settling on a model involving the four variables, $\cos(i)$, $(1 + \cos(s))/2$, z and z^2. The reason for expressing the $\cos(s)$ term in this way is to retain the connection to Equation (4). Regression involving these four variables corresponds to a model in which target radiance depends on the cosine of the solar incident angle, the cosine of the slope and elevation (true of (6)), but with no coupling between them (not true of (6)).

As expected from previous work, the cosine of the solar incident angle is the most important variable for all bands, followed by elevation z. The third most important term was the cosine of the slope. Unexpected to us, however, the correlation with $\cos(s)$ was consistently negative, once corrections for $\cos(i)$ and z had been applied. One possible explanation for this negative correlation is that it arises as an artifact of the correction applied for $\cos(i)$. (This could happen, for example, if the correlation with $\cos(i)$ were high but the relationship was nonlinear). The way to rule this out would be to repeat the analysis using targets of constant $\cos(i)$. Unfortunately, there were not enough data points of constant $\cos(i)$ to yield a statistically significant conclusion, one way or the other.

Another possibility is that treating skylight as a uniform hemispherical source is unrealistic. In order to test other models of sky radiance, the sky hemisphere was uniformly tessellated into discrete cells. A point source was associated with each cell, weighted according to sky radiance integrated over the solid angle subtended by that cell. A synthetic image then is generated and summed with those from all other point source locations. Because shadow calculation is included from each point source direction, the final sum excludes, at each point, directions obscured from the sky by adjacent terrain. It was a simple matter to replace the $(1 + \cos(s))/2$ term in Equation (6) with one corresponding to other analytic and measured models of skylight. When this was done, the negative correlation with skylight persisted in all cases tried. Occlusion of the sky by adjacent terrain means that the amount of sky seen at each surface element is no longer a local function of the slope s, as is the case in (4). Negative correlation with skylight can, in fact, be interpreted as positive correlation with adjacent terrain.

A numerical model of sky radiance was computed, based on a uniform hemispherical source but taking occlusion by adjacent terrain into account. The regression analysis was repeated and the numerical model produced a slightly higher coefficient of determination and a slightly lower standard error in all cases and for all bands. Radiometric correction was applied to each Landsat MSS band.

Standard statistical tests were applied to validate the analysis. Correlation coefficients for the regression analysis were computed. Examination of regression residuals as a function of the independent variables indicated no discernible trends. The results also were examined for correlation between the regression variables. (High correlation between regression variables would make the analysis unstable.) There is almost no correlation between $\cos(i)$ and the other three variables. Not surprisingly, significant correlation exists between $\cos(s)$ and z since the more rugged terrain tends to occur at higher elevations.

A practical concern is whether this method of radiometric correction improves the accuracy of standard spectral classification. A simple classification was performed, the details of which are reported in (Gray, 1986). A nearest centroid classifier was used based on four ground cover classes: forest, clearcut, water and alpine. A truth map was constructed from the forest cover map. The classification of the uncorrected Landsat MSS image resulted in a map accuracy of 51% for forest cover. Examination of the corresponding confusion matrix revealed deficiencies typical of remote sensing of forest cover in rugged terrain. That is, many forest slopes with southeast aspect were miss-classified as clearcut and many forest slopes with northwest aspect were miss-classified as water. The classification of the corrected Landsat MSS image resulted in a map accuracy of 80% for forest cover. Correction for $\cos(i)$ achieved the greatest increase in classification accuracy. Elevation correction, in addition, decreased the miss-classification of forest as water. Correction for skylight, while statistically significant in the regression, had little effect on the final classification.

Discussion

My research strategy has been to insist that parameters determining image irradiance be related directly to physical models. Idealized physical models correctly characterize simple worlds and can be elaborated as the need is demonstrated. Curiously, there has been strong debate in the remote sensing literature concerning the assumption of Lambertian reflectance. Usually, the debate concerns whether or not Equation (3) provides a good fit to the data. Any image irradiance equation necessarily encodes assumptions both about the BRDF of the surface material and about the distribution of the light sources. My work with St. Mary Lake indicates that models that consider only direct solar illumination are inadequate, regardless of the assumption made about the target BRDF. Sky radiance must be dealt with explicitly. The relative amount of direct solar illumination and diffuse sky illumination varies with slope and aspect. The spectral distribution also noticeably varies because skylight is stronger in the shorter wavelengths than is direct sunlight. In areas of rugged terrain, adjacent terrain makes a small, but nevertheless significant, contribution.

The development of additional test sites is essential for further work on radiometric correction. Comparison of physical models would benefit from *in situ* measurement of

atmospheric transmission and sky radiance at the time of satellite overflight.

The development and availability of appropriate software and hardware tools for image synthesis has been invaluable to our work. Visual comparison of a synthetic image and the corresponding real image often conveys a strong sense of the adequacy and robustness of the model being tested. One can look at the radiometric correction to see if shadows are removed, if known homogeneous areas appear homogeneous independent of slope and aspect and if atmospheric corrections are appropriate over the full range of elevations that occur in the scene. Local anomalies in methods that otherwise have good global performance become readily apparent. In my view, the local anomalies often provide the greatest insight.

Of course, formal evaluation criteria also need to be developed. The only obvious criterion is that the estimated bidirectional reflectance factor for Lambertian surfaces must lie between zero and one. The methods described generally achieve this except along seams of slight shadow miss-registration or along sharp ridges that have been smoothed in the DTM. Another possible criterion would be consistency in the estimated reflectance factor over time. Unfortunately, reflectance depends on many factors, including some like surface moisture that vary rapidly over the time period between successive Landsat overflights. It would be useful to test these methods on images acquired at different times on the same day.

Radiometric correction becomes increasingly important when dealing with images acquired from different sensors, at different times of day, and with both nadir and oblique viewing capability. Radiometric correction requires that the local incident and reflected beam geometry be known. It is therefore important that this local information be retained in geometric rectification.

On the other hand, it can be advantageous to obtain multiple images of a given scene under the same geometry but with differing conditions of illumination. This idea has been exploited in robot vision (Woodham 1980b) (Horn and Ikeuchi 1984). Different conditions of illumination provides additional constraint to help to separate the effects of topography from those of ground cover.

I conclude this discussion with an anecdote concerning one of my first Ph.D. students. During our regularly scheduled research meetings, he often would say, "You know the problem we discussed last week? Well, it's a lot more difficult than we thought". What was unique in his case was that he would be saying this with great enthusiasm. Invariably, he would be full of new ideas and new experiments to try. In short, he was ready for the challenge. It is in the same spirit that I now say to you, "You know the problem we've been working on for the last ten years? Well, it's a lot more difficult than we thought.".

Acknowledgement

Direct contributions to the work described here were made by many students, staff and colleagues, especially E. Catanzariti, M. Gray, S. Kingdon, T. Lee, J. Little, A. Mackworth, M. Majka and F. Wong. Help in the preparation of figures, used in the talk but not included in the paper, was provided by N. Holm. Several organizations contributed to the acquisition of test data, including the BC Ministry of Forests and MDA. This paper describes research done at the Laboratory for Computational Vision of the University of British Columbia. Support for the Laboratory's remote sensing research was provided, in part, by the UBC Interdisciplinary Graduate Program in Remote Sensing, by NSERC grants A3390, A0383, E0008 and SMI-51, and by the Canadian Institute for Advanced Research.

References

1. Cook, R.L. & K.E. Torrance (1982), "A reflectance model for computer graphics", *ACM Transactions on Graphics* (1)7-24.
2. Goel, N.S. & D.W. Deering (1985), "Evaluation of a canopy reflectance model for LAI estimation through its inversion", *IEEE Transactions on Geoscience and Remote Sensing* (23)674-684.
3. Gray, M.H. (1986), "Radiometric correction of satellite imagery for topographic and atmospheric effects", M.Sc. Thesis, UBC Faculty of Forestry, Vancouver, BC.
4. Hapke, B. (1971), "Optical properties of the lunar surface", in *Physics and Astronomy of the Moon (2nd edition)*, Z. Kopal (ed.), pp 155-211, Academic Press.
5. Horn, B.K.P. (1986), *Robot Vision*, MIT Press/McGraw-Hill, Cambridge, MA.
6. Horn, B.K.P. & K. Ikeuchi (1984), "The mechanical manipulation of randomly oriented parts", *Scientific American* (251)100-111 (August, 1984).
7. Horn, B.K.P. & R.W. Sjoberg (1979), "Calculating the reflectance map", *Applied Optics* (18)1770-1779.
8. Kimes, D.S. (1983), "Dynamics of directional reflectance factor distributions for vegetation canopies", *Applied Optics* (22)1364-1372.
9. Li, X. & A.H. Strahler (1985), "Geometric-optical modeling of a forest canopy", *IEEE Transactions on Geoscience and Remote Sensing* (23)705-721.
10. Little, J.J. (1982), "Automatic registration of Landsat MSS images to digital elevation models", *Proc. Workshop on Computer Vision: Representation and Control*, pp 178-184, Rindge, NH.
11. Mandelbrot, B.B. (1982), *The Fractal Geometry of Nature*, W.H. Freeman, San Francisco, CA.
12. Minnaert, M. (1961), "Photometry of the moon", in *The Solar System Vol. III: Planets and Satellites*, G.P. Kuiper & B.M. Middlehurst (eds.), pp 213-248, Univ. Chicago Press.
13. Nicodemus, F.E., J.C. Richmond, J.J. Hsia, I.W. Ginsberg & T. Limperis (1977), "Geometrical considerations and nomenclature for reflectance", NBS Monograph 160, National Bureau of Standards, Washington, DC.
14. Sjoberg, R.W. & B.K.P. Horn (1983), "Atmospheric effects in satellite imaging of mountainous terrain", *Applied Optics* (22)1702-1716.
15. Wendlandt, W.W. & H.G. Hecht (1966), *Reflectance Spectroscopy*, John Wiley & Sons, Inc., New York, NY.
16. Woodham, R.J. (1980a), "Using digital terrain data to model image formation in remote sensing", *Proc. SPIE* (238)361-369.
17. Woodham, R.J. (1980b), "Photometric method for determining surface orientation from multiple images", *Optical Engineering* (19)139-144.
18. Woodham, R.J., E. Catanzariti & A.K. Mackworth (1985), "Analysis by synthesis in computational vision with application to remote sensing", *Computational Intelligence* (1)71-79.
19. Woodham, R.J. & M.H. Gray (1987), "Analytic method for radiometric correction of satellite multispectral scanner data", *IEEE Transactions on Geoscience and Remote Sensing* (25)258-271.
20. Woodham, R.J. & T.K. Lee (1985), "Photometric method for radiometric correction of multispectral scanner data", *Canadian Journal of Remote Sensing* (11)132-161.

IDENTIFICATION - THE GOAL BEYOND DISCRIMINATION

The status of mineral and lithological identification from high resolution spectrometer data: Examples and challenges

J.F.Huntington, A.A.Green and M.D.Craig

CSIRO Division of Exploration Geoscience, North Ryde, Australia

BACKGROUND

A major task of the average practising geologist, whether in a government mapping agency or an exploration company, is the identification of primary earth materials (rocks and minerals) and their secondary weathering products (the regolith), primarily from their mineralogical and chemical composition. The three-dimensional disposition of these materials is also of major importance.

For decades geologists have successfully used remotely sensed images, including aerial photographs, to assist them in this process. This remote sensing contribution has been based on two qualitative aspects of photogeology, namely "discrimination" of image differences (using the normal criteria of tone, colour, texture, shape, context, drainage, resistance, etc.) and "interpretation" which is an experience-dependent process of identification based upon inference, deduction and, hopefully, field checking and iterative refinement.

There is now ample evidence to show that this interpretation process can and should be refined much further by making use of information about physical properties inherent in the latest high resolution, remotely sensed data sets that allows a much more rigorous "identification" of earth surface materials.

Improved identification, using the quantifiable, wavelength dependence of spectral reflectance of geological materials obtained by the latest high spectral resolution scanners, provides the possibility of a new generation of geological maps. In certain environments these quite new or revised maps will have an increased mineralogical emphasis, will be more reliable in their lateral continuity, will require less field-checking, and thus may be produced more cheaply and efficiently than existing maps made only by traditional means. From more reliable, up-to-date and useable geological maps can flow numerous cost savings in exploration such as the placement of expensive resources into the right parts of remote areas. We are in no way advocating machines taking over from field geologists, but rather helping them be more cost effective in their daily tasks.

The production of data with a greater physical and hence geological basis should also serve to increase the acceptance of remote sensing as a reliable geoscientific tool amongst the many geologists who still regard the subject as an interesting, but uneconomic, fringe discipline.

SPECTRAL REFLECTANCE: THE PHYSICAL BASIS FOR MINERAL MAPPING

The last fifteen years have seen the establishment of many libraries of spectral reflectance characteristics of geological materials at visible, near-infrared (VNIR), shortwave-infrared (SWIR) and mid-infrared (MIR) wavelengths (Hunt and Salisbury, 1970-1976; Salisbury et al., 1987). From such data it is possible to identify minerals using the characteristic patterns of absorption, reflection or emission due to charge transfer and electronic transitions in the transition elements (Fig.1), overtone bending-stretching vibrations in hydroxyl (Fig.2) and carbonate-bearing materials and fundamental Si-O vibrations in silicate materials (Fig.3). Not all minerals can be identified by these means. Table 1 summarizes the major groups for which identification is possible.

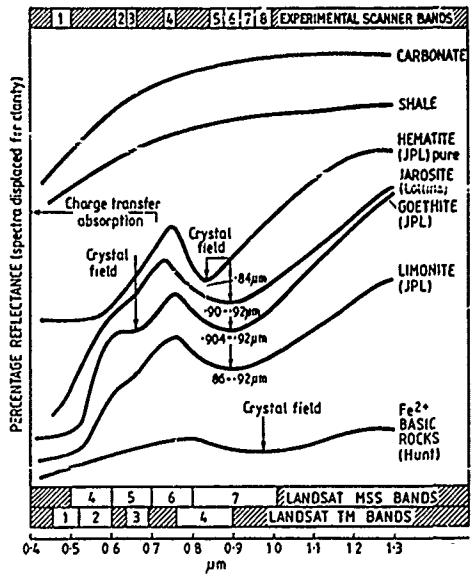


Figure 1. Reflectance spectra of the major iron oxide minerals in the VNIR portion of the electromagnetic spectrum.

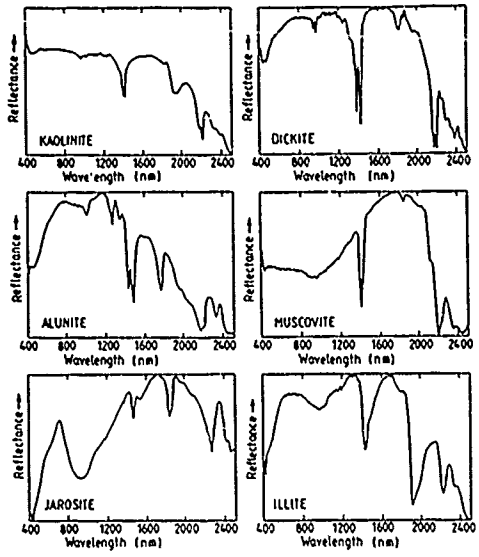


Figure 2. Reflectance spectra of selected OH-bearing minerals in the VNIR and SWIR portion of the electromagnetic spectrum.

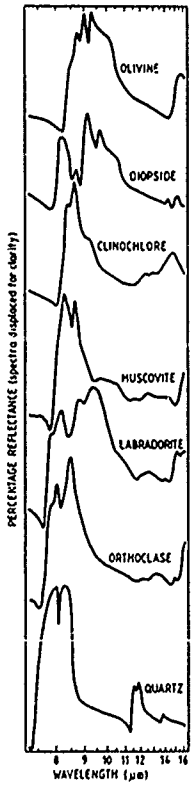


Figure 3. Reflectance spectra of selected silicates in the MIR portion of the electromagnetic spectrum.

TABLE 1 - WAVELENGTH AND MINERALOGICAL SIGNIFICANCE

REGION	WAVELENGTH (μm)	MAJOR MINERALS SENSED	
VNIR	0.40-1.20	Transition elements, Fe, Mn & Ni oxides	
		Haematite, goethite, lepidocrosite	
		REE-bearing minerals	
	0.50-0.80	Vegetation	
SWIR	1.30-2.50	Hydroxides, carbonates, & sulphates generally	
	1.47-1.82	Sulphates	- Alunite
	1.47-1.76	Sulphates	- Jarosite
	2.16-2.24	Al-OH minerals	- Muscovite,kaolinite
			- Dickite,pyrophyllite
			- Smectites,illite
	2.24-2.26	Si-OH minerals	- Opaline silica
	2.24-2.30	Fe-OH minerals	- Jarosite,hectorite,saponite
	2.30-2.40	Mg-OH minerals	- Chlorite,talc,epidote
			- Amphibole
	2.32-2.35	Carbonates	- Calcite,dolomite,magnesite
			- Siderite
MIR	8.00-14.0	Silicates	- Quartz,feldspars,pyroxenes
			- Olivines

The characteristics that permit identification of spectra are associations of wavelengths of absorption or emission, absorption depth and absorption width at a certain depth. These criteria have been used sucessfully in expert systems for automatic mineral identification from reflectance spectra (Lister et al., 1987; Horsfall et al., 1987; Ali et al., 1989) for use in field-portable spectrometers now under development or already on the market.

Although remote sensing instruments are as yet unable to measure reflectance directly, present models of reflectance (Green and Huntington, 1987) and modern pre-processing methods are getting closer to "calibrating" and "transforming" remotely acquired radiance measurements (that contain atmospheric and surface illumination effects as well as the all important mineral specific signals) so that they can be identified by the same basic mineral and vegetation reflectance criteria that are used in the laboratory. Various methods are available for this purpose. We have used the Least Upper Bound, Hull Quotient and Log Residual techniques quite effectively (Green and Craig, 1985), primarily on the grounds that whatever method is used can only be useful operationally if the "calibration" comes from the data itself or from the instrument. Dependence on ground-acquired information is not in our belief a practical operational solution. A comparison of several calibration methods is given by Roberts et al. (1985).

Our methods are by no means perfect and the fundamental problems they address, that of calibration of the physical attributes of reflectance, plus the additional problem of mixed pixels (most severe in cases of vegetation/mineral mixing) are being addressed by other keynote speakers at this conference. A third problem in this whole field is the paucity of suitable sensor systems around the world, not just for research but more importantly for operational applications.

MINERAL MAPPING SPECIFICATIONS

To achieve mineralogical identification of surface materials we must clearly and realistically understand the necessary instrument specifications, in terms of spatial and spectral resolution, radiometric sensitivity and data handling, that must be met if the product is to provide the real geological benefits sought.

Current geological experience suggests that in the visible and near-infrared between 15 and 20, 40-30 nm bands are adequate for transition element detection and for separating geological responses from those of dry and green vegetation. This is in fact less than most current spectrometers provide. Higher resolutions are likely to be required for vegetation studies or detection of some specific materials such as rare earth element bearing minerals. In the shortwave infrared about 64, 8 nm bands are required between 2.0 and 2.5 μm. The region between 1.9 and 2.0 μm, ignored by many people because it encompasses several atmospheric absorption windows, also needs to be sampled with about 16, 50 nm bands. This is a important region as it can assist in both mineral and vegetation identification and un-mixing at other wavelengths. The highest number of bands operating over the three regions at present is 220 provided by the AVIRIS (Porter and Enmark 1987).

There are no operational mineral-mapping spectrometers yet operating in the mid-infrared (this Division is currently testing a 100 channel, CO_2 laser spectrometer for this region) and it is too early to suggest specifications for this region until more laboratory studies have been completed (Salisbury et al., 1987) and until we better understand the effects of surface coatings on mineral detection. Current work is limited to interesting results coming from the six band TIMS scanner from youthful, well exposed terrains.

The critical point, however, is not just the number and position of bands but their radiometric sensitivity. Because of the very high interband correlation that results from many closely spaced bands, and the processing steps involved in transforming radiance to reflectance-like spectra, there must be no compromise with respect to sensitivity and precision. Airborne mineral mapping in the early 1980's with the GER Inc. profiling spectrometer suggested that NEΔρ of 0.1 % were necessary and achievable, particularly in mixed pixel and weathered environments. There is little value in increasing the current number of bands provided unless there is a concomitant increase in sensitivity, and indeed some would suggest we already have too many bands for the sensitivities currently available.

Another point about the treatment of these data is that although we have and need sensors with very large numbers of bands and hence an apparent data processing headache, our methods of analysis suggest that we do not need to process or keep all the data and that substantial reductions can be made by physically-based, feature extraction algorithms.

A case also exists for digital stereoscopic mapping to be incorporated into future mineral mapping systems to provide the all important landform/ structural context of the mineral determinations.

Having said all this we are still not sure what the natural limitations of spectral contrasts are in the natural environment. Clearly airborne spectra measured from other than perfectly exposed terrains suggest that airborne spectra are broader and noiser than their laboratory counterparts. The noise is a function of instrument sensitivity and can be improved with appropriate effort. The simplicity and broadness of airborne spectra probably reflects the effects of surface material mixing. We do not think that we have yet reached any limit imposed by the natural terrain but many more studies need to be carried out in less perfectly exposed terrains so that we can better understand the effects of mixing and the realistic specifications that are worth seeking.

Todays tools of identification are a few multi-band scanners and profiling and imaging spectrometers with a large number (24-576) of very sensitive spectral bands, supported by appropriate deterministic software tools. Such instruments include the AIS, AVIRIS, GER-64, and Geoscan II sensors. At present only one of these can be said to even approach an operational mineral-specific mapping tool.

EXAMPLES OF MINERAL IDENTIFICATION

Since about 1980 an increasing number of successful cases of remote mineral mapping have been conducted by the US and Australian research communities. At the same time the methods required to achieve success are slowly becoming better understood and more accepted, albeit still primarily amongst the research community.

In 1982 and 1983 KcKeon and Marsh and Marsh and McKeon respectively reported on a survey flown in December 1980 with the GER 576-channel profiling spectrometer that provided identification of alunite, illite (mica) and chlorite mineralogy along airborne traverses at Oatman in Arizona. Podwysocki et al. (1982) reported on a similar survey in which they were able to identify montmorillonite, kaolinite and alunite in an argillic alteration system at Marysvale, Utah. In 1984, Rowan et al. flew the same instrument over the rare-earth bearing Mountain Pass Carbonatite in California and reported picking up the very narrow (2-3 nm wide) but characteristic absorptions of neodymium over the mine and some undisturbed areas of carbonatite.

In 1983, using the same GER instrument and a 20 m pixel, we demonstrated the identification of carbonates, hornblende, epidote, and Ca and Fe-bearing smectites at Mary Kathleen in Queensland (Huntington, 1984), talc carbonates and chlorites at Kambalda in ultramafic terrains in Western Australia (Gabell, 1986), sericitic alteration in a porphyry system in Queensland (Gabell, 1986) and sericitic siltstones and epidotised basalts at Mt Isa, also in Queensland (Horsfall, pers.comm.).

Whilst these were important research results, in recent years this one-dimensional mineral mapping has been expanded with two-dimensional imaging systems such as the MEIS-II, AIS, AVIRIS, GER-64 and Geoscan Mark II capable of imaging up to 512 or more pixels of image width. These imaging systems have truly brought about the possibility of wide-area mapping of minerals.

Using data from these new spectrometers and new analysis methods one can pursue a hierarchical strategy that first "identifies" major felsic, mafic and ultramafic lithological groups based on "generic" mineralogical constituents such as: Fe_2O_3, Al-OH, Fe-OH, Mg-OH, CO_3, and various silicates groups, and then, where possible, seeks "identification" of particular mineral species. This strategy has already yielded new, previously un-mapped, geological and exploration information ready for integration with other geological knowledge.

In 1984, using the Canadian MEIS-II scanner fitted with 40 nm wide, custom-built filters centred over the crystal field absorption of iron oxides

between 747 and 986 nm (Fig. 1), we were able to produce absorption-wavelength and absorption-width images that specifically identified and mapped the distribution of haematite and goethite over gossans, laterites and iron formations in various parts of Australia (Fraser et al., 1985).

Flying the GER Inc 64-band scanner (Fig. 4) in Australia in 1987 illustrated that broad-scale lithological identification could be achieved in sedimentary and greenstone belts. For example, rock-types such as sandstones, basalts, felsic volcanics, peridotites, granites and talc-carbonates could be assigned to their correct felsic, mafic, ultramafic class from their generic mineral absorption characteristics. Further specific rock-forming and alteration minerals were also identified using the specific shapes of the absorption features mapped by this scanner. In particular a previously unrecorded zone of pyrophyllite alteration in a sericitic sandstone was identified, field checked and confirmed.

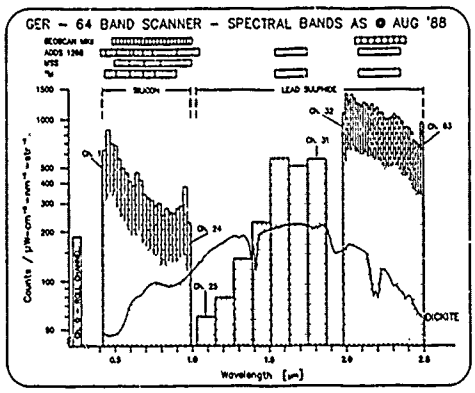


Figure 4. GER 64-band scanner spectral band configuration as at August 1988 with superimposed dickite spectrum for wavelength comparison.

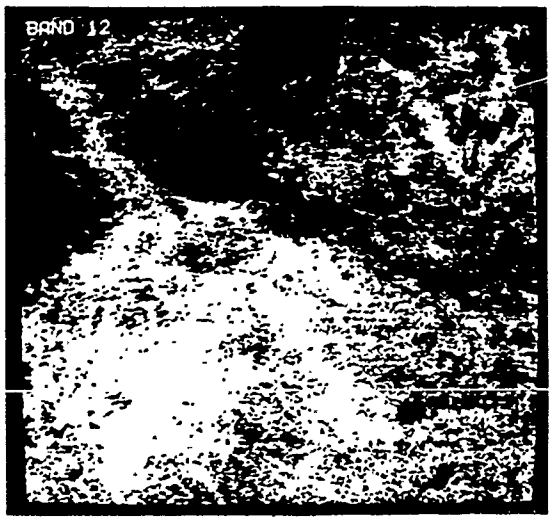


Figure 5. GER SWIR Band 12 (2.166 µm) log residual image from Oatman Arizona. Black responses indicate terrain with 2.166 absorptions due to the presence of alunite alteration.

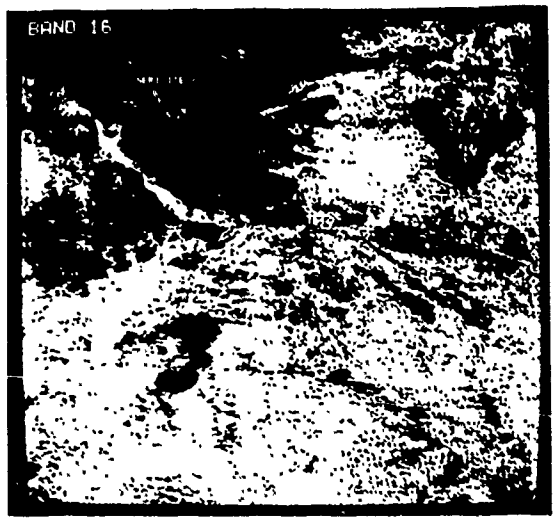


Figure 6. GER SWIR band 16 (2.22 µm) log residual image indicating absorptions due to Al-OH clays, including sericite, smectite and kaolinite. Absorption in the wide 2.166 alunite band also influences this band. Specific mineral identification uses other band responses not shown on this figure.

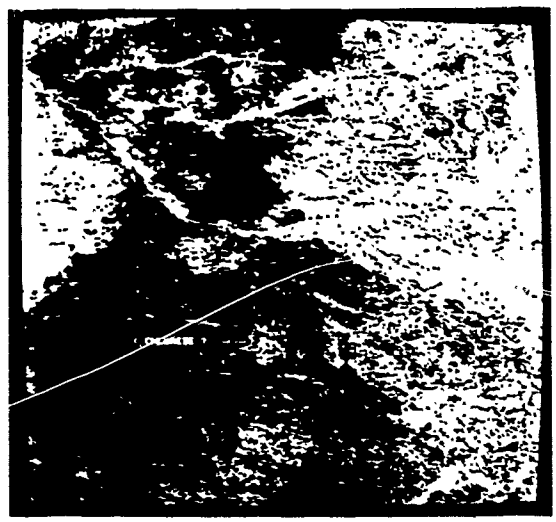


Figure 7. GER SWIR band 25 (2.35 µm) log residual image indicating in black areas of Mg-OH absorption coupled with some secondary absorptions from alunite and sericite.

At Oatman in Arizona full two-dimensional alteration mineral identification has also been achieved with this same 64-band scanner (Huntington et al., 1988). While we cannot adequately illustrate all the results in this paper, Figs. 5, 6 and 7 show just three mineral specific bands from the 32 bands available in the 2.0 to 2.5 µm region. These figures map the intensity of absorption at 2.166 µm indicative of alunite (Fig.5), 2.22 µm indicative of sericite (Fig.6) and 2.35 µm indicative of chlorite (Fig.7). Note that these identifications are not made from these single bands but from the wavelength, depth, width and shape characteristics of each spectrum for each pixel. Examples of some type mineral spectra extracted from the 32 band Oatman data set are shown

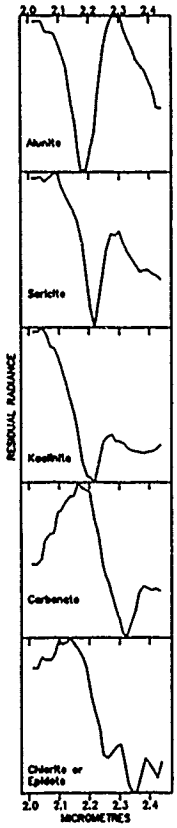


Figure 8. Type mineral spectra extracted from GER 32 band SWIR image data from Oatman, Arizona.

in Fig.8 and Fig.9 shows Al-OH spectra from 21 adjacent pixels from a sericitic/smectitic tuff horizon.

Armed, in advance, with such data, we believe the exploration geologist should be able to function with greater reliability and efficiency. Where primary mineralogy is obscured or replaced by secondary weathering products (as in some parts of Australia), the task is more difficult but research is showing that here also mineral mapping can make a contribution to geoscientific knowledge, for example evidence that some specific weathering minerals, say Fe-Mg bearing clays, can tell us about the original primary mineralogy.

Given adequate spectral and radiometric resolution the steps involved in identification from these types of high resolution data include "calibration" to reflectance-like spectra, feature coding of patterns of absorption wavelengths, depths and widths for each image pixel (Table 2) and some form of classification of these features against known spectral features and feature mixtures. Finally these identifications must be viewed in the landscape context in which they occur since the processing steps involved in identification so often remove or suppress the spatial and geomorphological information.

The significance of adequate radiometric sensitivity and precision is amply illustrated by these data. After the necessary processes of de-correlation and reflectance transformation have been applied to the 12-bit raw data, the residual spectra fit almost exactly into a 256, 8-bit range, suitable for direct image display.

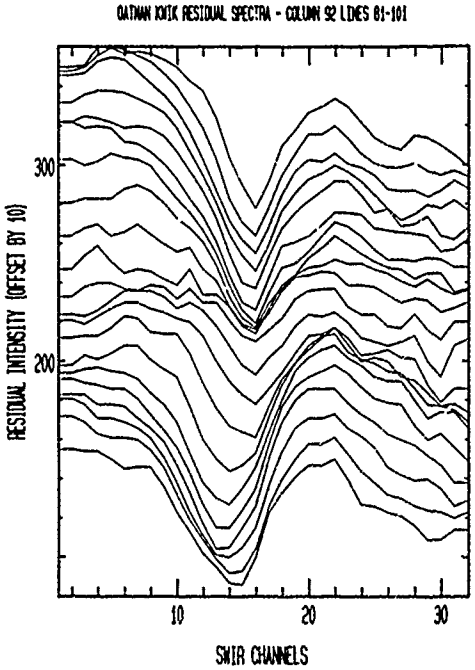


Figure 9. Twenty-one SWIR residual spectra from adjacent image pixels crossing an alunite/sericite mineral boundary from an argillic alteration zone at Oatman, Arizona. The upper 11 spectra are primarily sericitic, the lower 10 with minima at a short wavelength are of alunite and alunite/sericite mixtures. Channel 12 approximates 2.166 μm, 16 approximates 2.22 μm, 25 approximates 2.35 μm.

TABLE 2. EXAMPLE OF SPECTRAL FEATURE EXTRACTION FOR THE JAROSITE SPECTRUM SHOWN IN FIGURE 2

FEATURE WAVELENGTH (nm)	RELATIVE DEPTH	QUOTIENT DEPTH	FEATURE DEPTH	WIDTH IN CHANNELS
433	0.784	0.373	0.373	43.3
845	0.834	0.396	0.161	64.8
885	0.957	0.455	0.032	26.3
932	1.000	0.475	0.475	218.5
991	0.937	0.445	0.054	37.3
1474	0.399	0.190	0.167	20.8
1842	0.424	0.202	0.199	27.8
2226	0.718	0.341	0.341	56.1
2409	0.294	0.139	0.139	38.7
2460	0.234	0.111	0.066	18.9

CONCLUSIONS AND RECOMMENDATIONS

In conclusion, therefore, it has been demonstrated that the following specific minerals have been "identified" and mapped (along with their host lithologies) from various airborne spectrometers even, in some cases, in the presence of up to 40 % vegetation cover:

Kaolinite	Chlorite
Sericite	Pyrophyllite
Alunite	Jarosite
Buddingtonite	Haematite
Goethite	Talc

Various smectites	Epidote
Amphiboles	Fe/Mg clays after garnet
Quartz	Carbonates
REE-containing minerals	Illite

Clearly, apart from helping to make more informative geological maps that portray lithological constituents much better and more uniformly across the country (rather than just named formations that tell one nothing about their composition), this approach to remote sensing allows us to target certain specific assemblages of (alteration) minerals associated with mineral and petroleum deposits as a focus for exploration and can help us to understand the weathering of primary minerals.

Despite one and a half to two decades of mineral reflectance knowledge and nine years of mineral mapping demonstrations it is extremely lamentable that so little is being done in this field, that so few operational systems exist and that so many practising geologists, companies and governments are so unaware of the potential advantages and cost savings that could accrue from an increased and optimal use of and research towards physically-based remote sensing.

The future of these developments, at least for the next five years, needs to be focused on operational, airborne tools for exploration and mapping, built for and operated by private enterprise who can and should be making immediate use of the results produced. Much of the onus is on private industry to "come up to speed" and to be defining what they want of remote sensing, rather than leaving it to government agencies who, understandably, are working to different agendas. Mineral mapping is possible but requires a stronger commitment and better collaboration, than currently exists between the technological investment industry, the geological user industry and the research community.

REFERENCES

Ali, K., Horsfall, C.L. and Lister, R. (1989) Towards knowledge- based identification of mineral mixtures from reflectance spectra. Knowledge-Based Syst., 2, 5-13.

Fraser, S.J., Huntington, J.F, Green, A.A., Stacey, M.R. and Roberts, G.P. (1985) Discrimination of iron oxides and vegetation anomalies with the MEIS narrow-band imaging system. Proc. 4th ERIM Thematic Conference on Remote Sensing for Mineral Exploration, April 1-4, 1985, San Fransisco, pp.233-253.

Gabell, A.R. (1986) High resolution remote sensing applied to mineral exploration in Australia. Unpublished PhD thesis, University of Adelaide.

Green,A.A. and Craig,M.D. (1985) Analysis of aircraft spectrometer data with logarithmic residuals. Proc. AIS Data Analysis Workshop, Vane,G. and Goetz,A. (eds), April 8-10, 1985, JPL Publication 85-41, pp.111-119.

Green,A.A. and Huntington,J.F. (1989) Remote sensing for surface mineralogy. Proc. Exploration 87, 3rd. Decennial International Conference on Geophysical and Geochemical Exploration for Minerals and Groundwater, Toronto, Canada, September 27-October 1, 1987.

Horsfall, C.L., Lister,R., Buda,R., Green,A.A. and Craig,M.D. (1987) Expert system software for computer identification of minerals from their short-wave infrared spectra. Proc. 4th Australasian Remote Sensing Conference, Adelaide, September 14-18, 1987, pp.642-654

Hunt, G.R. and Salisbury,J.W. (1970-1976) Visible and near-infrared spectra of minerals and rocks. Modern Geology, Volumes 1-5.

Huntington , J.F. (1984) Interpretation of visible, near and short wavelength infrared, field and airborne spectra from Mary Kathleen, Queensland, Australia. Proc. 3rd Thematic Conference on Exploration Geology, Colorado Springs, April 16-19, 1984, p.227.

Huntington, J.F., Green, A.A. and Craig, M.D. (1988) Mineral and lithological mapping with the new GER 64-band scanner, U.S. and Australian examples. Summaries of the 6th Thematic Conference on Remote Sensing for Mineral Exploration, Houston, Texas, May 16- 19, 1988, pp.39-40.

Lister,R., Ali,K., Buda,R., Horsfall,C.L., and Buntine,W. (1987) GOLD: an expert system for mineral identification from reflectance spectra. Australian Joint Artificial Intelligence Conference, Sydney, pp.29-43. (Also in Gero,J.S and Stanton,R. (eds), Artifical Intelligence Developments and Applications, North-Holland, Netherlands, 1988, pp.33-48).

Marsh, S.E. and McKeon, J.B. (1983) Integrated analysis of high- resolution field and airborne spectrometer sata for alteration mapping. Econ. Geol., 78, 5, 618-632.

McKeon,J.B. and Marsh,S.E. (1982) Integrated analysis of high resolution field and airborne spectrometer data from an epithermal gold district. Proc. 2nd Thematic Conference on Remote Sensing for Mineral Exploration, Fort Worth, Texas, December 6-10, pp.151-172.

Podwysocki,M.H., Segal,D.B. and Collins,W.E. (1982) Mapping the distribution and mineralogy of hydrothermally altered rocks by using airborne multispectral scanner and spectrometer data, Marysvale, Utah. Proc. 2nd Thematic Conference on Remote Sensing for Mineral Exploration, Fort Worth, Texas, December 6-10, pp.173-174.

Porter,W.M. and Enmark,H.T. (1987) A system overview of the airborne visible/infrared imaging spectrometer (AVIRIS). Airborne Visible/Infrared Imaging Spectrometer (AVIRIS): A description of the sensor, ground data processing facility, laboratory calibration and first results. Vane,G. (ed.) JPL Publication 87-38, pp.3-12

Roberts,D.A., Yamaguchi,Y. and Lyon,R.J.P. (1985) Calibration of airborne imaging spectrometer data to percent reflectance using field spectral measurements. Proc. 19th International Symposium on Remote Sensing of Environment, Ann Arbor, Michigan, October 21-25, 1985, 10 pages.

Rowan,L.C.,Collins,W.E, Kingston,M.J. and Crowley, J.K. (1984) Spectral reflectance of the carbonatite complexes at Mountain Pass, California and Iron Knob, Colorado. Proc. 3rd Thematic Conference on Remote Sensing for Exploration Geology, Colorado Springs. April 16-19, 1984, pp.217-218.

Salisbury, J.W., Walter, L.S. and Vergo, N. (1987) Mid-infrared (2.1-25 micrometre) spectra of minerals: First edition. USGS Open File Report 87-263.

Towards Quantitative Observations from Synthetic Aperture Radar

R. Keith Raney

Canada Centre for Remote Sensing
and
RADARSAT
2464 Sheffield Road, Ottawa, Ontario, Canada

Abstract

Synthetic aperture radar (SAR), based on wavelengths of several to many centimetres, have inherent properties that both limit and extend quantitative sensor performance when compared to optical systems. Against this background, quantitative SAR remote sensing is explored at three levels: system; setting; and significance. Notable technical and quantitative results have been obtained at the first two levels. The third level, however, demands more than technical achievement. To this end, an opportunity is outlined for a Canadian initiative that would result in significant consequences derived from quantitative SAR remote sensing.

Introduction

At Arecibo, Puerto Rico, there is a famous radio and radar astronomical antenna which is one third of a kilometre in diameter. The diffraction limited angular acuity of that antenna is about the same as that of the human eye, which works best, of course, at optical wavelengths. To most people, the huge disparity between optical and microwave wavelengths, illustrated by the eye/Arecibo contrast, sets radars apart from other remote sensing systems. In partial consequence, there are those who secretly believe that the capacity for quantitative remote sensing by radar suffers in proportion to its wavelength.

The perception is particularly seductive with Synthetic Aperture Radar (SAR), for which technologists have gone to great lengths (about 16 kilometres for the SAR on Seasat, figuratively speaking!) to synthesize large apertures. A large aperture is desirable because diffraction limited resolution is proportional to wavelength and inversely proportional to aperture size, whether that of the actual antenna or emulated in computer memory as for SAR. Aperture synthesis is technology intensive, and has been perceived as the principal quantitative achievement of remote sensing SAR systems during the past three decades.

If data corresponding to the full aperture is integrated by a SAR processor for finest resolution, the resulting image is corrupted by a noise known as "speckle". Speckle is a natural consequence of all coherent imaging systems, and may be seen in laser images and holograms as well as radar data. In SAR, however, speckle seems more of a nuisance than in optical images. It is characterized by a standard deviation that is as large as its average value, and its spatial correlation matches the system resolution. When viewed, the image seems dominated by multiplicative noise. In fact, for such an image, the most likely observed reflected power at any point in the image is zero, and the Fourier transform of the image generally carries more information about the sensor than about the scene! In themselves, these are not desirable properties of an earth observation remote sensing system. In partial consequence, there are those who secretly believe that the capacity for quantitative remote sensing by SAR suffers because of speckle.

Consideration of quantitative aspects of SAR earth observation is enlightening when set against a background of the apparent wavelength and speckle disadvantages so characteristic of such systems.

Objectives

The first objective of this paper is to examine the measurement potential of a SAR system. It is ironic that the most exciting results of recent years depend on both radar wavelength and speckle to achieve notable quantitative observations unique to SAR.

The second objective is to explore aspects of the utility of SAR. Even if "quantitative" in theory and in a laboratory environment, remote sensing data must be transformed into information and put to work in a real **setting** for it to be useful.

The third objective of this paper is to extend the concept of "quantitative" to embrace consequences. This idea is considered under the heading of **significance.**

Finally, an **opportunity** is described for the remote sensing and related communities to address a substantive issue of global importance based on quantitative SAR observations.

Background

The history of SAR since it was first conceived by Carl Wiley in 1951 is shaped by several watershed events. Seasat, which operated from 4 July to 10 October, 1978, marked the beginning of a remarkable decade of increasing interest in and capability of SAR systems. The key innovation in 1978 was the advent of digital SAR processing, as opposed to optical processing which had been the standard until then. In almost all regards, optical processing, being analog, is not capable of providing the stability, control, and repetivity of digital processing upon which most results in quantitative observations by SAR depend.

Experiments based on the two Shuttle Imaging Radar flights (November, 1981 and October, 1984), and on aircraft SAR systems (JPL, CCRS, Intera, etc.) continue to extend the

technical and programmatic domain of SAR. There are two additional Shuttle radar flights planned (1992-3), and airborne SARs are increasing in number and sophistication. Three earth resource SAR satellites are planned to be launched starting in 1990.

System

Quantitative SAR performance starts with the system. A SAR system includes the usual radar hardware (transmitter, receiver, and antenna), the processor ne ded to form an image from the ensemble of signals, the interaction of the microwaves with the reflecting surface, and post-processors used for image manipulation. These elements are set against a conceptual theoretical framework, an essential part of a SAR "system" taken in the larger sense.

There are three aspects to quantitative SAR system performance: geometric; radiometric; and interferometric.

Geometric performance of a SAR conceptually is very different from that of nearly all other sensing systems. Unlike optical systems, for example, in which object position is determined by angular measures with respect to the frame of reference of the instrument, SAR image geometry depends on i) range delay, a function of the speed of light, a known quantity, and ii) Doppler shift, a function of sensor vehicle speed, another accurately known quantity. Thus, object position for a SAR is determined primarily by sensor vehicle position. Sensor angular uncertainties do not degrade the geometric error budget for a SAR. Thus, all else equal, SAR systems are capable of more accurate image positioning than are other remote imaging systems. These geometric properties of SAR are preserved and exploited best by modern digital processing methodology.

Radiometric SAR performance depends on all aspects of the "system", including scatterer reflectivity, antenna angular orientation, processor settings, etc. Amplitude calibration of SARs is challenging at best, and progress is not rapid. Absolute calibration for SIR-C (1992) is to be on the order of 3dB, which might be compared to the state of the art 15 years ago of about 4dB. Relative calibration is generally much better than 1dB, and is improving. Most interesting quantitative SAR results to date depend primarily on relative calibration.

Interferometric performance of a SAR refers to a system's ability to support operations in the complex image domain, in which image phase is of paramount importance. Until recently, most SAR R&D concentrated on the detected image, and so both sensors and processors were developed that disregarded interferometric SAR performance. However, the most exciting advances in quantitative SAR work, some of which should see economic payoff in the 90's, depend on phase preserving properties in the complex image domain. It is at this level that the importance of system concept is evident.

"Conventional" digital SAR processing, which has served the Seasat SAR decade so well, is based on impulse response magnitude criteria. In search of speed, the resulting design trade-offs led to image phase irregularities, most of which were of no consequence until recently. Now on the threshold of new quantitative techniques that require or benefit from accurate phase structure in the complex image, standard processors no longer suffice. New approaches to SAR processing are emerging with phase preserving properties. Based on wavenumber domain methods introduced to the reflection seismics community by Stolt in 1978, Rocca and others first applied it to radar processing in 1987. Using improvements unique to SAR, CCRS and MDA are developing a new class of phase preserving wavenumber domain SAR processors. From the point of view of quantitative complex image exploitation, we are entering a new decade.

There are several areas of SAR system level work of importance that depend on complex image phase. These include calibration, time-base interferometry, elevation interferometry, and polarimetric synthesis. In the case of polarimetry, the relative (complex) calibration between channels is critical, with notable recent results being reported at IGARSS.

Polarimetry, at the system level, deserves special mention, largely because a configuration is required which goes beyond the single wavelength and polarization implementation now planned for ERS-1, J-ERS-1, and RADARSAT. The first polarimetric work in the SAR context was by Zebker, Van Zyl, and colleagues at the Jet Propulsion Laboratory. The new JPL aircraft SAR is designed around polarimetry, and it is a key feature of SIR-C. In polarimetric synthesis, images that correspond to all possible combinations of transmit and receive polarizations are available from one polarimetric data set. Typically, four data paths are required (for each frequency), hence the term "quadrature polarimetry". Each data stream corresponds to a unique pair of transmit and receive polarizations, drawn from the horizontally or vertically polarized antenna used for transmission, and for reception, respectively. The SAR data is gathered in time multiplex, processed to the complex image stage, and, with the matrix operator known for the desired polarization, manipulated to synthesize the corresponding image.

At the user end, work station techniques and standards are being developed at many centres world wide. Many are designed exclusively for polarimetric investigations. JPL should be recognized and complimented for their leadership in this area, particularly for making their data and software available for polarimetry synthesis work by other investigators. The entire remote sensing community benefits from such unselfish thinking.

Setting

SAR systems are becoming more elegant in concept and performance, but quantitative *results* are obtained only through observations of a real world setting. There have been notable recent accomplishments, to which the several hundred SAR papers at this level for IGARSS'89 attests. Indeed, with the potential of complex imagery being more deeply explored, new quantitative applications are rapidly developing. Many of these should result in tangible economic benefits during the 90s.

The traditional approach to SAR image interpretation has been to look for reflectivity signature variations as a function of SAR system parameters, and as a function of scene characteristics. Important results have been obtained in most application areas, including ice, oceans, agriculture, forestry, and terrain mapping.

The setting is ocean wave observation by SAR. Both from spacecraft and from aircraft, wave spectral estimation by SAR has important practical potential benefits, but the problem is compounded by the motions of the sea which complicate and sometimes limit the ability of a SAR to determine the directional wave spectrum. Recent progress has been made, with the result that data from airborne SARs may be processed in innovative ways to more accurately estimate wave spectra, and to unambiguously determine the wave propagation direction. It is expected that progress on wave height estimation will follow. The work has impact on current and new radars for both aircraft and spacecraft.

The setting is forest mapping. Within the restraint of single frequency, single polarization SAR, an important system variable that may be exploited in applications is incidence angle. Using SIR-B data at several different incidence angles, Cimino and colleagues have shown that different forest species

and different forest structures for a single species may be differentiated. Since RADARSAT will be a system employing a wide variety of incidence angles, the approach will be of use as a tool of potential operational significance in the 1990s.

Two new quantitative methods based on appropriate operations on sets of complex SAR imagery deserve special notice: interferometry, and polarimetry.

The setting is terrain height estimation. Goldstein and his colleagues at JPL have proposed and demonstrated an innovative approach to this application. By linearly combining two complex images of the terrain, each taken from different but nearby spatial positions, interference fringes are formed which can be transformed into an elevation contour map. The two images need not be simultaneously obtained. Data from a single satellite SAR but from two similar but different orbits may be used. The method works precisely because the wavelength of SAR is well suited to the purpose, and because the complex reflected phase is preserved in each image. If a real image were to be formed, the seemingly random phase in each image would become "speckle", but for correlated pairs of complex images, the phase *difference* expressed in the interference fringes carries radar range information accurate to a fraction of a wavelength. Knowing the geometry of the image pair, this is sufficient to find the elevation of each scatterer. The resulting elevation contour scale depends on SAR and imaging geometry, but elevation sensitivity on the order of tens of meters is attainable.

The setting is signature analysis, and hence, scatterer classification. First the JPL group, and now many investigators, are exploiting the benefits of quadrature polarization data sets to address this class of questions. When all combinations of polarization are available, the data may be scanned to derive a "polarization signature", and to compare it to theoretical signatures that would result from selected scattering models. In forestry, for example, this approach allows reflections from the terrain, from the tree trunks, and from the canopy to be separately observable. In more general terms, several classes of scattering simultaneously occurring in each resolution cell may be quantitatively differentiated. Since radar reflectivity is dominated by the relative size and orientation of the incident wavelength and scattering elements, and since the polarization synthesis is phase sensitive, the quantitative benefits of the technique is a direct positive consequence of radar wavelength and speckle.

There are many other specific cases that could be cited. The examples above have been chosen to illustrate specific technical capabilities of SAR systems, and the rapid rate of progress in quantitative SAR observations in practical settings. There remains one more level to be explored.

Significance

Quantitative observations by SAR in practical settings are necessary for significant results, but are not sufficient, at least as viewed from a global and economic perspective. *In order for SAR observations to be significant, they must amount to something.* Individual research, no matter how inspired, falls short of this test. Demonstration projects, by themselves, would be encouraging, but are not "significant" in the larger sense. For SAR to be a significant quantitative and economic tool for the 90s, it needs to see substantive and large scale use.

There are examples of significant SAR employment, but most are military. In civilian remote sensing, perhaps the best examples are found in airborne radar mapping applications. The STAR-1 (Intera Ltd of Calgary) has flown millions of miles in mapping missions throughout the world. Its performance is certainly quantitative, taking advantage of the geometric stabilities inherent to SAR imaging and post-mission rectification to provide excellent products. Clients with interest in specific settings from the tropics to the Arctic have benefitted. The results have been significant.

The work of CCRS, JPL, ERIM, and other institutions is an essential foundation and background for development of SARs such as the STAR-1. New systems, new concepts, and new applications come from these important centres that play such a key role in enhancing quantitative remote sensing SAR performance. Likewise, the role of industry is acknowledged, as in the development by MDA of the first digital SAR processor for Seasat. However, moving beyond technical excellence (system) and demonstrated effectiveness (setting) towards global consequences (significance) is daunting, particularly for technical institutions considered individually.

Late in 1990 or early 1991 the first SAR bearing remote sensing satellite since Seasat will be launched. The European Space Agency's ERS-1 will carry a radar whose performance will be approximately similar to that of Seasat, except at C-Band instead of L-Band. The Japanese will launch the SAR satellite J-ERS-1 in 1992, and Canada should see RADARSAT launched in 1994, a C-Band multi-mode SAR having a five year design life. Other polar orbiting SARs including polarimetric systems are under discussion for 1997 and beyond. For remote sensing, the decade of the 90s will be the decade of SAR satellites capable of quantitative performance. But from a global perspective, as economic tools, will they be significant?

Opportunity

For largely domestic purposes, Canada plans to use SAR data from RADARSAT together with data from other sources in an on-line real-time information service to the navigation and ice community. The systems are being designed now. Remote sensing and environmental scientists are actively engaged in studying C-Band scattering properties together with detailed parameterization and modelling of the pertinent environments through major field projects such as LIMEX. Institutional arrangements are being created for data flow, and data exchange protocols are being negotiated. Data recipients are participating in data use and analysis based on aircraft SAR programs, both experimental and operational. In short, there is a long range, inter-disciplinary, and inter-institutional program designed to support future significant RADARSAT SAR data utilization for Canadian ice related services. This is a small example of what could be done.

One notable opportunity open to the remote sensing community at large is tropical deforestation. As well as being significant and deserving urgent attention, it is one of the global problems to which radar remote sensing is directly applicable. Tropical forest inventory and its annual change are only approximately known. Since radar wavelengths penetrate fog and cloud cover, satellite based imaging radar is the ideal tool to gather the needed information.

These views generally are accepted in principal, but most of us think it is someone else's problem. There is at this moment no global initiative planned to come to grips with the issue.

This is the opportunity for Canada, and for the remote sensing community represented here at IGARSS '89. An initiative could be taken, starting from basic scientific measurements, to culminate in an operational international tropical forest inventory and monitoring system ten years from now. The plan would have to be long range, inter-disciplinary, and inter-institutional as in the example of Canada's ice information

program. In addition, it must be international, and should include as an active partner one or more countries with tropical forests. The primary data base would come from the sequence of SAR satellites now foreseen. Such a project would need to begin soon to build the required scientific and institutional foundations.

Canada is uniquely positioned to motivate such a project. Since RADARSAT is planned to start operations in 1994, Canada could make an offer now to provide, *gratis*, and annually, a base set of tropical forest inventory data from RADARSAT for the life of the mission. The offer, if made to the United Nations, for example, could serve to encourage an international mobilization in response to the issue. I believe that the remote sensing community, and concerned individuals in all nations, would welcome and build on this initiative, and would willingly participate in the efforts needed to achieve significant consequences from earth observation by SAR.

Closing

There have been notable achievements in recent years in quantitative SAR observations. With the new radars and processors now being developed, the decade of the 90s will see substantive benefits in specific application areas. However, for global issues for which SAR is an appropriate remote sensing tool, significant and economic consequences will follow only if long range and large scale preparations are made. It will not happen by itself.

The author is indebted to many colleagues whose work, directly and indirectly, make the substance and the optimism of these remarks possible. It is not practical to list all pertinent references, but the Digest for IGARSS'89 serves as a good starting point.

SIMPLE MODELS FOR COMPLEX NATURAL SURFACES: A STRATEGY FOR THE HYPERSPECTRAL ERA OF REMOTE SENSING

John B. Adams, Milton O. Smith and Alan R. Gillespie

Department of Geological Sciences, University of Washington, Seattle, WA 98195

Abstract

We describe a two-step strategy for analyzing multispectral images. In the first step, the analyst decomposes the signal from each pixel (as expressed by the radiance or reflectance values in each channel) into each of the individual components that are contributed by spectrally distinct materials on the ground, and those that are due to atmospheric and instrumental effects and to other factors such as illumination. In the second step, the isolated signals from the materials on the ground are selectively edited, and recombined to form various kinds of unit maps that are interpretable within the framework of field units.

Keywords. multispectral images, hyperspectral systems, mixture models

Introduction

New developments in technology have brought us to the threshold of the hyperspectral era in which it is possible to measure the same target in hundreds of narrow wavelength intervals (e.g., Goetz et al., 1985). The remote-sensing community, however, is not yet prepared to process and interpret hyperspectral data, and then to integrate the measurements from multiple instruments on a platform such as the NASA Earth Observing System (EOS). Even if we assume that future advances in computer technology will make it possible to retrieve. archive and process the large amounts of data that will flow from future remote-sensing systems, there is the equally challenging problem of how to validate and interpret the data.

The purpose of this paper is to outline a new strategy for the remote sensing of natural surfaces using multispectral imaging systems aboard satellites and aircraft. The approach has been tested on multispectral images of a variety of natural land surfaces ranging from hyperarid deserts to tropical rain forests. Data were analyzed from Landsat MSS and TM, the airborne NS001 TM simulator, Viking Lander and Orbiter, AIS and AVIRIS. Because the approach is not restricted by wavelength or by instrumentation, it is effective with the four channels of Landsat MSS and with the 224 channels of AVIRIS.

Field Mapping and Image Analysis

Those who have mapped geology, soils, or vegetation in the field are especially aware of the enormous complexity of many natural surfaces. One of the most important skills that is required for field mapping is the ability to organize and simplify the observable information, and to present a condensed and edited version of the data as a map. Such maps are a fundamental means of conveying information, and can be considered simple models of complex surfaces. It is important not to confuse this kind of mapping, to which we refer simply as "field mapping", with quantitative mapping or with mapping the positions of features in a Cartesian framework, as in topographic and cartographic work.

Field mapping is done by creating "units," which usually represent materials that have distinctive and significant properties. Map units typically have "type areas" on the ground where the distinctive nature of the materials can be referenced. Units are carefully selected as part of an overall organizational scheme to convey certain important information to the reader of the map. For example, a geologist might organize an area initially into volcanic rocks, alluvial deposits and æolian dunes. With further work each of these units might be subdivided into other units, such as individual volcanic flows and alluvial fans of different ages, and dunes of different forms.

On any given map, field units are few in number, typically between three and ten; maps having hundreds of units are unintelligible. Field units always are distinctive; they can be told from other units. If one unit grades into another, the gradational nature of that boundary must be indicated on the map. Field units have characteristic properties, which may include species composition, mineral composition, age, texture, internal homogeneity or heterogeneity. They always are defined with a particular purpose in mind, that is, to convey certain information and to omit certain information. For example, geologic maps commonly lump surficial deposits into a single unit, and entirely omit units of soils and vegetation. Vegetation maps typically omit geological units. Soils maps usually do not show geologic or vegetation units.

It is notoriously difficult to describe how to map units in the field, because defining units involves the judgement, experience and objectives of the mapper. Thus, each person mapping in the field answers the question: "What do you want to show?" The mapper must organize, simplify and choose. In establishing a unit in the field the mapper implies that it is meaningful within a defined geological, ecological or other frame of reference.

The image analyst, especially one using hyperspectral images, is confronted with problems similar to the ones encountered in field mapping. There is a large amount of data that must be organized to convey selected

This work was supported by NASA grant NAGW-85.

information. Because our conventional scientific frame of reference comprises field observations and laboratory measurements, the results of organizing the remote-sensing data would correspond ideally to the kinds of units that we would map on the ground. The image analyst, however, has different tools to work with than the field or laboratory worker, and, therefore, the criteria for selecting units often are different. Nonetheless, the products of the image analysis must be understandable to a mapper in the field, otherwise the conventional frame of reference is lost.

Basic Tools from the Landsat Era

Most of the methods that are used for processing and interpreting multispectral images were developed in the 1970's for application to Landsat MSS, which had four spectral channels. Excluded from consideration here are various restoration, filtering and enhancement techniques that apply to individual channels. Instead we focus on the methods for displaying and interpreting the channel-to-channel variations in the images. The basic tools include three-color enhancements, channel-to-channel ratioing, ratio color compositing, principal-components analysis (PCA) and assorted classification schemes. These methods still are used widely for Landsat, SPOT and other systems having only a few channels.

The basic tools can be applied to images having hundreds of channels; however, only a small fraction of the data can be treated at one time. For example, only three channels can be displayed on a color monitor. Ratios are limited to two channels, although by using color composites three ratios can be displayed at a time. When all of the channels cannot be examined or displayed simultaneously then choices must be made as to which to use and which to ignore.

PCA is a computationally straightforward way to describe the variance in a four-channel Landsat image (four-dimensional space), and to rotate the data according to the eigenvectors. With hundreds of channels PCA is computationally expensive, and the results may be difficult to interpret. There is also the disadvantage that, regardless of the dimensionality of the image data, the orthogonal axes of statistical variance do not have consistent and simple equivalence to field units. By examining the data structure of many multispectral images in an interactive computer display (Possolo et al., 1989) we have found that the most interesting axes of variance in terms of image interpretation commonly are separated by small oblique angles.

Classification methods (e.g., parallelepiped, maximum-likelihood, etc.) are among the most commonly used basic tools. Clusters of points in multidimensional space are defined, and corresponding pixels on images are "alarmed" or "painted." Classifications may be guided by known spectral properties, or they may be defined statistically. There typically will be fewer classes for any given scene when it is viewed by a camera system having three or four channels than when it is viewed by a hyperspectral system, because more spectral variation can be detected by using more channels. However, one can generate a large number of classes even with the simplest multispectral system. With a hyperspectral system the potential number of classes is greatly increased; however, simply increasing the number of classes is not our objective.

Our objective is to be able to classify images within the conventional frame of reference of field and laboratory observations, and to choose from among these classes certain ones for presentation in maps, while editing others that, for example, only show variations in lighting, or different proportions of the same materials.

A New Strategy

Hyperspectral measurements can bring the image analyst closer to describing the nature of the materials on the ground; however, there still is the need for a way to separate and edit portions of the data. We may choose to ignore the soil and the rocks if we are mapping vegetation in the field; but a multispectral image of the same area records the entire signal. A choice to ignore part of the signal only can be made at a later time by the image analyst. If the pixel size in the image is small enough so that each pixel typically encompasses only vegetation or only soil/rock, the analyst could selectively display the vegetation pixels, and ignore the soil/rock pixels (and *vice versa*). A more realistic example, however, is the one for which the spatial resolution is at a scale of meters to tens of meters, and the pixels encompass areas of both vegetation and soil/rock. Now the analyst needs to separate from each pixel that part of the signal that expresses the vegetation or the soil/rock.

The presence of shading and shadows, which may be related to the vegetation or soil/rock roughness elements, topography, solar elevation and other factors, will further complicate the comparison of pixel signatures with field units. The field mapper will not define a "shade/shadow" unit because illumination variations are not a tangible part of the sene. A multispectral image, of course, will record all of the variations in illumination, and the analyst needs to be able to separate that part of the signal that is contributed by tangible materials.

We propose a two-step strategy for analyzing multispectral images. In the first step, the analyst decomposes the signal from each pixel (as expressed by the radiance or reflectance values in each channel) into each of the components that are contributed by spectrally distinct materials on the ground, and those that are due to atmospheric and instrumental effects and to other factors such as illumination. In the second step, the isolated signals from the materials on the ground are selectively edited and recombined to form various kinds of unit maps that are interpretable within the framework of field units.

New Tools for Image Processing: Image Endmembers

Consider a multispectral image of a hypothetical scene that includes vegetation and soil. We will use this simple example throughout the paper for illustration; however, the methods discussed have been demonstrated on many natural scenes of considerable complexity, and are not restricted to vegetation and soil (Smith and Adams, 1985; Adams et al., 1986; Ustin et al., 1986; Smith et al.,1989; Possolo et al., 1989). The objective is to isolate the spectral contributions of vegetation, soil and other sources in each pixel. Our approach is to define a set of spectra from the image, called "image endmembers," which when mixed using an appropriate rule, reproduce all of the pixel spectra.

Rarely are natural surfaces uniformly illuminated; therefore, to account for natural variations in illumination, we select the spectra for the vegetation and soil image endmembers from areas that receive maximum illumination. We then introduce an image endmember that we call "shade." As an endmember, shade can mix in all proportions with each of the other endmembers or with their mixtures, thereby modeling the spectrum of the endmember material when it is not fully illuminated. Thus the endmember "shade" accounts for variations in lighting at all scales due to changes in incidence angle, and variations at all scales that are caused by shadows, including multipixel shadows cast by topographic features and subpixel shadows cast by trees, bushes or other objects having

roughness and texture.

Each of the spectral endmembers must be relatively free of contamination by other materials or other effects. It is unrealistic to expect to find a pixel that has absolutely no effects of shading and shadows at all scales; nevertheless we might find one in which these effects are minimal. Similarly, in our example we seek the spectra of relatively pure areas of vegetation and of soil. An image endmember having, for example, 70% vegetation and 30% soil would not be satisfactory, whereas one having only 3% soil might be, depending on the objectives of the analysis. The problem is simplified if we have knowledge from field observations that the spectra of the endmember pixels are representative of the materials that we want to isolate. The shade endmember typically is derived from a pixel in a shadow cast by a topographic feature. Dark areas such as a blacktop road or a body of clear water under some circumstances can be a proxy for shade.

Once we have defined the spectra of the image endmembers we can calculate the spectra of any mixtures of the endmembers by applying an appropriate mixing rule. The objective is to solve for the fractions of each of the endmembers in each of the pixels. A general equation for mixing by area is:

$$DN_c = \sum_{i=1}^{N} F_i DN_{i,c} + E_c \quad \text{and} \quad \sum_{i=1}^{N} F_i = 1 \qquad (1)$$

where DN_c (digital number) is the relative uncalibrated radiance in channel c of an image pixel; F_i is the fraction of endmember i; $DN_{i,c}$ is the relative radiance of endmember i in channel c; N is the number of endmembers; and E_c is the error for channel c of the fit of N spectral endmembers. Equation 1 converts the image DN values in each channel for each pixel to the equivalent fractions of each of the endmembers. It also requires that the fractions of the endmembers sum to one. For the specific example used above, $N = 2$: F_1 and $DN_{1,c}$ might refer to vegetation, and F_2 and $DN_{2,c}$ might refer to soil.

We now have constructed a model consisting of three image endmember spectra which are allowed to mix with each other to form other spectra. We wish to determine how well this model (the ensemble of spectra plus Equation 1) accounts for the actual spectral variation in the acquired data. If the model does not fit the data well our strategy provides us with ways to find out whether we have missed some important part of the signal (another endmember), whether one or more of the original endmembers is inappropriate, or whether part of the spectral variation in the image is caused by atmospheric, instrumental or other effects that are extrinsic to the surface.

The model is tested in three ways. First, we compute the root-mean-squared (rms) error in the fit of the model to the image data. This parameter is based on the E_c term of Equation 1, squared and summed over all M image channels:

$$\varepsilon = \left[M^{-1} \sum_{c=1}^{M} E_c^2\right]^{1/2} \qquad (2)$$

The error is computed separately for each pixel. It is a measure of the spectral residue that cannot be explained by the mixing model. If ε is small, then the model may be a good one; is ε is large, the image endmembers or mixing rules may have been chosen or characterized incorrectly. Spatial patterns and variability in the rms error can be examined in an image by encoding values of ε as DN. We display the higher values of ε as lighter.

Second, we display the fractions of the image endmembers, F_i, as "fraction images," by scaling values of F_i to fit in the 8-bit DN range. Endmember fraction images permit us to examine variability in the model in a spatial context, and to relate spatial patterns in the fractions to conventionally enhanced images and to other types of information. For example, the high fractions of vegetation, and of soil, should form patterns on the image that are consistent with ground or other observations. The fraction of shade is closely related to the topography; therefore, the complement ($1-F_i$) of the shade-fraction image should mimic the topography, illuminated as during image acquisition, but devoid of spectral and most albedo information. The construction of fraction images is a powerful test of the validity of the model that exploits the interpretive skills of the analyst: if patterns of F_i in the fraction images "don't make sense" then the model may have been constructed incorrectly.

We display the "fraction image" by arbitrarily encoding F_i as DN such that $DN = 100\ (F_i+1)$; thus, $F_i = 0$ is indicated by a DN of 100, and $F_i = 1$ by a DN of 200. Because realistic values of endmember fractions will range between zero and unity, the correctly modeled data in the fraction images will lie in the range $100 \le DN \le 200$.

For the third test, we compute a parameter we term the "fraction overflow". As noted above, fractions of endmembers in the scene must range from zero to unity. However, if the model is inadequate to describe the image, the computed image-endmember fractions may fall outside this range, even though they sum to unity as required by Equation 1. In this situation, ε may nevertheless be small, indicating a good fit.

Out-of-range endmember fractions may be encountered if areas were incorrectly chosen: for example, a vegetation reference area might actually contain a small area of visible substrate, or a soil reference area might be partially vegetated. Even so, fractions will not be overrange unless a more extreme or pure pixel is encountered in the image. If an area of 50% vegetation is used as a reference area, a completely vegetated area will have $F_i = 2$. In this instance, because the fractions must sum to unity, the sum of the shade and soil endmember fractions must be -1.

In the fraction images, out-of-range fractions are encoded as values outside the range $100 \le DN \le 200$: negative fractions are indicated by $DN < 100$, and superunity fractions by $DN > 200$. We display the "fraction overflow" by by "alarming" or "color coding" pixels for which values of F_i are unrealistic. "Fraction-overflow" images thus created highlight areas and components in the scene for which the model is ambiguous or otherwise inappropriate.

If the initial model does not fit the data satisfactorily, as judged by the three tests described above, we may need to change the reference areas used to define the image endmembers, or we may need to select one or more additional endmembers. Potential new areas for defining endmembers are identified on the rms-error (ε) image as light pixels. The various new endmembers are not discriminated in the ε image, except in that they are together distinguished from well modeled pixels. Light pixels in the ε image may be all due to one unmodeled endmember, or to twenty. We may choose to ignore some pixels that have a high rms error, for example those representing a road or a lake, if we are interested primarily in the rest of the scene. If the model does fit we may proceed with analysis, keeping in mind that the endmembers that were used represent broad classes of materials, and may not yet be suitable for specific identifications.

In the example so far we have converted an image from radiance (DN) to fractions of endmembers. Two of the endmembers represent known materials on the ground, while the third, shade, was introduced to account for variable illumination and for shadows. Using equation 1 we quantified the fractions of the endmembers at the sub-pixel scale, avoiding the problem of defining large numbers of arbitrary classes consisting of simple mixtures. Now that the multispectral image data are expressed as fractions of endmembers we can select and delete information, which is necessary to arrive at units that are consistent with the framework of field mapping.

New Tools for Image-Processing. Displays of Fraction Images

Endmember fractions can be displayed in a variety of ways. Individual fraction images can be color coded by fraction steps, or three endmembers can be displayed by assigning each to one of the red, green or blue (RGB) color guns. Significantly, combinations of some fraction images can be displayed while leaving out others. For example, in the hypothetical image we can delete the fraction of shade by rescaling all fractions so that vegetation and soil sum to unity. (When we refer to vegetation cover on the ground we do not count the fraction of shade). Thus, a pixel having 0.33 vegetation, 0.33 soil and 0.33 shade would convert to 05 vegetation and 0.5 soil. An image of the rescaled fractions can be used to display quantitatively the amount of vegetation relative to soil by color coding selected fraction-intervals.

If there are two or three soil endmembers, we can rescale their fractions to exclude vegetation and shade, thereby producing a fraction image that shows the relative proportions of the soils alone. This also is a way of removing the vegetation contribution to the image, and has been dubbed "spectral defoliation" (A. Green, personal communication). In a similar way we can exclude soils and shade, and isolate two or more vegetation endmembers.

A particularly useful display for classifying vegetation is a three-color fraction image where R = (1 - F_{shade}), G = F_{veg}, and B = F_{soil}. Forested areas (or any green vegetation having a high fraction of shade) are green, and low trees, shrubs and grasses (or any green vegetation having a lower fraction of shade) are yellow (a mixture of green and red, where red is high in the absence of shade).

Fraction images are a powerful tool for isolating parts of the image data and for quantifying the amounts of the materials that are present. This is the same kind of selective organization that characterizes field mapping. With this approach the image interpreter does not have to convert mentally between radiance and materials on the ground. Classification is done in terms of the endmembers, and not in terms of statistical groupings. Furthermore, the fraction images can be based on all of the spectral channels. With more channels, more endmembers can be defined (one more endmember than the number of channels), and the uniqueness of the spectral identity of each endmember may be enhanced. Thus, the approach lends itself to hyperspectral data as well as to simpler systems such as Landsat.

New Tools for Image-Processing: Reference Endmembers

In the above examples we obtained endmembers from pixels in the image itself. If there are no areas in the image that consist of the desired materials, for example in a semiarid scrub where every pixel has vegetation, soil and shadow, the endmembers must be derived from laboratory or field reference spectra We term these endmembers "reference endmembers." To use reference endmembers we need to convert the image data to reflectance. Atmospheric and instrumental calibration coefficients rarely are available to correct an image to within a few percent reflectance; therefore we must use a part of the scene as a reference. Traditionally this is done by finding pixels that are entirely occupied by a known material. However, the same calibration can be done if known materials are mixed, by applying an appropriate mixing algorithm. Thus, if we know the laboratory or field reflectance spectra of representative materials on the ground, we can calculate the expected reflectances of all pixels that contain mixtures of these materials. The discrepancies between the calculated reflectances expressed as DN and the measured pixel radiances in DN are caused by a combination of factors that include the spectral solar irradiance, atmospheric attenuation and scattering, instrumental gains and offsets, and lighting artifacts that have "corrupted" the reflectances of the materials on the surface.

If we do not know the spectra of the materials on the ground then we can test for the combination of the best endmembers derived from laboratory or field spectra and the most plausible gains and offsets for each channel that are needed to make the endmembers fit. These gains and offsets combine all of the corrupting effects.

We have developed a method to solve simultaneously for endmember fractions and for the combined gains and offsets that bring the image data into "alignment" with the reference spectra. The equation is.

$$G_c DN_{i,c} + O_c = \sum_{r=1}^{N} F_r R_{r,c} + E_c \quad \text{and} \quad \sum_{r=1}^{N} F_r = 1 \qquad (3)$$

where $R_{r,c}$ is the reflectance in channel c for reference endmember r, and $DN_{i,c}$ is the uncalibrated radiance in channel c for image endmember I, and we are solving for the fractions, F_r, the gains, G_c, and the offsets O_c. As in equation 1 the fractions of the endmembers sum to unity. Applying the coefficients G_c and O_c to the acquired radiance data calibrates them to reflectance measurements.

Our strategy uses Equation 3 initially on a small subset of a larger image, for which it is safe to assume uniform atmospheric conditions. We select a simple area where we know the endmembers, or can narrow the probable choices of materials. Once we find the gains and offsets for each channel we apply them to the rest of the image. After an image is calibrated to reflectance it is easier to test additional models with reference spectra. Calibrating, modeling and testing are done iteratively.

Narrow Absorption Bands

One of the main reasons for developing hyperspectral systems was to be able to detect narrow absorption bands in the spectra of minerals and other materials, because such bands often are diagnostic of chemical composition and structure, and thus are useful for identification. If the absorption bands are strongly expressed they may be evident in the earliest stages of data analysis, even prior to correcting the data to reflectance or deconvolving mixtures Methods for detecting and characterizing absorption bands have been developed by several workers , but will not be reviewed here.

Our experience has been that many of the strong absorption bands that are seen in laboratory spectra are weak or undetectable when the same materials occur on natural surfaces. Absorption bands in any one material may vary significantly in strength depending on such factors as particle size and the presence of other materials that are strongly absorbing. Within a pixel-sized area a material,

which in pure form in the laboratory would show characteristic absorption bands, may have weak or indistinct bands that are of the same magnitude as the effects of atmospheric and instrumental variations Our strategy requires a further step to detect narrow absorptions that depart from the continuum by only a few percent. We achieve this by examining the fit of the endmembers to the image data in each channel.

So far we have described an approach that gives equal weighting to all channels, and is relatively insensitive to three or four channels out of hundreds. Because the reference endmembers are selected for their overall fit in all channels they may not fit exactly in any few channels. A narrow absorption band may occur in the image data, but not be present in a reference endmember that otherwise fits the overall data well. Or, an absorption band (weak or strong) may occur in the spectrum of a reference endmember, but that band is not present (or detectable) in the image data.

To examine the fit of the model channel by channel, for each pixel we estimate a model DN from the sum of the products of the reference endmembers and the model fraction, F_r. From this value is subtracted the observed DN_c, to give for each pixel a residual **N**-vector comprised from the components $\Re_c$. The residual for each channel, $\Re_c$, is found according to:

$$\Re_c = \sum_{r=1}^{N} F_r R_{r,c} - DN_c \qquad (4)$$

The residual may be encoded as DN and displayed as an image. Positive residua indicate that, in channel **c**, the endmember model specifies higher reflectances than were actually observed. Negative residua indicate that the image DN are higher than the modeled value.

We know the detailed spectra of each of the reference endmembers. If we are searching for a mineral absorption band at, say, 2.3 μm, we know whether we have included such a band in the model, and which reference endmember contains it. There are several possibilities. If we have modeled the absorption feature, and if it is actually in the scene, values of $\Re_c$ at the wavelength of the feature will be small. If the feature is absent, they will be large and and negative. On the other hand, our model might not contain the absorption feature. In that case, if it is present in the image, values of $\Re_c$ will be large and positive; if it is absent, they will be small.

If the model and the measurements do not agree, the wavelength at which the disagreement is worst may be determined by inspection of $\Re_c$ for several channels in the vicinity of the feature. For example, in our model of the hypothetical image above, there might be an absorption band at 2.3 μm in the spectrum of the soil endmember, but none in the spectrum of the vegetation or the shade. If an image of $\Re_c$ = 2.3 μm shows that all pixels have negative residua, we would conclude that for some reason the surface materials do not show the same absorption band at 2.3 μm as appears in our soil reference endmember Whether the apparent absence of the band on the ground is the result of a significant compositional difference is a question for interpretation. Similarly, if the residual image shows pixels that have positive values, we might conclude that the corresponding areas on the ground had a stronger absorption band at 2.3 μm than the reference soil. Thus we are able to learn a great deal about the fit of our model, and the spectral identity of misfit features, by inspection of images of $\Re_c$.

To improve further our confidence that the absorption band that we seek is a feature of the surface and not an atmospheric or instrumental effect, we inspect the shape of the absorption feature in the image by examining the residual images for several channels near the center of the band. A large residual in only a single channel is suggestive of instrumental noise, whereas residua that systematically increase to a maximum, and then decrease are suggestive of the characteristic pattern of a mineral absorption band: weak absorption in the wings, and increasing absorption toward the band center. To determine whether the absorption is occurring on the surface or in the atmosphere, we examine the patterns on the residual images. Absorptions on the ground typically exhibit a spatial pattern that has meaning within the context of the image. Thus, for example, we would not expect the 2.3-μm band in our hypothetical image to extend over areas of 100% vegetation as well as over areas of soil. Broad or diffuse patterns that are not spatially organized may indicate atmospheric absorptions; random patterns, line patterns, or periodic geometrical patterns are typical of noise or instrumental effects.

It happens that the human eye-brain system is extraordinarily sensitive to subtle spatial patterns, and we often can detect spatial patterns in the residual images that are caused by absorptions on the ground, even when these patterns occur in the presence of stronger superposed patterns of lines or random noise. The approach, therefore, lets us detect weak absorptions at the level of the system noise. Furthermore, by presenting the data in image form we are less likely to mistake atmospheric absorption bands or instrumental pseudobands for bands in materials on the ground.

Identifying Materials

When analyzing multispectral images having only a few channels we are accustomed to the idea of discriminating among materials on the ground. It is generally understood that with only three, four or six channels reflectance spectra are so undersampled that it is not possible to identify most materials uniquely. The advent of hyperspectral images does not mean, however, that now it is possible to identify materials routinely. We discussed above that mixtures of materials may mask the absorption bands that are needed for unique identification. Some materials, of course, cannot be identified uniquely by their spectra even under optimum laboratory conditions. For most materials, however, spectral identification and spectral detectability depend on the context, that is, on what other materials are present, and on the conditions under which the measurements are made (Shipman and Adams, 1987). In one context a material may stand out clearly, whereas in another the same material may be indistinguishable from its surroundings. To interpret multispectral data we need to be aware of the conditions under which we can expect to detect each material. Thus, identification and detectability are topics that require consideration of spectral mixtures.

If it is important to consider the spectrum of each material in context with those of the other materials that are present, it follows that an exploration strategy needs to include the broader context, rather than searching for just one material at a time. A search that is narrowly focused toward finding material **x** may yield quick results, perhaps by finding a strong, characteristic absorption band in the radiance data. If the same search yielded no diagnostic spectral information the results would be negative, without leading to the next logical step. Material **x** might actually be present, although to detect its weak signature it might be

necessary to examine the total spectral signature, rather than just the absorption band.

Consider green vegetation. The strong absorption bands of chlorophyll are diagnostic in most natural contexts. This material is so important for terrestrial image interpretation, and so distinct spectrally in the visible and near-infrared, that all earth-observing multispectral systems have placed channels at wavelengths that respond to the maximum spectral contrast of chlorophyll. In fact, two channels, one where absorption is near a maximum, and another on the long wavelength wing of the bands, often suffice for discriminating green vegetation. Hence there are various vegetation indices that are based on radiance in two channels.

But even the strong signature of chlorophyll becomes masked by other spectra when green vegetation comprises less than about 20% of a pixel. Thus, ratio-based vegetation indices do not work well, or at all, in areas of sparse vegetation. It is even more difficult to detect and quantify vegetation by these means when the substrate is spectrally variable. We have found that vegetation can be detected and quantified at the 5-10% level of cover by using the strategy outlined above, where green vegetation is treated as one endmember that mixes spectrally with others (Smith et al., 1989). This approach opens the way to improved mapping of vegetation cover in arid lands.

We also have quantified vegetation cover in a desert scrubland, using an hyperspectral AIS images spanning only the wavelength region from 1.2 to 2.4μm, a region that does not include the diagnostic chlorophyll bands. In this wavelength region we used the continuum of the vegetation spectrum as a spectral endmember (Smith et al., 1988). The continuum part of a spectrum commonly contains important information, and should not be discarded just because narrow absorption bands are not present.

The analysis of vegetation imaged by AIS shows that strong and narrow absorption bands may not always be encompassed in hyperspectral data, even though materials having such bands may be present in the scene. A narrowly focused analytic approach that seeks only such bands may not apply in these situations, and in any case such approaches may overlook critical information in the rest of the data.

Surprises

A successful strategy for analyzing and interpreting multispectral images must allow for the unexpected. Without any analytical strategy discoveries are left to chance. If the strategy is too narrow it will yield only "yes" or "no" results. A classic example of a spectral surprise was the discovery of the mineral buddingtonite at Cuprite, Nevada, using AIS data (Goetz and Srivastava, 1985). The mineral had not been detected in the field because it was visually identical to other minerals present. Its presence was revealed by absorption bands in the near infrared in the AIS data; however, the bands did not correspond to reference spectra then available. Field checking and laboratory analyses subsequently demonstrated the presence of buddingtonite, until then an obscure mineral to most spectroscopists.

The approach that we have proposed is well suited to reveal surprises. If there is an unexpected material within our hypothetical image there are two possibilities. One is that the spectrum of the mystery material is indistinguishable from mixtures of vegetation, soil and shade in all channels, in which case it will not be revealed by any spectral processing, and it only could become evident by its spatial pattern or by other means. The second possibility is that the material is not modeled by a mixture of the spectral endmembers. In this case the pixel(s) having the poor fit will be revealed in the rms-error image and/or in the fraction-overflow image. These pixels can then be examined to see whether they have the spectrum of a pure material in the reference library, or whether they fit as mixtures of different suites of endmembers. The images highlighting the anomalous pixels also would show any characteristic spatial pattern. If the spectral properties were not understandable in terms of the available reference materials, and the area could not be identified by shape or outside information, a field check might be warranted.

Conclusions

We have outlined an approach to the processing of multispectral images, including hyperspectral data, in which a scene is modeled as fractions of a set of spectral endmembers. Fraction images are produced that can be edited and displayed to form map units that are interpretable in terms of a conventional compositional frame of reference of field and laboratory measurements. With further development the approach has the potential of being implemented, in whole or in part, in real time aboard sensor platforms. The advantages would be that the data volume would be significantly reduced by transmitting a few fraction images rather than transmitting the output from hundreds of channels, and, at the outset, the analyst would focus on interpreting images that were understandable in terms of materials on the surface.

References

1. J.B. Adams, M.O. Smith, and P.E. Johnson, "Spectral mixture modeling: A new analysis of rock and soil types at the Viking Lander I Site," *J. Geophys. Res.*, vol. 91, pp. 8098-8112, 1986.
2. A.F.H. Goetz, G. Vane, J. Solomon, and B.N. Rock, "Imaging spectrometry for Earth remote sensing," *Science*, vol. 228, pp. 1147-1153, 1985.
3. A.F.H. Geotz and V. Srivastava, "Mineralogical mapping in the Cuprite mining district, Nevada," Proc. Airborne Imaging Spectrometer Data Analysis Workshop, *JPL Pub. 85-41*, pp. 22-31, 1985.
4. A.M.G. Possolo, J.B. Adams, and M.O. Smith, "Mixture models for multispectral images," *J. Geophys. Res.*, submitted.
5. H. Shipman and J.B. Adams, "Detectability of minerals on desert alluvial fans using reflectance spectra," *J. Geophys. Res.*, vol. 92, pp. 10391-10402, 1987.
6. M.O. Smith and J.B. Adams, "Interpretation of AIS images of Cuprite, Nevada using constraints of spectral mixtures," Proc. Airborne Imaging Spectrometer Data Analysis Workshop, *JPL Pub. 85-41*, pp. 62-68, 1985.
7. M.O. Smith, D. A. Roberts, H.M. Shipman, J.B. Adams, S.C. Willis, and A.R. Gillespie, "Calibrating AIS images using the surface as a reference," Airborne Imaging Spectrometer Workshop III, *JPL Pub. 88-30*, pp. 63-69, 1988.
8. M.O. Smith, S.L. Ustin, J.B. Adams, and A.R. Gillespie, "Variation of vegetation along environmental gradients in desert regions, Owens Valley, CA: I. Estimating vegetation abundance from remotely sensed images," in preparation.
9. S.L. Ustin, J.B. Adams, C.D. Elvidge, M. Rejmanek, B.N. Rock, M.O. Smith, R.W. Thomas, and R.A. Woodward, "Thematic mapper studies of semiarid shrub communities," *Bioscience*, vol. 36, pp. 446-452, 1986.

INVESTIGATION OF THE RELATIONSHIP OF SAR HH AND VV BACKSCATTER TO SURFACE ROUGHNESS AND DIELECTRIC CONSTANT

S. D. Wall and J. A. VanZyl
Jet Propulsion Laboratory
California Institute of Technology
Pasadena CA 91109 USA
(818)354-7424
TELEX67-5429

With the advent of polarimetric radar instruments it is sometimes possible to employ all possible combinations of both transmit and receive polarizations to infer target properties from remote sensing SAR data. However, remote sensing missions limited by either downlink data rate or high-reliability technology must rely on a correspondingly limited data set. Because most models are underconstrained even with full polarimetric data, the data-limited case becomes one of particular concern.

Planetary missions are generally faced with both data and technology limitations. If they are to produce quantitative estimates of surface properties, planetary scientists must employ and verfiy models which can be inverted to predict statistical properties of targets given the data available. The Magellan mission will map 80% of the surface of the planet Venus with S-band, HH-polarized SAR during its first full cycle of the planet's rotation. Near-nadir S-band data will be acquired simultaneously by an altimeter. Magellan will have the ability to acquire VV data during extended-mission cycles. It may also then acquire data at multiple incidence angles.

In this paper, predictions based on a second-order Bragg-Rice electromagnetic interaction model will be compared to actual results obtained in the recent NASA Mojave Field Experiment. Targets of low, intermediate, and extreme roughness in the California Mojave Desert were overflown by the JPL Airborne Imaging Radar, a full polarimetric SAR operating at L, C, and P band. Target surface roughness and dielectric constant were measured simultaneous with the overflight. HH and VV images were later reconstructed and

calibrated to simulate data expected from the Magellan (Venus) extended mission. Ground measurements were then used to predict HH and VV returns. Results will be compared with predictions, and methods for inversion of the model to use Magellan data will be discussed.

The research described in this paper was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

RADAR SCATTERING CLASSIFICATION MAPS FROM MULTIFREQUENCY IMAGING RADAR POLARIMETRIC DATA

by Howard A. Zebker, Jakob J. van Zyl, and Tom G. Farr
Jet Propulsion Laboratory
California Institute of Technology
300-235, 4800 Oak Grove Drive
Pasadena, Ca 91109
(818) 354-8780

We have utilized a three-wavelength (6cm, 24cm, and 68cm) imaging radar polarimeter on board a NASA DC-8 aircraft to classify terrain according to its multifrequency polarimetric scattering behavior. Specifically, we measure the amount of depolarization at each of the radar wavelengths and classify each pixel in the corresponding radar image according to which wavelength yields the greatest depolarization.

The apparent depolarization of the radar echoes is related to the pixel to pixel variation in observed scattering properties. This variation may be due to a number of different physical phenomena, including multiple scatter from rough surfaces, volume scattering from subsurface layers or vegetation canopies, or gross meter-scale variation in the physical properties of the the surface. The multiwavelength feature of the polarimeter permits estimation of these effects, and hence inferred surface parameters at different length scales or subsurface parameters at different depths.

We have applied our classification algorithms to radar images of sea ice and also land terrain consisting of alluvial surfaces of different ages. We find that in the former case we can readily distinguish multiyear from first year ice, while in the latter there is a correspondence between the age of the alluvial surface and the classification produced by our algorithm. In the multiyear ice instance, we believe that volume scattering from centimeter scale oblong inclusions leads to the observed result. For the alluvial surfaces, it is likely that vegetation plays a significant role. However, at present our models are incomplete, and we do not in every case know if we are observing the expression of surface or subsurface parameters.

APPLICATION OF A SAR IMAGE SIMULATOR TO THE STUDY OF THE POLARIZATION SIGNATURE OF MAN-MADE TARGETS

J.M. NASR

AEROSPATIALE DS
STS/M
BP 96 - Route de Verneuil
78133 LES MUREAUX Cedex
FRANCE

ABSTRACT

A new approach to use radar polarimetry to enhance the detectability of an artificial man-made target in a SAR image is proposed.

The method is based upon the use of a SAR image simulator to produce the target polarization signature. Once this signature has been obtained, it is possible to evaluate an optimal polarization of the electromagnetic wave which will maximize the target echo. Then, by synthesis, one can compute from polarimetric data (real multipolarization SAR or simulation products) the target optimal image.

The results show once again the interest of polarimetry for SAR remote sensing.

KEYWORDS: Polarimetry, SAR, Target detection

1. INTRODUCTION

The purpose of this paper is to show that it is possible to enhance the detectability of a particular target in a SAR image if one disposes of a multi polarization SAR antenna providing for two orthogonal linear emitted polarizations the received co and cross polarized signals.

This can be done using the following algorithm which can be divided into three major steps.

A: First of all, we must compute the polarization signature of the target. The inputs are obtained either from the HH, HV, VH and VV images of the target obtained from a real multipolarization SAR, or from a SAR simulator. The above multipolarization data enables the computation of the target scattering matrix and the polarization signature.

B: Second of all, we evaluate, from the polarization signature, the optimal polarization angle to be used to obtain the best echo on the image. The criteria will be properly defined.

C: Finally, we compute the target image from the multipolarization data with the polarization evaluated in the second step.

The result is the optimal image, for the given target, in terms of detectability. The following part of this paper will describe these steps. We shall show some simulation results at the symposium.

2. NOTATIONS

In this paper, we shall use the following notations.

2.1 SAR and frame notations

The reference frame will be linked to the SAR geometry. Fig. 1 shows the major elements.

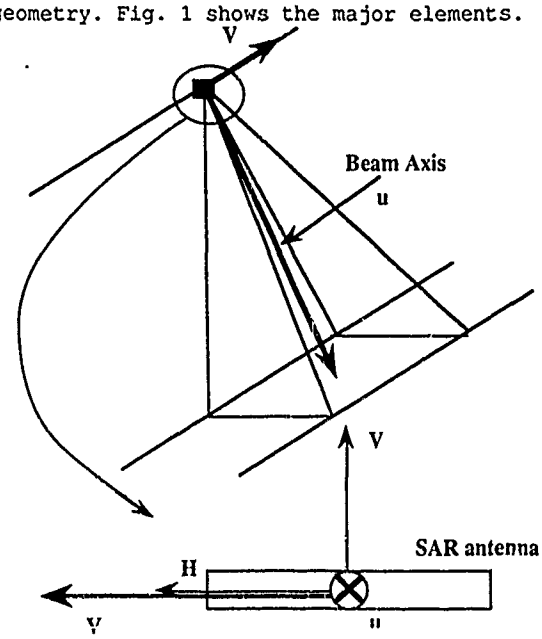


Figure 1: SAR and antenna geometry

At the top of fig. 1, one can notice a classical SAR geometry with the beam axis directed by vector u and the satellite velocity V. The bottom of fig. 1 is an enlargment of the antenna zone. In the antenna plane, we defined two vectors:

Vector H: Parallel to the velocity
Vector V: Orthogonal to the velocity and to the beam axis.

2.2 Polarimetry notations

Since we work with linear polarization, the electric field **E** lies in the HV-plane and the angle θ between H and **E** is called polarization angle. Classicaly, in a multipolarization SAR, the radar emits in H and V and receives the copolarized and crosspolarized signals. Thus, we obtain four images:

EMISSION	RECEPTION	RESULT
H	H	HH
H	V	HV
V	H	VH
V	V	VV

With these four results in module and phase (complex numbers) it is possible to obtain the scattering matrix of the imaged target (Ref.4) which can be written as:

$$\begin{pmatrix} EsH \\ EsV \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} EiH \\ EiV \end{pmatrix} \qquad (1)$$

with

$$a_{11} = \frac{\sqrt{\sigma_{HH}}\ e^{j\varphi_{HH}}}{\sqrt{4\pi}\ r} \qquad a_{12} = \frac{\sqrt{\sigma_{HV}}\ e^{j\varphi_{HV}}}{\sqrt{4\pi}\ r}$$

$$a_{21} = \frac{\sqrt{\sigma_{VH}}\ e^{j\varphi_{VH}}}{\sqrt{4\pi}\ r} \qquad a_{22} = \frac{\sqrt{\sigma_{VV}}\ e^{j\varphi_{VV}}}{\sqrt{4\pi}\ r}$$

where
- σ_{HH}, σ_{HV}, σ_{VH}, σ_{VV} are the target Radar Cross Sections (RCS).
- φ_{HH}, φ_{HV}, φ_{VH}, φ_{VV} are the associated phase shifts.
- Eix ($x \in \{ H ; V \}$) is the projection of the incident electric field on the x-axis.
- Esx ($x \in \{ H ; V \}$) is the projection of the scattered electric field on the x-axis.

Once this matrix is obtained, it is easy to simulate any incident and scattered field because the scattering matrix is a fully representation of the behaviour of the target. Assuming an emission with a polarization angle φ, **Ei**, and a reception with a polarization angle ψ, we then can write the following formula using plane rotations:

$$\begin{pmatrix} Es\psi \\ Es\psi\perp \end{pmatrix} = \begin{pmatrix} \cos\psi & \sin\psi \\ -\sin\psi & \cos\psi \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix} \begin{pmatrix} Ei\varphi \\ 0 \end{pmatrix}$$

Esψ is the received field on the polar ψ and Esψ⊥ is the cross polarized field. Eiφ is the emitted field on the polarization φ. Figure 2 represents these fields.

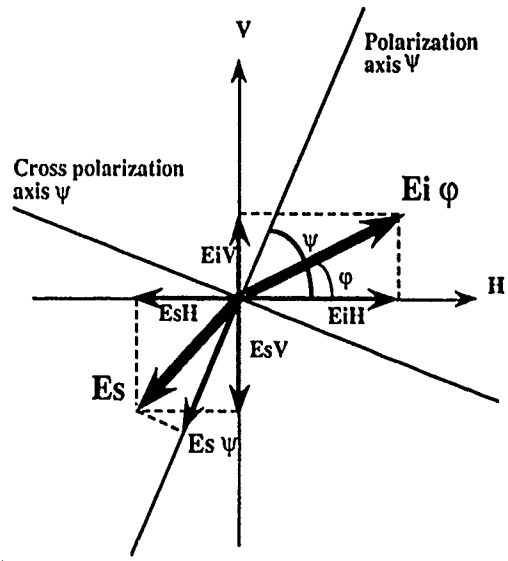


Figure 2: HV multipolarization frame.

3. POLARIZATION SIGNATURE

3.1 Target description

This is the first step of our algorithm. Since we want to study a particular target for which we do not dispose of multipolarization data, we use a SAR simulator to produce the four images HH, HV, VH and VV. This simulator is fully described in ref.2 and 3. The target can be decomposed into a collection of Elementary Geometric Reflectors (EGR) backscattering the radar wave after one, two or three internal reflections. Therefore, it becomes necessary to study the signature of each EGR if we want to manage correctly the behaviour of the whole target.

In this study, we shall consider the following EGR:

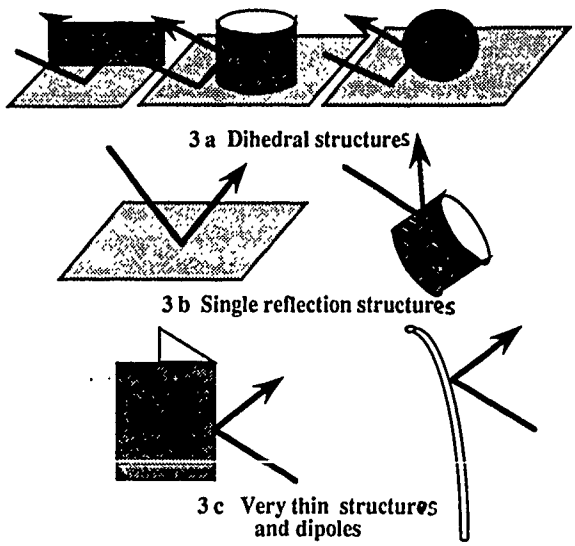


Figure 3: Three different EGR types

Type a: It concerns the dihedral structures as dihedral corner reflectors or group cylinder-plane.

Type b: It concerns the strutures which backscatter the radar wave after a single specular reflection and do not depolarize the incident wave.
Type c: It concerns the very thin structures like wedges or thin wires which behaves almost like dipoles in terms of polarization.

3.2 EGR Polarization signature

In this study, the polarization signature is represented by a polar coordinates defined curve $r=f(\theta)$ in the HV plane. The extremity of the ray-vector is function of the polarization angle θ and represents the relative power, with respect to the emitted power, received on the emission-like polarization. The scales are graduated in %.

3.2.1 Type a

The polarization signature of a dihedral structure can be derived from the scattering matrix. For an incidence perpendicular to the dihedral axis, a_{21} and a_{12} have zero values. The diagonal terms depends strictly on the Fresnel reflection coefficients (Ref.4). The polar coordinates curve can be written:

$$r = A \left(K1 \cos^2\theta + K2 \sin^2\theta \right) \qquad (3)$$

where A is the RCS power factor,
K1 and K2 are function of the Fresnel reflection coefficients.

One will notice in figure 4 two polarization signatures corresponding to two different reflectors. In these examples, the dihedral axis is parallel to the H direction.:

- Curve 4 a) is the polarization signature of a perfect conductor dihedral corner reflector. The incidence angle is equal to 23° (ERS-1).

- Curve 4 b) is the polarization signature of a real dihedral corner reflector. The incidence remains equal to 23° and the permittivity of the two planes is $\varepsilon = 8\ \varepsilon o$.

3.2.2 Type b

The type b EGR backscatter the radar wave after a single reflection. Some of them (flat plate on normal incidence, ellipsoid etc..) do not depolarize the incident field. Thus, in this case, the polarization signature is a circle.

3.2.3 Type c

These reflectors are very different from the others. Indeed, they have generally one optimal polarization direction. For a dipole, this direction is parallel to the dipole axis and the curve can be expressed by:

$$r = A \cos^2\theta \qquad (4)$$

Where A is the RCS power factor.

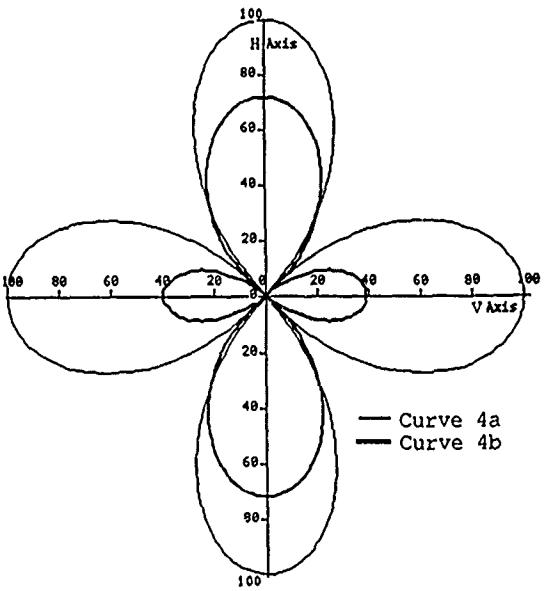


Figure 4a) and 4b): Relative polarization signatures for two dihedral corner reflectors. Two nulls can be noticed on the like polarization.

4. OPTIMUM POLARIZATION EVALUATION

4.1 Method description

The aim of this process is the computation of an optimal polarization axis for the detection and the identification of a given fully described target (dimensions, position, attitude and dielectric nature). The solution will be function of all the target parameters.

First of all, we must divide the target into a collection of Elementary Geometric Reflectors, each one of them having a different behaviour with respect to the radar illumination.

Then, we must find a criterion to define the optimal polarization to be used for a particular target. An obvious criterion would be to maximize the sum of the EGR polarization signatures. However, this may lead to an ambiguous position. For example, if the target is composed of two orthogonal equal power dipoles lying in a plane perpendicular to the incidence direction, the polarization signatures will be perpendicular. Thus, the signatures sum will be a circle ($r = A \cos^2\theta + A \sin^2\theta = A$) and the maximum is undetermined. Moreover, if the dipoles are different, the maximum will be relative to the greatest one and the less powerful will be completly invisible.

Thus, the criterion we have chosen is not related to maximum power considerations. It can be expressed with two conditions:

C1: All the EGR must be visible
C2: Once C1 is realized, we must maximize the EGR signatures.

The condition C1 refers to a threshold in terms of signal to noise ratio. It means that under a certain value, the EGR can be visible

but hidden in the noise (speckle etc...). The method used to find the best compromise is based upon these two conditions.

Assume a target to be composed of n independent EGR: it is then possible to trace the normalized polarization signatures of these EGR on the HV plane. Then, for each polarization angle θ, we determine the smallest value among the n signatures. Finally, we obtain a polar coordinates defined curve $s = f(\theta)$ which can be written as:

$$s = f(\theta) = \text{Min } (Sig_i(\theta);\ i \in \{1, \ldots, n\}) \quad (5)$$

where $Sig_i(\theta)$ is the i^{th} EGR polarization signature. Then we compute the maximum of s when θ ranges from 0 to 2π. Based on the value of θ for this maximum, an estimate of the optimal angle θ_{OPT} can be derived as follows:

$$s(\theta_{OPT}) = \text{Max } (s(\theta);\ \theta \in [0, 2\pi[) \quad (6)$$

Once θ_{OPT} is obtained, we can evaluate the value of each $Sig_i(\theta_{OPT})$ for $i \in \{1, \ldots, n\}$ and verify if they remain above the threshold. If it is not the case, we must accept to loose one EGR (the worst case) and to run the estimation process again.

C1: All the EGR must be visible
C2: Once C1 is realized, we must maximize the EGR signatures.

The condition C1 refers to a threshold in terms of signal to noise ratio. It means that under a certain value, the EGR can be visible but hidden in the noise (speckle etc...). The method used to find the best compromise is based upon these two conditions.

Assume a target to be composed of n independent EGR: it is then possible to trace the normalized polarization signatures of these EGR on the HV plane. Then, for each polarization angle θ, we determine the smallest value among the n signatures. Finally, we obtain a polar coordinates definec curve $s = f(\theta)$ which can be written as:

$$s = f(\theta) = \text{Min } (Sig_i(\theta);\ i \in \{1, \ldots, n\}) \quad (5)$$

where $Sig_i(\theta)$ is the i^{th} EGR polarization signature. Then we compute the maximum of s when θ ranges from 0 to 2π. Based on the value of θ for this maximum, an estimate of the optimal angle θ_{OPT} can be derived as follows:

$$s(\theta_{OPT}) = \text{Max } (s(\theta);\ \theta \in [0, 2\pi[) \quad (6)$$

Once θ_{OPT} is obtained, we can evaluate the value of each $Sig_i(\theta_{OPT})$ for $i \in \{1, \ldots, n\}$ and verify if they remain above the threshold. If it is not the case, we must accept to loose one EGR (the worst case) and to run the estimation process again.

4.2 One result

It concerns a group of two dihedral corner reflectors in the HV-plane. Figure 5 shows this geometry:

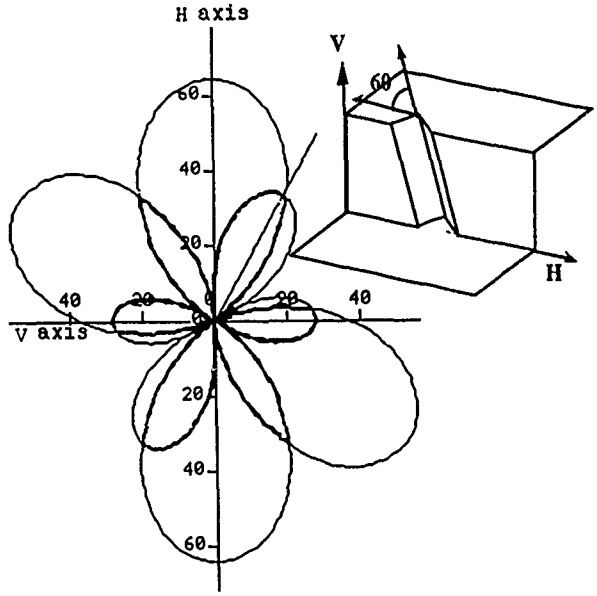


Figure 5 and 6: Two dihedral corner reflectors structure and best polarization angle.

For this geometry, the best polarization angle is 60° and the EGR are attenuated by:

DCR 1: 47%
DCR 2: 63%

5. OPTIMAL SIMULATION

This is the last step of our process. Once the best polarization angle for the given target is obtained, it is easy to simulate the image of this target with this polarization. Thus we are in the optimum conditions if we want to detect this particular object.

We shall show some results at the symposium since they are not ready for presentation at the time of the writing of this article.

6. REFERENCES

1. **EVANS** D.L.,"Radar Polarimetry: Analysis, Tools and Applications". IEEE trans.on geosc. and rem. sensing. Vol 26 n°6 pp 774-789 - Nov 1988.

2. **NASR** J.M.,**VIDAL-MADJAR** D.,"Image Simulation of geometric targets for spaceborne SAR".Proc. IGARSS'1988 - Edinburgh

3. **NASR** J.M.,"Simulation d'images de cibles geometriques pour radar à synthèse d'ouverture". Doctoral thesis - PARIS 7 - Feb 1989.

4. **VAN ZYL** J.J., **ZEBKER** H.A., **ELACHI** C.,"Imaging radar polarization signatures: Theory and Observation".Radio Science Vol. 22 n°4 pp 529-543, Aug. 1987.

CLASSIFICATION OF EARTH TERRAIN COVER USING MULTIFREQUENCY IMAGING RADAR POLARIZATION DATA

J. J. van Zyl and C. F. Burnette
Jet Propulsion Laboratory
4800 Oak Grove Drive
Pasadena, CA 91109
TEL (818) 354 - 1365
FAX (818) 354 - 3437
TELEX 67 - 5429

There has been considerable interest in applying the additional information that radar polarimetry could provide to classification of earth terrain cover. For example, earlier results of an unsupervised classification algorithm showed that polarization information allows one to classify scattering behavior based on the number of reflections that the wave suffered before returning to the radar receiver. Also, it has been shown that when accurate training data are available, a supervised Bayes classifier yields better performance than the unsupervised classification algorithm and simpler discriminates such as single channel amplitudes, or the phase difference between two channels.

In this paper we compare the results of supervised classification algorithms based on the Bayes technique (maximum likelihood) applied to the first order statistics of the measured signals, as well as a simpler minimum distance technique applied to the second order statistics of the measured signals. Both techniques are applied to single frequency and multiple frequency (wavelengths of 67, 24 and 6 cm) imaging radar polarimetry data sets acquired over various different earth terrain covers such as sea ice, desert areas, grasslands, forests and urban areas.

Our results show that the two algorithms yield very similar results, but that the minimum distance method uses considerably less computer time than the Bayes method. Also, results are considerably more accurate when the full three-frequency set is used than when any single frequency set is used. In some cases, like desert terrains, most information seems to be contained in a relatively small subset of the feature space. This means that the dimension of the minimum distance feature vector could be reduced, resulting in an increase in speed without a significant loss in classification accuracy.

SEGMENTATION OF MULTIFREQUENCY POLARIMETRIC RADAR IMAGES TO FACILITATE THE INFERENCE OF GEOPHYSICAL PARAMETERS

C. F. Burnette, P. C. Dubois and J. J. van Zyl

Jet Propulsion Laboratory
California Institute of Technology
4800 Oak Grove Drive
Pasadena, CA 91109

ABSTRACT. Geophysical parameters, such as r.m.s. height, correlation length and dielectric constant, may be derived for each resolution element of a scene using model inversion techniques, although in practice this is prohibitive because of the amount of computation involved. Segmenting the images into classes with similar geophysical characteristics allows the model inversion to be performed on each class rather than each pixel. We use an unsupervised clustering algorithm to segment multifrequency polarimetric radar data from the NASA/JPL airborne SAR. Twenty-two parameters are evaluated for their discriminatory capability for each pixel of an image. A clustering analysis is then performed using different subsets of these parameters. This analysis relies on data taken as part of an intensive field experiment during the summer of 1988 in the vicinity of the Pisgah lava flow in the Mojave Desert in southern California. As part of the experiment, extensive ground truth was acquired, including dielectric constant and topography measurements. Segmentation results show good agreement with these measurements.

Keywords. Clustering, Polarimetry, Statistical Separability

INTRODUCTION

There is considerable interest in deriving geophysical information, e.g. surface roughness, dielectric constant and soil moisture from remote sensing data. Since radar backscatter from rough surfaces is influenced by both the geometrical and electrical properties of the surface, SAR data may allow inference of physical surface parameters. Model inversion techniques achieve this derivation but require a large amount of computation when applied on a pixel to pixel basis. Segmenting the images into classes with similar geophysical characteristics allows the model inversion to be performed on each class rather than each pixel.

The new JPL Aircraft SAR acquires fully polarimetric data at C-, L- and P-bands [1]. For each resolution cell, the resulting data set consists of three symmetrical Stokes matrices, one for each frequency. There are nine independent parameters in each Stokes matrix, for a total of twenty-seven independent parameters per resolution cell. Therefore, the first step in a segmentation process consists of reducing the dimensionality of the data, i.e., selecting a limited set of parameters among the original twenty-seven which contain most of the information for discrimination purposes among differing geologic surface types. Several parameters appear to contain little information for classifying geologic surfaces, while others discriminate well between surface types. A clustering algorithm is then used to segment images using the reduced set of polarization parameters. The final step, inversion of geophysical parameters for each class, is presented in a related paper [2].

DESCRIPTION OF DATA

The Aircraft SAR system outputs scattering matrices directly, but further data volume reduction transforms the data into compressed Stokes matrix format [3]. The following analysis was performed on the standard JPL polarimetric compressed data files. A four stage calibration process, including phase, cross-talk, amplitude and radiometric calibration, was performed on the data before analysis [4].

The JPL Aircraft SAR system was flown on a NASA DC-8 over the Pisgah lava flows and the adjacent dry lake bed in the Mojave desert in June 1988. Our analysis is done on a 10 x 8 km image containing 1024 x 800 4-look pixels. The L-band total power image is shown in Figure 1, although data from C- and P-bands are also used. The L-band data contains interference patterns noticeable as two vertical bar-shaped boxes in the lower part of the image. In the studied data set, a wide variety of geologic surfaces are encountered. Six distinct classes of surface type were identified previously during field work. For each class, a typical site, approximately 16 pixels square, is selected. The sites are labeled as follows: site 1 is Playa (dry lake bed), site 2 is Pavement (cobble), site 3 is Alluvial, site 4 is Phase I Lava (smooth lava), site 5 is Phase III Lava (intermediate lava) and site 6 is Phase II Lava (rough lava). A complete description of the lava types and their origin is given in Wise [5]. The sites are shown in Figure 2.

Similar sites, each approximately 10 m on a side or one pixel, were extensively studied during the Mojave Field Experiment which was simultaneous with the radar overflight. Soil samples were taken and a precise description of the geology was noted for each site. Soil moisture and dielectric constant measurements were acquired on the non-lava sites. Surface roughness profiles were acquired with a ground template as well as with a helicopter-borne stereo camera.

Because of its diversity, the Pisgah location has been extensively studied in the past. Aerial photography, Landsat images, SPOT images and Thermal Infrared Mapping Spectrometer (TIMS) data of the area are also available.

POLARIMETRY PARAMETERS

An attempt is made in reducing the dimensionality of the multifrequency, polarimetric data by selecting the most significant parameters for the purposes of discriminating between geologic surface types. Even though the complete polarimetric scattering



Fig. 1 L-band total power image of the Pisgah scene

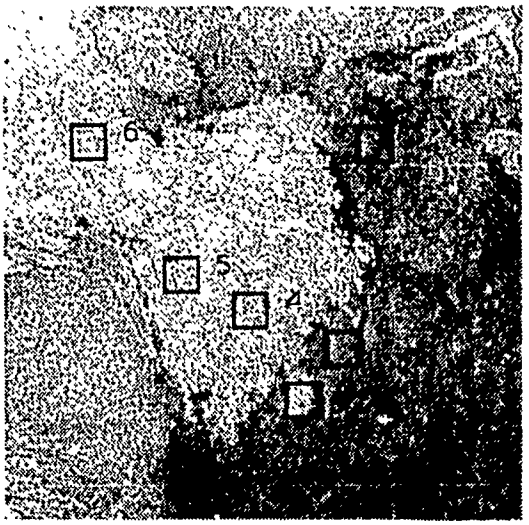


Fig. 2 Pisgah scene showing six selected sites representative of different surface types.

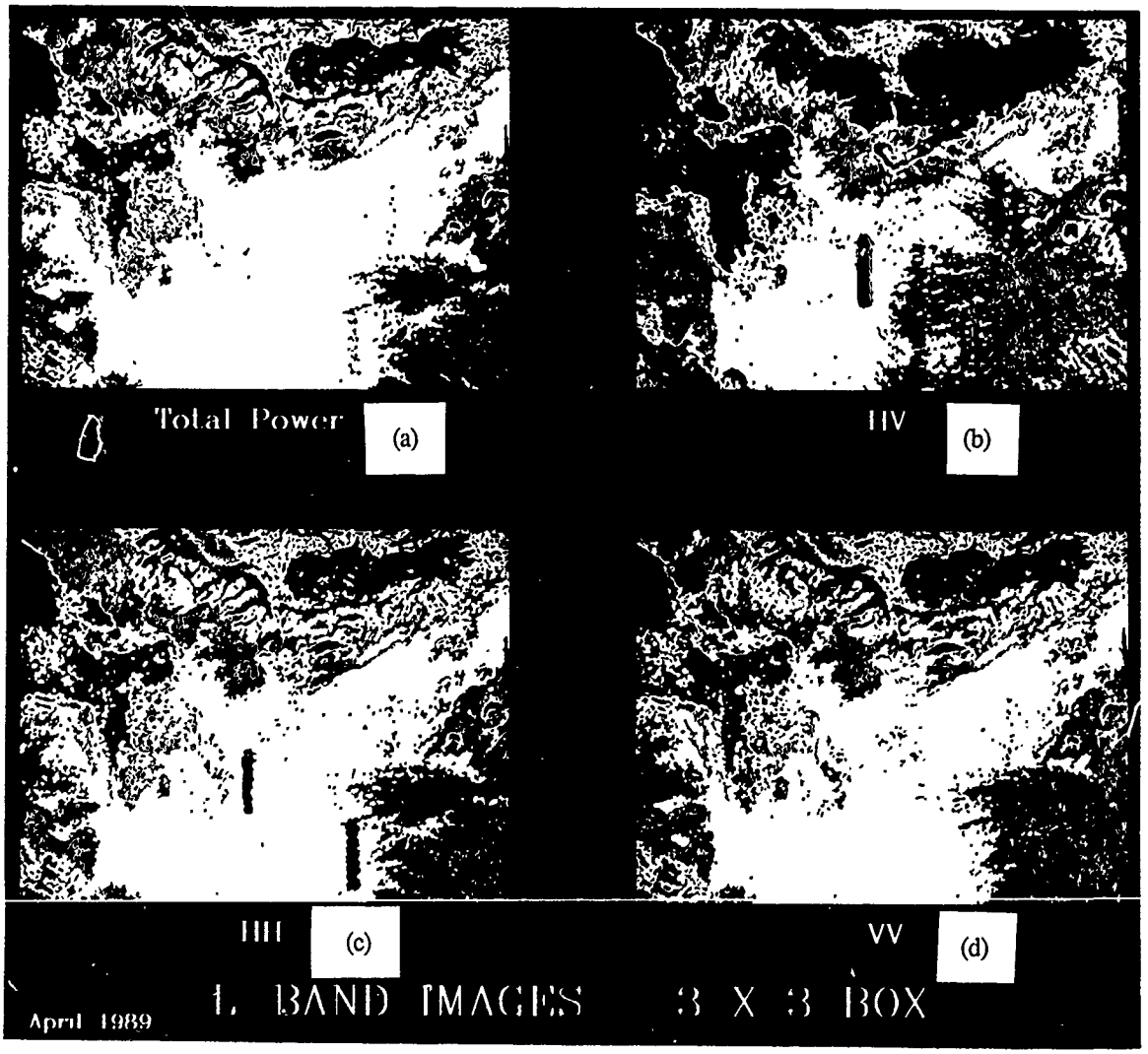


Fig. 3 L-band clustering results with one input parameter and 3 x 3 box size. The input for a) is Total Power, b) is HVHV*, c) is HHHH* and d) is VVVV*.

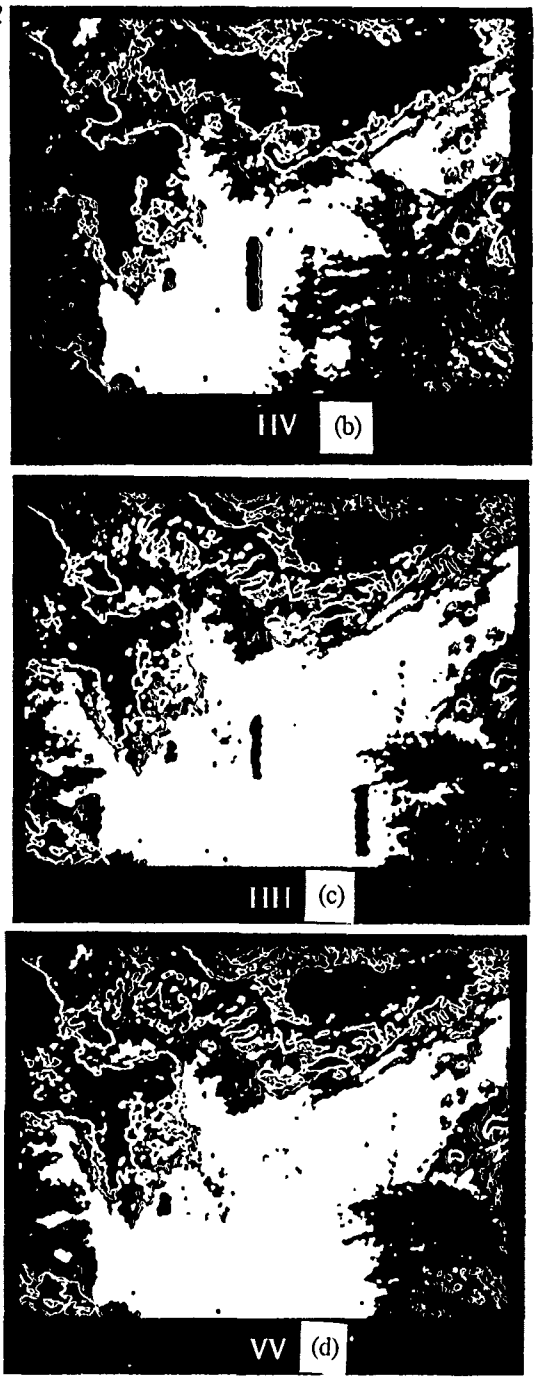


Fig. 4 L-band clustering results with one input parameter and 5 x 5 box. The input for b) is HVHV*, c) is HHHH* and d) is VVVV*.

behavior of a resolution cell is described by the nine independent elements of its Stokes matrix, twenty-two parameters are studied. They display the information contained in the Stokes matrix in different ways to emphasize diverse aspects of the scattering behavior. The twenty-two parameters may be categorized in four sets. Set 1 consists of the scattering matrix cross-product terms and the total power (span of the scattering matrix) for a total of ten elements. Set 2 includes the normalized Stokes matrix terms for a total of nine elements. Set 3 is the two phase difference terms and set 4 is the coefficient of variation, which is defined to be the minimum possible power in the co-polarized or cross-polarized configuration divided by the maximum possible power in co-polarized or cross-polarized configuration.

An analysis is performed on these parameters to determine which are the best at discriminating between the six selected sites. Three different methods are used in this analysis. First, the histograms of each parameter for each of the sites are compared. Second, statistical separability is used to compute the interclass distances based on each parameter [6]. The Jeffries-Matusita (J-M) distance appears to be the most reliable indicator of separation between two distributions. And finally, images of each parameter are created [7].

Among parameters, a good discriminator results in large interclass separation, whereas a poor discriminator presents similar distributions across classes, small interclass distances and practically random parameter images. In set 1, the largest interclass distance is obtained with HVHV*, followed closely by HHHH*, VVVV* and the Total Power. The parameters of set 2 do not discriminate as well as those of set 1. The best discriminators in set 2 are F_{12}/F_{11}, F_{33}/F_{11} and F_{44}/F_{11} where F_{ij} are the Stokes matrix elements. In set 3, Arg(HHHV*) is randomly distributed regardless of the surface type. This is consistent with predictions by most theories that the co- and cross-polarized terms of the scattering matrix are uncorrelated. Arg(HHVV*) shows distribution dependence on surface type. The coefficient of variation is consistently a good discriminator in all three frequencies.

CLUSTERING METHOD

The best of the discriminators described above are used to generate a segmentation map using an unsupervised classification, or clustering, algorithm. The clustering analysis is done using the Isodata algorithm [8]. Isodata is an iterative algorithm where a set of cluster centers are redefined every iteration until convergence. The cluster centers are redefined in an attempt to maximize intercluster distance and minimize the distance between the elements of a cluster and its center, or overall cluster size. The algorithm also includes heuristics to lump, eliminate and/or split clusters based on several user-specified parameters. Once the cluster centers have been selected, each pixel is then assigned to the nearest cluster center using a Euclidean distance measure. For each pixel, the algorithm uses both the mean and standard deviation of the pixel values from an N x N box around the central pixel.

CLUSTERING RESULTS

At this stage, only parameters from set 1 and set 2 have been used as input to the clustering algorithm. Cross-product parameters yield better results than relative Stokes matrix parameters, which result in noisy segmented images. The L-band classification images for Total Power, HVHV*, HHHH* and VVVV* are shown in Figure 3. The vertical interference bars are clearly visible in the HVHV* and HHHH* images. The HVHV* image gives a good lava type discrimination.

Clustering was performed using two different box sizes, 3 x 3 and 5 x 5. Figure 3 shows the segmentation results obtained by using 3 x 3 boxes. Figure 4 shows images obtained using the same parameters but with a 5 x 5 box. The larger box size yields better results, i.e., the resulting image has fewer isolated points and well defined borders between classes.

Running the clustering algorithm without the standard deviation information results in a slightly noisier image. There are more isolated points especially in the bright areas. The standard deviation provides a smoothing effect, although the mean is the main driver of the classification.

Multiparameter clustering was also tested with a 3 x 3 box. Images were created using Total Power, HVHV*, HHHH* and VVVV* for all three frequencies, shown in Figure 5. The image in Figure 5a is created by using the Total Power from C-, L- and P-bands as input for the clustering algorithm. Similarly, Figures 5b, 5c and 5d are

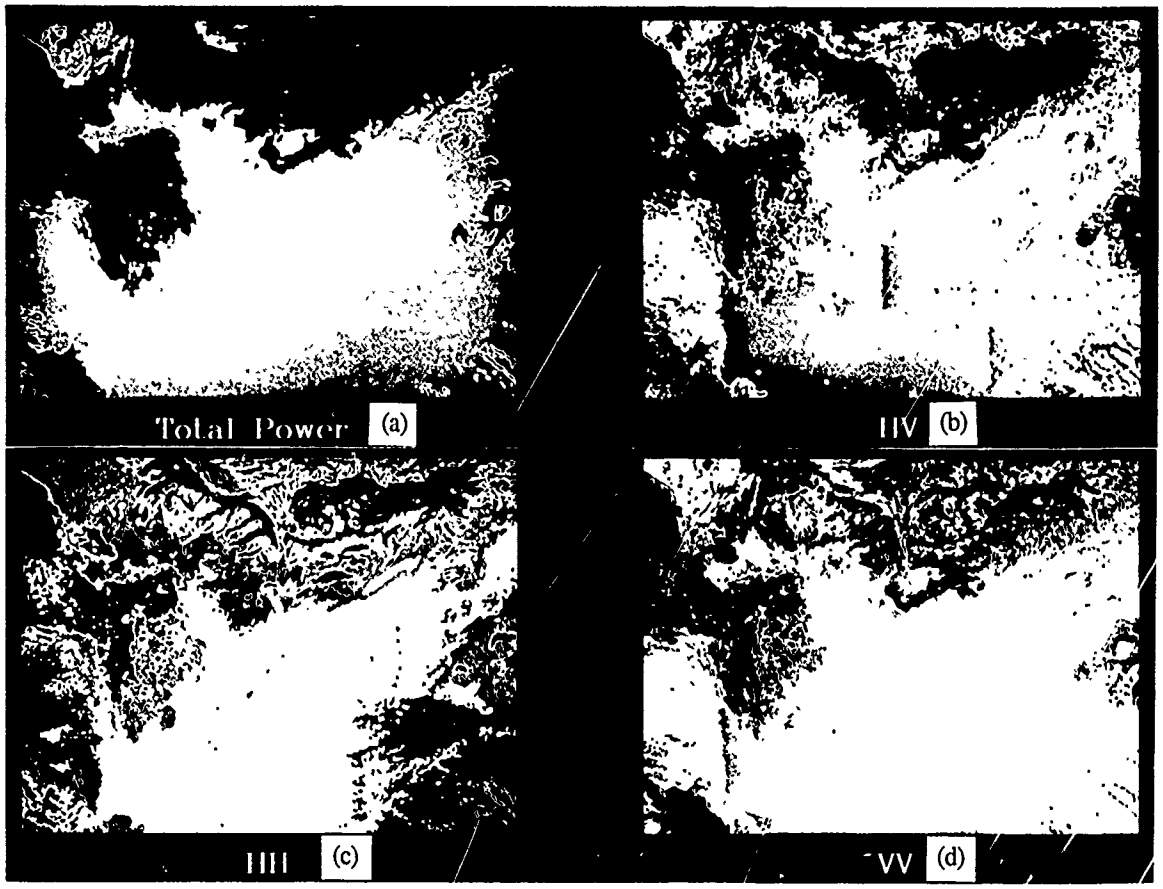


Fig. 5 Clustering results with three input parameters, one from each frequency. The input for a) is Total Power, b) is HVHV*, c) is HHHH* and d) is VVVV*.

also calculated using three frequency information. Comparing Figures 5c and 3c indicates that adding additional parameters in the clustering algorithm improves the performance.

CONCLUSION

The clustering algorithm assigns a class value to each image pixel and these classes seem to correspond to well-defined geological surface types. The initial results of the segmentation are already usable for model inversion [2].

It is apparent from the figures that the algorithm has a tendency to underestimate the number of classes in the darker areas. An investigation is under way to determine the best method for increasing the number of classes in the darker image areas. The algorithm capacity will be increased to handle more than three input parameters at once, but the result is not expected to improve significantly.

These results were derived for geologic applications and the same approach could be used to study other earth terrain cover, but there is no guarantee that the analysis would yield the same conclusions.

The parameters from set 2 are normalized, meaning that the brightness has been divided out. Judging from their poor performance, it appears that the brightness information is important for classification of geologic surfaces. Unfortunately, different geological surface types may have the same brightness. Including data from other sensors in the classification process can resolve the ambiguity [9].

REFERENCES

[1] Held, D. N. *et al.*, The NASA/JPL Multifrequency, Multipolarization Airborne SAR System, IGARSS '88 Proceedings, 1988.

[2] van Zyl, J. J., C. F. Burnette and T. G. Farr, Inference of Geophysical Parameters from Multifrequency Polarimetric Radar Observations and Model Inversion, IGARSS '89 Proceedings, 1989.

[3] Dubois, P. C. and L. Norikane, Data Volume Reduction for Imaging Radar Polarimetry, IGARSS '87 Proceedings, 1987.

[4] van Zyl, J. J., Calibration of Polarimetric Radar Images using only Image Parameters and Trihedral Corner Reflector Responses, Submitted to IEEE Transactions on Geoscience and Remote Sensing, 1989.

[5] Wise, W. S., Geologic Map of the Pisgah and Sunshine Cone lava fields, NASA Technical Letter, 11, 1966.

[6] Swain, P. H., Fundamentals of Pattern Recognition in Remote Sensing, Chapter 3 in Remote Sensing: The Quantitative Approach, McGraw-Hill, pp. 164-174, 1978.

[7] Dubois, P. C. and C. F. Burnette, Investigation of Polarimetry Parameters, JPL Internal memorandum, April, 1989

[8] Ismail, M. A. and M. S. Kamel, Multidimensional Data Clustering Utilizing Hybrid Search Strategies, Pattern Recognition, Vol. 22, No. 1, pp. 75-89, 1989.

[9] Evans, D. L. *et al.*, Incorporation of Polarimetric Radar Images into Multisensor Data Sets, IGARSS '89 Proceedings, 1989.

COMPARISON OF SEA ICE TYPE CLASSIFICATION USING POLARIMETRIC AND NON-POLARIMETRIC SYNTHETIC APERTURE RADAR

D.P. Winebrenner
D.A. Rothrock
H.L. Stern

Polar Science Center
Applied Physics Laboratory
University of Washington
Seattle, WA 98195
(206)543-6613

We first present a comparison of first-year/multi-year classification results based on polarimetric, multi-frequency data versus those based on single-polarization C-band data, using sea ice images obtained with the Jet Propulsion Laboratory aircraft SAR in the Beaufort Sea in March 1988. We examine the effects of adding polarization ratios and the degree of polarization at C-band to backscattering strength as classification clues. We compare classification results with airborne passive microwave data when possible.

Secondly, we present a new algorithm for the discrimination of thin, flat ice from open water in leads and polynyas. Both types of surfaces typically appear dark in single frequency, single polarization SAR images because both are relatively smooth and backscatter little radiation. However, the dielectric constants of sea ice and sea water differ markedly, making the ratio of like-polarized backscattering cross sections differ (theoretically) between surface types. The variation is robust with respect to natural roughness variations between and within surface types. We will present a comparison of experimental and theoretically expected signature behavior from JPL SAR images containing open water and new ice and give error estimates for threshold-based classification based on observed variances of the polarization ratio.

Key Words: SAR, sea ice type classification

Theoretical studies of backscattering at L-band from new sea ice types and from open water indicate that rough surface backscattering dominates the return observed using SAR in both cases. In this case, the theoretical ratio of backscattered powers at VV and HH polarization depends primarily on the dielectric constant of the scattering material. Since there is a large difference in dielectric constants between sea water and sea ice at L-band, this ratio may be useful for discrimination of new ice from open water (Winebrenner et al., 1989).

Figures 1 and 2 show HH-polarized, L-band SAR images of San Franscisco and adjacent ocean areas and of a region of the Beaufort Sea, respectively. Both images were obtained using a JPL polarimetric airborne SAR, the former image by the first JPL polarimeter operating aboard a CV-990 aircraft in 1986, the second by the second JPL polarimeter aboard a DC-8 in 1988. The reason for using this pair of images is that presently available sea ice images from the Beaufort Sea do not contain any areas in which open water is known to be present. In fact, environmental observations at the time strongly suggest that little open water is likely to be found in images from this time and place. Because we are concerned with new ice/open water discrimination, we included in our study areas from the San Fransisco image which are certainly composed entirely of returns from sea water. Selected areas in the Beaufort Sea image are located in an apparently new lead which backscatters little power relative to other ice in the image. For the reasons cited above, and because of the signature behavior described below, we believe these areas contained new sea ice, possibly mixed with small areas of open water. Caution should be used in interpreting analysis of data from two different imaging systems since the relative calibration between VV and HH channels may have differed between systems. Results presented below nonetheless provide some insight related to multi-polarization new ice/open water discrimination.

The specific regions of the two images used in this study are the areas enclosed by white rectangles (including the areas covered by the rectangles themselves). Each rectangle in the Beaufort Sea image is exactly 10 pixels on a side (a region approximately 66m by 120m), while the size of the rectangles in the San Francisco image varies but always contains more than 100 pixels. The incidence angles in the selected regions varied from 25 to 32 degrees for the samples of backscattering from sea water (Figure 1) in the San Francisco image, and from 29 to 36 degrees for sea ice backscattering samples in the Beaufort Sea image.

Figures 3a and b show histograms of the ratio of powers backscattered at VV and HH polarizations, computed pixel by pixel for one study area in each image. It is clear that while the means differ substantially, computing the ratio signature pixel by pixel is a noisy estimation technique. We obtained more stable results by averaging VV and HH backscattered powers over each individual study area to estimate average powers, and then computing the ratios of these average powers. The results, versus incidence angle, are shown in Figure 4. Data points marked W are from the San Francisco image and definitely contain sea water. The theoretical signatures based on Bragg rough surface backscattering (Winebrenner et al., 1989) for sea water (relative dielectric constant 70 + i80) and for sea ice (relative dielectric constant approximately 3.5 + i0.25) are plotted as solid lines. Finally, regions from the Beaufort Sea image are labeled I since they are likely to contain sea ice. The study areas from the Beaufort Sea display a noisy but ice-like signature. The areas from the San Francisco image show larger polarization ratios as are expected for sea water and follow the theoretical trend but substantially exceed the theoretical values for sea water. The reason for this, whether theoretical or instrumental, remains to be determined.

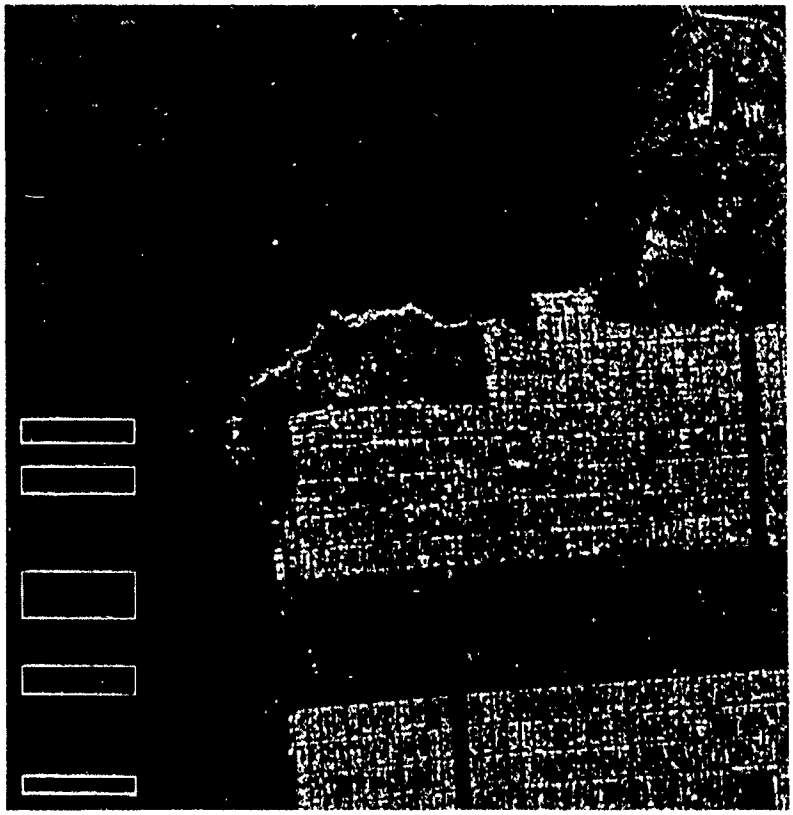


Figure 1. JPL L-band polarimeter image of San Francisco and adjacent ocean areas.

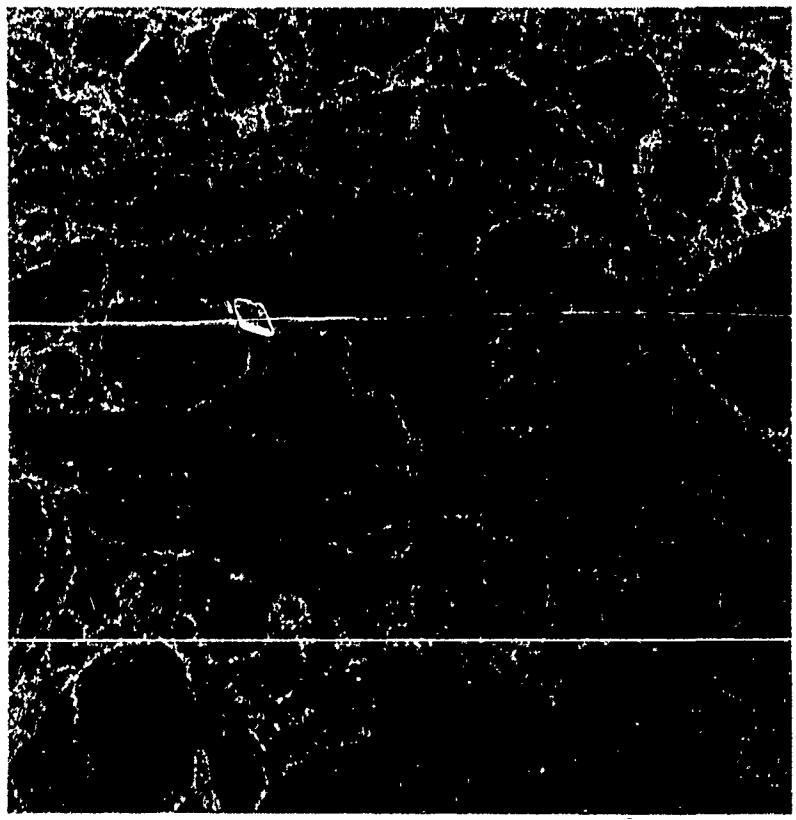


Figure 2. JPL L-band polarimeter image of an area in the Beaufort Sea.

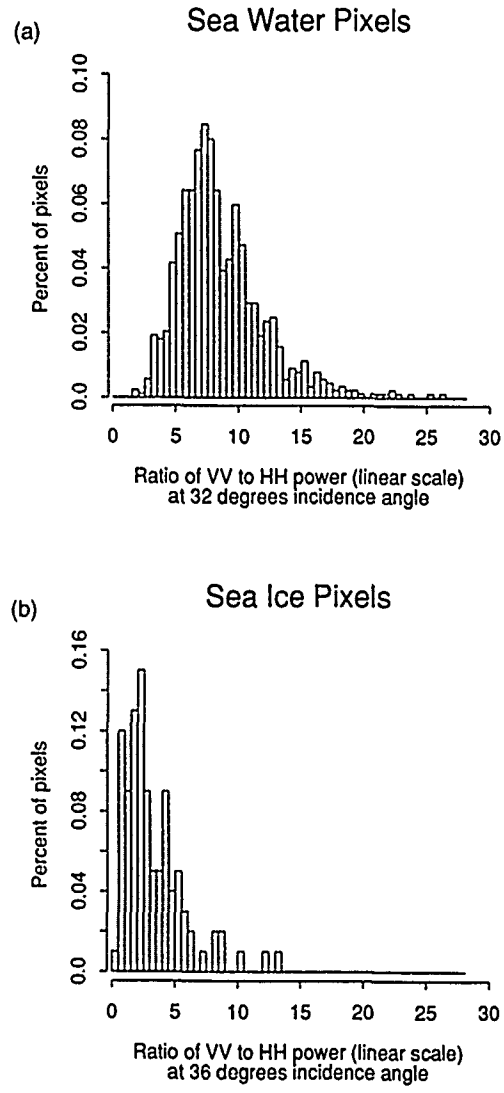


Figure 3. Observed distribution of polarization ratio signatures for particular test areas in a) Figure 1 and b) Figure 2.

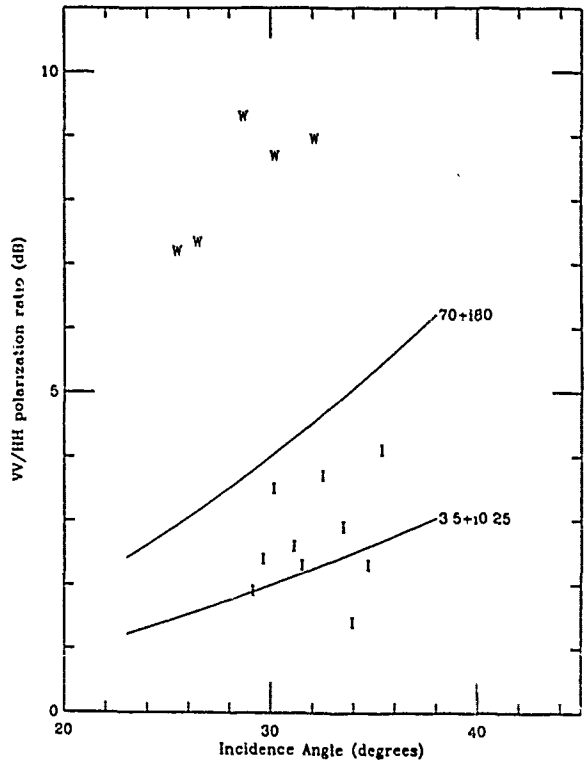


Figure 4. Comparison of observed and theoretical polarization ratio signatures for areas containing sea water and (probably) new sea ice.

Acknowledgements: The authors would like to thank those at the Jet Propulsion Laboratory who helped in providing the data used here, in particular John Crawford, Ben Holt, and Howard Zebker.

REFERENCES

Winebrenner, D.P., L. Tsang, and R. West, "Sensitivities for Two Polarimetric Backscattering Models to Geophysical Parameters," to appear in the Proceedings of the NATO Advanced Workshop on Direct and Inverse Methods in Radar Polarimetry (held 19-23 Sept., 1988), W.-M. Boerner, ed., 1989.

Detection of Dielectric Cylinders on a Rough Surface Using a Multi-Frequency Polarimetric Radar

D. Kähny*, S. Riegger, W. Wiesbeck

Institut für Höchstfrequenztechnik und Elektronik
University of Karlsruhe, Kaiserstr. 12 D-7500 Karlsruhe, FRG

ABSTRACT

This paper will present an approach for the separation of scattering components using polarimetric radar data, measured in laboratory [1]. An artificial rough surface and dielectric cylinders were used as targets. They were measured separately and combined.
By comparison of the polarimetric signatures (amplitude ratios, phase differences [2], polarization signatures [3] and characteristic polarization states [4]) of the different measurements, the authors will show methods to detect dielectric cylinders on the rough surface. Criteria for the separation of scattering components with different characteristics, using laboratory measured results and evaluating complex polarimetric signatures, will be given.

KEYWORDS

Rough Surface Scattering, Polarization, Polarimetric Signature, Radar

1. INTRODUCTION

The feasibility of information extraction from remotely sensed radar images depends significantly on the measurement parameters, as frequency and polarization. A lot more information can be extracted from polarimetric signatures [1,2,3] than from pure RCS-values. The intention of this paper is to show how such information can be extracted from wideband laboratory data for dielectric models.
For real vegetated areas the backscattered signal originates not only from the vegetation itself but also from the underlying rough soil surface. So there are different components contributing to the scattering behavior. The idea of the authors was to simulate the main components of a vegetated area, such as forest stands, by a dielectric model. By measuring the backscattering behavior of the model and evaluating polarimetric signatures out of these data

- separation criteria for the different components,
- tools for the identification of the components

can be derived.
These separation criteria and the identification parameters, respectively, can be entered into a data base for further analysis.
The laboratory measurements were performed with a wideband polarimetric radar set-up, described in [1], over a frequency range from 2-20 GHz.

2. DEFINITIONS

The fundamental radar scattering behaviour of any target can be described by its complex RCS-matrix:

$$|\sigma_{HH}|\, e^{j\varphi_{HH}} \begin{bmatrix} 1 & \frac{|\sigma_{HV}|}{|\sigma_{HH}|}\, e^{j(\varphi_{HV}-\varphi_{HH})} \\ \frac{|\sigma_{HV}|}{|\sigma_{HH}|}\, e^{j(\varphi_{HV}-\varphi_{HH})} & \frac{|\sigma_{VV}|}{|\sigma_{HH}|}\, e^{j(\varphi_{VV}-\varphi_{HH})} \end{bmatrix}$$

It is obvious, that in the monostatic case the matrix is described by 5 real values, neglecting the absolute phase. For the data evaluation in section 4 the amplitude ratios will be used for comparative purposes.
From the scattering matrix the characteristic polarization states (CPS) [4] can be deduced. They deliver the information for which polarization states of the antennas a target is optimal detectable or invisible. The mathematical derivation of the CPS is

given in detail in [4] and will not be treated further. A diagram for visualizing the CPS, based on a projection of the Poincaré sphere [5], is given in fig. 1.

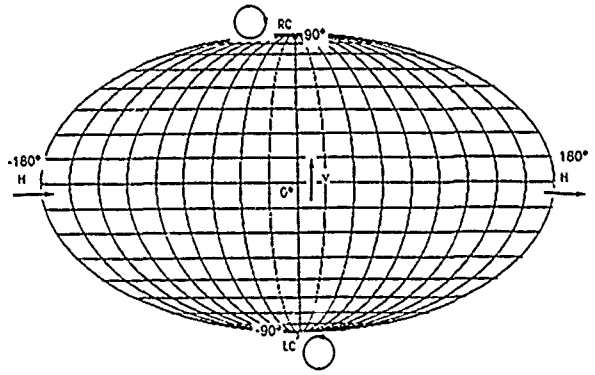


Fig. 1 Poincaré representation of a polarization state.

In this diagram a polarization state is given in spherical coordinates. The longitude angle corresponds to 2 times the rotation angle of the polarization ellipse, and the lattitude angle to 2 times the ellipticity angle of a polarization state [1]. The equator corresponds to linear, the poles to circular polarization.

3. Measured Objects

The artificial statistical rough surface was built of a polyethylene foam disc of 1 m diameter. Fig. 2 shows the digitized surface.

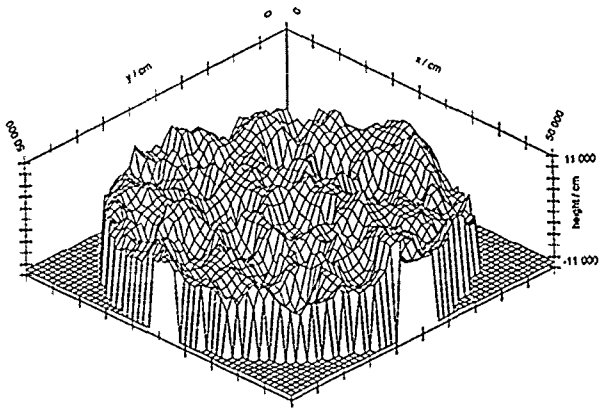


Fig. 2 Digitized rough surface model

Its statistical parameters are:

- mean height: $\mathbf{h_{mean}}$ = 8.4 cm
- standard deviation: $\Delta\mathbf{h_{std}}$ = 2.93 cm
- correl. length: $\mathbf{l_{corr}}$ = 5.3 cm
- mean surface slope: $\bar{\varphi}$= 48°

The dielectric constant ε was 1.3.

The material of the dielectric cylinders was polyvenylchlorid plastic with the complex dielectric constant of ε=3+j0.07, a height of 20 cm and a width of 2 cm.

Fig. 3 shows the measurement geometry.

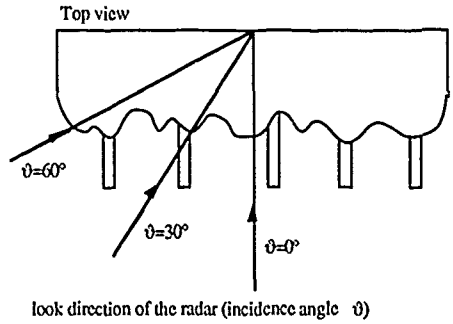


Fig. 3 Measurement Geometry

The measurements presented in section 4 were all undertaken for an incidence angle ϑ=30°. To get a statistical ensemble of measured data for one incidence angle, different footprints were taken by rotating the surface around its symmetry axis.

4. Measurement Results

For the principal understanding of the scattering behaviour of a rough surface fig. 4 presents the magnitude of the different components of the RCS-matrix for the pure rough surface averaged over 10 footprints. $|\sigma_{vv}|$ and $|\sigma_{hh}|$ show a almost identical behaviour versus frequency, due to the low dielectric constant. $|\sigma_{vh}|$ is increasing with frequency. At 17 GHz it is almost as big as $|\sigma_{hh}|$ and $|\sigma_{vv}|$.

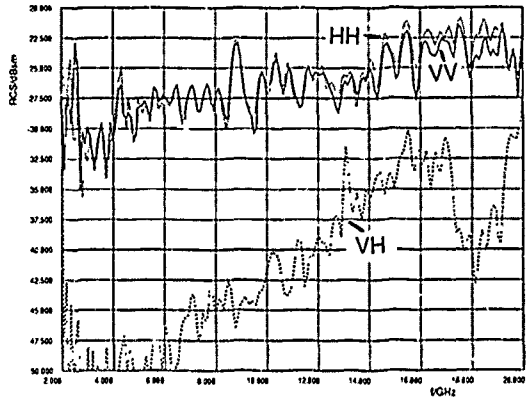


Fig. 4 $|\sigma_{hh}|$, $|\sigma_{vv}|$ and $|\sigma_{vh}|$ for the dielectric rough surface

In a second step a set of 7 cylinders was added to the surface and identical measurements as for the pure surface were performed.

In fig. 5 the amplitude ratios $|\sqrt{\sigma_{vv}}|/|\sqrt{\sigma_{hh}}|$ are compared.

For the pure surface this ratio is approximately 1. But with cylinders, especially in the frequency range

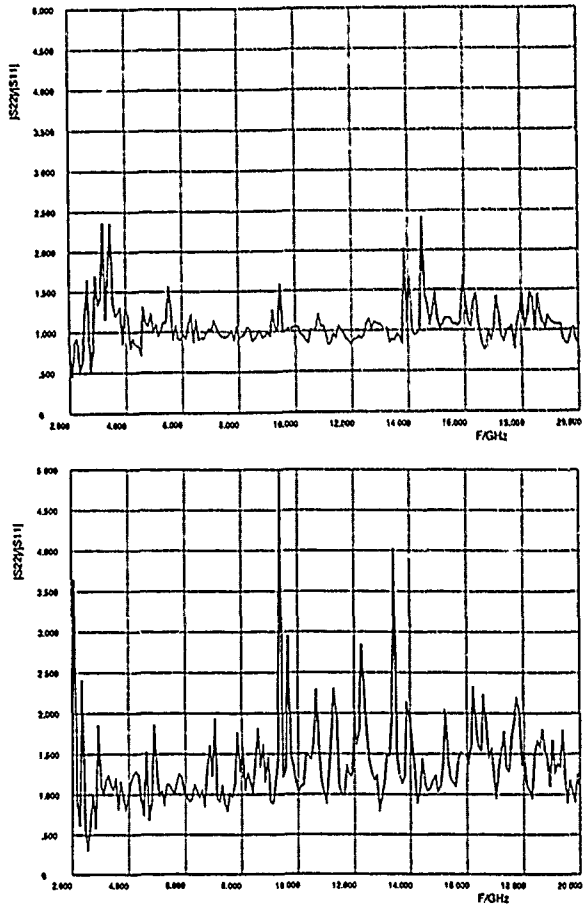


Fig. 5 $|\sqrt{\sigma_{vv}}|/|\sqrt{\sigma_{hh}}|$; rough surface (top) ; rough surface with 7 cylinders (bottom)

from 9 to 15 GHz, larger ratios occur due the cylinders.

The polarization signatures for the rough surface at 5 GHz (fig.6) shows a typical rough surface behaviour. The maximum co-polarization return results for linear polarization and the cut versus the rotation angle is smooth. For the measurement with cylinders the maximum occurs for a rotation angle of about 45°, due to the cylinders.

At 17 GHz (fig. 7) the rough surface behaviour is still the same. But for the combined measurements the maximum occurs no more for linear, but for elliptical polarization, due to multiple reflections.

Regarding now the cross-polarization null-pairs (XPOL-nulls) of the CPS (see [4]) for the measurements at 5 GHz (fig.8), the difference is more clearly. XPOL-null means for a radar, measuring cross-polarization, that for this polarization state of the antennas minimum power will be returned. It can be shown, that one of the XPOL-nulls is identical to the COPOL-max, for which a radar measuring co-polarization will receive maximum power.

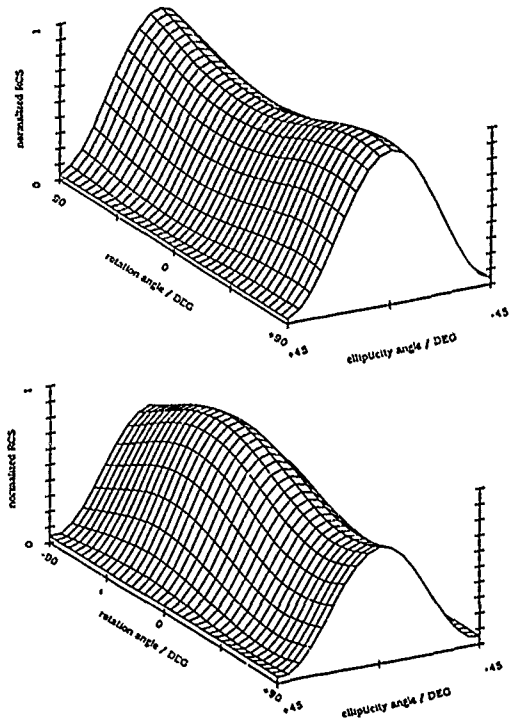


Fig. 6 Polarization Signatures (f=5 GHz); rough surface (top); rough surface with 7 cylinders (bottom)

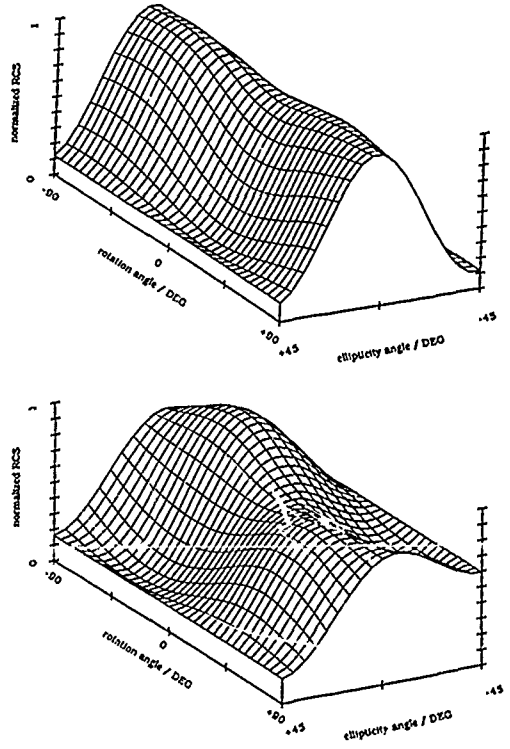


Fig. 7 Polarization Signatures (f=17 GHz); rough surface (top); rough surface with 7 cylinders (bottom)

The XPOL-nulls for the rough surface are distributed along the equator (linear polarization). So no significant rotation angle dependency for this target is observed. For the combined measurement, XPOL-nulls concentrate on a certain area of the sphere. Here the single cylinder behaviour dominates the scattering behavior.

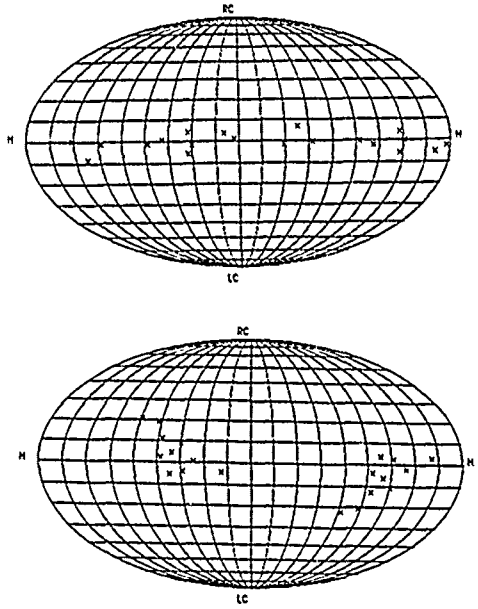


Fig. 8 XPOL-null pairs f=5 GHz; (rough surface top; rough surface plus cylinders bottom)

These signatures are highly frequency sensitive. But by calculating the frequency dependent statistical parameters of the CPS it is well suited to find separation criteria.

5. CONCLUSION

Out of the results in the previous section three main criteria for the detection of the cylinders on the rough surface can be derived:

1. amplitude-ratio signatures
2. polarization signature
3. statistical distribution of the optimum polarization states for different measurement samples and different objects

The amplitude ratio signatures give hints on the frequencies where other signatures should be evaluated e.g. CPS. It is also a measure for the symmetry of a target.
The polarization signature describes the complete polarimetric characteristic of a target for one frequency. By comparison with signatures from well known targets the target scattering mechanism can be identified.
The CPS, which represents points of the polarization signature for special polarization states (maxima or minima in the 3D-diagrams), give nearly the same information as the polarization signature. But this 2D-representation is easier to handle for further statistical treatment.
From the resulting information a data base will be built up for further analysis used together with a classification algorithm (solution of a multiclass problem).
All the above signatures are highly frequency sensitive. By using special frequencies (optimum frequencies for detection of a special target), it is possible to optimize classification tasks.
Further targets and theoretical results will be investigated in future. They will be analyzed in the same way as described in this paper. Additionally other types of signatures will be analyzed.

Acknowledgement : This work was supported by the European Communities Joint Research Center (JRC), Ispra/Italy under contract no. 3314-87-12 ED ISP D.

6. REFERENCES

[1] D. Kähny , S. Riegger, W. Wiesbeck; "Basic Polarimetric Measurements on Monostatic or Bistatic Images," *NATO ASI-series*, Dr. Reidel Publ. Comp., 1989 (in press)

[2] Boerner W.-M. et. al.; "Interpretation of the Polarimetric CO-polarization Phase Term in Radar Images Obtained with The JPL Airborne L-Band SAR System"; *IEEE Trans. on Geosci. a. Remote Sensing: GE-25*, pp. 77-82, Jan. 1987.

[3] Zebker H. A. ; Van Zyl J. J. ; Held D. N. ; "Imaging Radar Polarimetry From Wave Synthesis"; *Journal of Geophysical Research, Vol. 92, No. B1*, pp. 683-701, Jan. 1987.

[4] Agrawal A.P.; Boerner W.-M.; "Redevelopment of Kennaugh's Target Characteristic Polarization State Theory Using the Polarization Transformation Ratio Formalism for the Coherent Case"; *IEEE Trans. on Geosci. a. Remote Sensing: GRS-27*, pp. 2-14, Jan. 1989.

[5] Waniliek G.; Stock D.J.R.;"Use of Polarimetric Information in CFAR- and Classification-Algorithms"; *Proc. of the Int. Conf. on RADAR* , Versailles, France, April 1989 (in press)

POLARIMETRIC MATCHED FILTER FOR COHERENT IMAGING

Alexander B. Kostinski, Brian D. James and Wolfgang-M. Boerner

University of Illinois at Chicago
UIC-EECS/CL, M/C 154
840 W. Taylor St., SEL-4210
Chicago, IL 60680-5480
Tel: +1(312)996-5480
Fax: +1(312)413-0024

In this paper we focus on image-contrast optimization between two rough-surface classes. Our approach is based strictly on polarimetric filtering, and therefore, no digital image-processing techniques are employed. The approach is tested on a complete polarimetric synthetic aperture radar (POL-SAR) image of the San Francisco Bay area. The data have been taken with the National Acronautics and Space Administration - Jet Propulsion Laboratory CV-990 L-band POL-SAR system, where eight real numbers (complex elements of a 2 x 2 polarization scattering matrix) are associated with each image pixel. Optimal transmitted polarizations (corresponding to maxima or minima of reflected energy) are found for each image pixel, and the results are analyzed statistically via a set of joint two-dimensional histograms. This is done for both of the rough-surface classes. The image response to the "optimal" incident polarization is then simulated digitally by adjusting the receiver polarization according to the modes of the histograms. The corresponding images are computed and displayed with significant image-contrast improvement.

DATA MANAGEMENT IN THE FIFE INFORMATION SYSTEM

Donald. E. Strebel[1], Jeffery. A. Newcomer[2], James. P. Ormsby[3],
Forrest. G. Hall[3], Piers. J. Sellers[4]

1) VERSAR, Inc., 9200 Rumsey Rd., Columbia, MD 21045
2) ST Systems Corp., Lanham, MD
3) NASA Goddard Space Flight Center, Greenbelt, MD
4) COLA, Dept. Meteorology, Univ. Md, College Park, MD

ABSTRACT

The First ISLSCP Field Experiment (FIFE) was conducted during 1987 on a 15x15 km area in central Kansas, including the Konza Prairie Natural Research Area. The experiment goal of understanding and modeling the effects of vegetation on the heat and mass fluxes in the atmosphere/land surface boundary layer required a flexible and responsive information system with georeferenced data to permit the necessary coordinated analyses. The FIFE Information System (FIS) met this goal in a variety of ways, depending on the type of data and the type of scientific investigation involved.

The FIS has served the FIFE investigators as a tool for experiment design, as well as for organizing and manipulating the complex data set. Two specific requirements were rapid and simple access to specialized user specified data sets and quick evolution as the science needs changed. As a result, the FIS differs in several important ways from "classical" data systems, covering the spectrum of conceptual, managerial, and organizational issues.

Our experience indicates that the design, development, and operation of an information system supporting such an experiment must be flexible and under direct day-to-day control of scientist/users. This experience has also led to a number of observations (both technological and organizational) for designing and building both future experiment support data bases and long term data systems supporting scientific research.

KEY WORDS: FIFE, INFORMATION SYSTEMS, ISLSCP

FIFE

The goal of the International Satellite Land Surface Climatology Project (ISLSCP) is to develop techniques to determine surface climatology from satellite acquired data. The First ISLSCP Field Experiment (FIFE) was designed to collect a coordinated data set useful for developing and validating the models this goal requires. Particular attention was directed toward the mass and energy balances at the land surface/atmosphere boundary, including the role of biology in controlling the interactions there. The experiment was also to explore the use of satellite observations to infer climatologically significant land-surface parameters related to these interactions. For further details, see Sellers, et al. (1988).

An integral part of the FIFE Science Plan (Schmugge and Sellers, 1986) was a data system which could serve the FIFE investigators as a tool for experiment design as well as for organizing and manipulating the complex data set during and after the data collection effort. A dedicated, remotely accessible data system was set up at NASA's Goddard Space Flight Center (GSFC) to meet this requirement. In the following sections we describe the general features of the system, what kinds and amounts of data are involved, and how the data are organized and used. In addition, we discuss some general data system management issues and how the FIS experience may bear on future data system efforts.

FIS

The decision to develop a dedicated FIFE Information System (FIS) was made in October 1986 after a thorough review of available alternatives (e.g. NASA's Pilot Land Data System, Wharton and Newcomer, 1988). The three objectives of FIS were to: a) capture and preserve the FIFE data, b) distribute the data to the investigators as rapidly as possible (preferably by interactive remote access), and c) function as an experiment design and analysis tool. These objectives were driven by experiment deadlines, which implied a responsiveness and flexibility not available using standard development methodologies. Existing technology (hardware and software) was used, and the effort was organized in a way that could tolerate the unavoidable changes in priorities, configurations, and designs

which one expects from an active experiment.

FIS is an active exchange mechanism for the cooperative analysis of data. The first data submission typically occurs with the data in a very preliminary state, perhaps simply converted to physical units with no checking. In this framework, there are three general processing stages for data submitted to FIS.

In the first stage, a data set is inventoried and archived, processed as necessary into FIFE standard formats (e.g. tapes or online tables), and associated with preliminary documentation. At this stage, FIS personnel try to establish an active scientific communication with the investigator, which not only educates both, but creates important lines of communication for serving users.

The second stage in the life of the data base is vigorous use, analysis, revision, and update of the data base information. During this stage, a particularly important feature of the structure of the FIS online data base is the quality code and revision date tag which are included with each data record.

In the third stage, the data system receives "derived" and refined "key" data sets for final archiving. This activity has not yet started for FIS, but it is anticipated that it will be a major function for the final two analysis years of FIFE. The integrated FIFE data set, it appears, will be useful to explore many scientific questions which are beyond the scope and resources of the FIFE project itself.

DATA TYPES, VOLUMES, STANDARDS

The volume of the digital data has exceeded 90 gigabytes and will probably approach 100 gigabytes as the data processing and reduction continue. Most of these data are the "raw" (Level 0) imagery, as indicated in Table 1. A reduction factor of about 20 is achieved as the imagery is screened for suitability for further analysis during processing to Level 1 (reformatted, with location information added) or Level 2 (geometrically and/or radiometrically corrected). In addition, there are over 70 online tables of data, inventory, and reference information, occupying approximately 150 megabytes of disk storage. An archive of analog data, primarily maps, photography, and video tapes, is also part of the data system.

It is probably obvious that the combination of data set diversity and volume could only be handled by establishing some input and output

TABLE 1
IMAGE DATA IN FIS

TYPE	NUMB	THOUSAND MBYTES
ASAS	-	-
GAC	1765	2.210
GOES	18269	0.018
LAC	1501	64.701
NS001	592	12.874
PBMR	-	-
TIMS	614	7.388
TM	16	1.336
SPOT/XS	37	0.953
SPOT/PAN	3	0.103
TOT		89.583

standards. At the same time, we faced the constraint of avoiding a rigidity which would discourage submission of data to the system or act as a barrier to access by potential users. Thus, we looked for common or near-universal capabilities in defining working standards in three areas: formats, measurement units, and coordinate systems.

ACCESSING THE DATA: COORDINATE REFERENCING

The time and three-dimensional spatial location of each measurement is crucial to the derivation of the energy and mass fluxes through the FIFE site. Thus, the data system was required to provide coordinate information with all data sets. In the FIS, this information is associated with each data set as it is assimilated into the system. The methods range from storing explicit values for point source data to appending georeference files as additional layers to image products. The method used depends largely on the physical dimensionality of the measurement technique (Table 2).

For point (0-D) data, a reference data base table is used to label each of the FIFE sites with an identifier and store its geographic position, elevation, slope, and aspect. Linear (1-D) data sets are inventoried with a starting and ending

TABLE 2
DATA DIMENSION VS REFERENCE METHOD

DIMENSION	EXAMPLE	METHOD
0	SOIL MOISTURE AT A GIVEN POINT	SITE ID #
1	TRACE GASES OVER TRANSECT	PAIRED SITE #'S
2	MULTISPECTRAL IMAGERY	LAT,LONG FILES
3	AIRCRAFT FLUX MEASURMENTS	X,Y,Z COORD

geographic location for the total transect and the postion (distance) of each actual measurement along the transect is kept in the data records with the variables measured. For area (2-D) data such as images, minimum and maximum values of latitude and longitude are stored in inventory tables along with platform heading, altitude, and ranges of view and solar angles. When images are processed, point by point latitude and longitutde information is calculated and appended as separate files to the data tape. Some volume (3-D) data have been included in the data base by including an altitude and latitude and longitude coordinates for the beginning and end of the flight line. An alternate method may be considered for the derived flux data sets, e.g. listing the coordinates of three corners of the rectangular volume being sampled.

UTILIZATION

A "rapid prototype" phase of the FIS provided critical support for the two experiment design workshops, held in December, 1986 and March, 1987. Pre-field season monitoring started in January, 1987 with the acquisition of satellite images and conventional weather data. Routine FIS processing and data distribution began soon thereafter, with emphasis on products (e.g. GIS data) useful in experiment planning. Between IFC's, some specific data sets were given rapid turn-around to provide "quick look" analyses of field data.

Tne online data base was opened for remote access by the beginning of the October 1987 field campaign, just under one year from the decision to build FIS. User logins climbed rapidly during the Fall of 1987 to well over 100 per month, and this level has been maintained through 1988 and into 1989. During this period the primary use of the online porition of FIS has been to obtain supporting data sets interactively and to obtain information about processed image tapes, which are then distributed by mail. The interactive usage shows increasingly intense periods. Use of the analog data and documentation in the archive located at GSFC is also becoming more important and frequent.

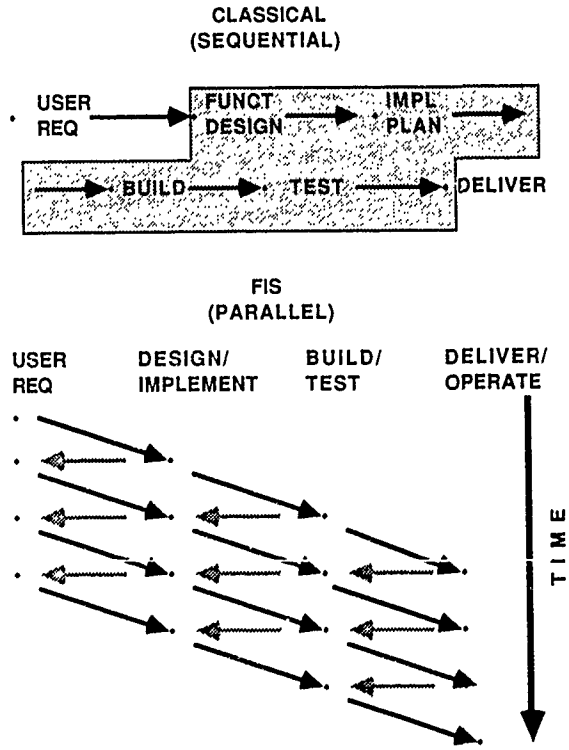
The integrated analysis and modeling phase of the experiment is just beginning. In addition, plans are being completed to open FIS for access by investigators outside of the FIFE project. Both of these factors are expected to subject the system to even more vigorous usage.

ORGANIZATION & MANAGEMENT

The objective of supporting an active field experiment imposed requirements to provide rapid and simple access to specialized user-specfied data sets and to evolve quickly as the science needs changed or new instruments or procedures were developed. This led us to apply concepts from complex system theory instead of using the standard software project management approaches. In this view, the information system is an integrated "organism" composed of people, hardware, and software which must adaptively respond to a fluctuating mix of input data and output demands.

Developing such a system required an approach different from the traditional sequential method of systems development (see Figure 1). FIS was built in an applications environment with constant review and input from practicing scientists,

Fig. 1 In a classical approach, users are largely excluded from boxed processes. In the FIS/complex systems approach users are involved at each step (gray arrows), in an ongoing iterative process. This can tolerate relatively vague, difficult to specify, or changing requirements, and transition naturally from development to operations.



including those doing the field experiment. As the realities of data collection modified the investigators' requirements and data, the data system contents and development priorities were restructured accordingly. It should be emphasized that this does not represent poor planning - experimental science is a dynamic, exploratory process. A data system which has an active (vs. archival) role must necessarily be equally dynamic, which the sequential development model cannot provide.

DATA SYSTEMS FOR FUTURE EXPERIMENTS

The successes of FIFE are already suggesting new experiments. A likely future experiment would involve two or more FIFE-type sites, instruments with increased spectral resolution (i.e. data volume), and additional groups of investigators representing new disciplines (e.g. tropospheric chemistry). A data volume of 5-10 times FIFE's could be anticipated for such an experiment in the early 1990's, i.e. 500 to 1000 Gigabytes. In addition, the experiment would require more extensive geographical data, a large initial image base for planning, more decentralized data base facilities, and more active field support.

The FIS experience points the way to handling the increased data volume and integration problems. The technological capacity to meet this demand is now available "off the shelf". We conceive of investigators with high powered workstations supporting a suite of interactive data base tools like those developed for FIFE. The PI workstations then become individual elements of the complete data system, as distinct from a "window" to a massive central facility. In addition, similar fixed and mobile workstations would be available in the field. A central experiment-specific database/archive would serve as the focus for the cooperative organization, distribution, and quality assurance of the data, as well as be the home base for a scientifically and technically well versed data system development/user support organization.

The most critical feature, though, will be the managerial organization of the effort as discussed above. It appears that the design, development, and operation of an information system supporting such an experiment must be flexible and under direct day-to-day control of scientist/users. This is necessary to incorporate diverse data types in a systematic way as they become available, to add scientific rigor by identifying data gaps, and to provide real-time quality assurance, and to incorporate (and change as understanding evolves) scientifically meaningful organization into the database tables and user interface. Equally important, this management approach emphasizes that the data system and its coordinator (the experiment "Information Scientist") must understand and respond rapidly to the evolving requirements characteristic of active scientific research.

CONCLUSION

The FIS has successfully completed the inital data absorption task and is in the midst of restructuring and revision. The next requirement is for public access to the data and a permanent archive. A policy for open access by interested investigators will be put in place during 1989. It appears that archiving can also be handled with an extention of the procedures already being used. An archive could take many forms, including a central facility like NASA's Pilot Land Data System, a long term archive at NSF's Konza Prairie LTER, or even distribution of the data archive on CD-ROM or other media by a private organization.

We expect future experiments to require much the same approach as was taken by FIS, up to and during the Eos era (NASA, 1986). That is, such active experiment support data system efforts must be accommodated separately from the centralized activities of EosDIS (Dutton, 1989, Chase, 1989).

ACKNOWLEDGEMENTS

This work was partially performed under contract to the National Aeronautics and Space Administration by employees of VERSAR, Inc., STX, SAR, RMS, and SSAI.

REFERENCES

1. Chase, R. R. P.,"Toward a Complete Eos Data Information System", IEEE Trans. on Geosci. & Rem. Sens., 27:2, pp. 125-131, 1989.

2. Dutton, J. A.,"The Eos Data and Information System. Concepts for Design", IEEE Trans. on Geosci. & Rem. Sens., 27:2, pp. 109-116, 1989.

3. NASA,"From Pattern to Process: The Strategy of the Earth Observing System", Eos Science Steering Committee Report, National Aeronautics and Space Administration, Vol. II, 1986.

4. Schmugge, T., and P. J. Sellers, "FIFE Science Plan", NASA Technical Report, 1986.

5. Sellers, P. J., F. G. Hall, G. Asrar, D. E. Strebel, and R. E. Murphy, "The First ISLSCP Field Experiment (FIFE)", Bull. Am. Met. Soc., 69:1, pp. 22-27, 1988.

6. Wharton, S. W. and J. A. Newcomer,"Requirements for Ongoing Development of the Pilot Land Data System (PLDS)", Proc. IGARSS'88, vol. 1, pp.85-88, 1988.

INCORPORATION OF POLARIMETRIC RADAR IMAGES INTO MULTISENSOR DATA SETS

D.L. Evans, J.J. van Zyl, M. Kobrick, C.F. Burnette, S.D. Wall
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109
(818) 354-2111
TELEX 67-5429

Many surface process studies require the ability to characterize the physical nature (e.g. morphology, roughness, dielectric properties) as well as the composition of the Earth's surface. The complement of instruments currently planned for the future, for example, the Shuttle Imaging Radar (SIR-C), the Earth Observing System (Eos) Synthetic Aperture Radar (SAR), High-Resolution Imaging Spectrometer (HIRIS), Japanese Intermediate Thermal Infrared Radiometer (ITIR), and the proposed Thermal Infrared Ground Emission Radiometer (TIGER) will provide the basic measurements required for this sort of characterization over large areas. Airborne sensors that are currently operational provide end-to-end prototypes for these instruments. Thus, analysis tools developed for aircraft data sets can be applied to future data sets as well.

A Geologic Remote Sensing Field Experiment (GRSFE) is currently planned for July 1989. Current plans for this experiment are to acquire coordinated Airborne Visible and Infrared Imaging Spectrometer (AVIRIS), Thermal Infrared Multispectral Scanner (TIMS), and NASA/JPL multipolarization aircraft SAR data, with associated field measurements. These data will be archived on a CDROM and distributed to the scientific community. GRSFE will concentrate on sites with minimal to no vegetation cover and flat to medium relief. Sites with increasing complexity will be selected for later experiments based on results of GRSFE.

To prepare for GRSFE, over two dozen scientists from Washington University, Arizona State University and JPL participated in a coordinated aircraft deployment and field measurement campaign in the Mojave Desert from May 29-June 3, 1988. Diurnal thermal data were collected with TIMS and L-, C-, and P-band polarimetric aircraft SAR data were acquired at three incidence angles over Pisgah Crater, which is approximately 80 km east of Barstow, CA. Five "modeling sites" were selected which ranged in roughness from 2 to 70 cm r.m.s. Field measurements included meteorological measurements, surface and subsurface temperature measurements over the diurnal cycle, soil complex dielectric constant, and vegetation density. Stereo photographs were acquired with a helicopter-mounted camera system for derivation of microtopographic profiles of surface roughness. In addition, samples were collected for laboratory XRD analysis, soil moisture, and whole-rock dielectric constant measurements.

Some of the data acquired during the Mojave Field Experiment are being used to verify radar backscatter models of surface roughness and dielectric constant; determine the multisensor signatures of desert surfaces of differing ages; and evaluate the effects of surface modification processes on these signatures. We will describe the data analysis technique development associated with these activities, emphasizing registration of imaging radar polarimeter data sets, and multisensor image analysis. It should be noted that while this investigation focuses on geologic process studies, the algorithms and software are of general use in analysis of aircraft or spaceborne SAR data.

FUZZY LOGIC AND KNOWLEDGE-BASED GIS: A PROSPECTUS

Yee Leung

Department of Geography
The Chinese University of Hong Kong
Shatin, Hong Kong

Abstract. Except for some experimental cases, the design of present-day GIS is based on Boolean logic, a rigid two-valued mathematical system which gives no room for imprecision in information and human thought processes. Though they seem to be quite satisfactory in performing data acquisition, input, storage, retrieval, analysis, and output, they are far from a desirable system with human intelligence and expertise. It is argued in this paper that fuzzy logic, in addition to providing a more adequate logical system, appears to be instrumental to the design of knowledge-based GIS. It serves as a basis for data manipulation, knowledge representation, and inference with geographic databases.

Keywords. fuzzy logic, fuzzy queries, knowledge representation, inference, expert systems

I. INTRODUCTION

Present-day GIS is a georeference system for the acquisition, storage, retrieval, manipulation, analysis, and display of data. Though they appear to be quite satisfactory in performing most of the above functions, they are still just a relatively sophisticated data processing system with conceptual shortcomings and very limited human intelligence and experience. Automation of data input, retrieval, analysis, and output is still primitive in terms of capturing the complicated and imprecise real world and replicating human wisdom and expertise.

At present, human intelligence has not been effectively integrated into data input, retrieval, analysis, and output. These operations are usually driven by menus or procedures which leave very little room for flexibility and automation of users' expertise. Available GIS do not provide facilities to aid users to design their own decisionmaking systems in geoprocessing. These systems are too mechanical to have any values in complex decisionmaking environment.

With respect to logical foundations, present-day GIS are predominantly based on Boolean logic. It is especially so in data retrieval and overlay operations. The all-or-nothing mathematical system of Boolean logic has however been demonstrated to be inadequate for spatial analysis and planning in general (Leung 1988) as well as spatial classification and boundary delimitations in particular (Leung 1984, 1985, 1987). It is well-known in the GIS community that overlay by Boolean logic generally ends up with loss and/or misrepresentation of information. It forces artificial precision on our otherwise imprecise information and thought processes. It arbitrarily screens out intrinsic imprecision in our database which might be essential in our analysis. It fails to communicate the extent of imprecision or error to users. Therefore, results of overlay unravel more questions about data quality and boundary mismatch than they solve (McAlpine and Cook 1971, Burrough 1986). Similar problems also exist in other data analysis and cartographic modeling exercises. It is thus imperative to have a more appropriate logical system for the design of GIS. Such a system should be able to handle and to communicate precise and imprecise information. It should also be instrumental in the design of intelligent GIS.

A basic requirement of a knowledge-based GIS is the ability to process natural languages in an appropriate manner. Information and knowledge representations, inferences, input and output controls, and interfaces all involve the use of natural languages which are intrinsically imprecise. Fuzzy sets theory in general and fuzzy logic in particular are instrumental to process natural language. Inspite of this necessity, only a handful of knowledge-based GIS incorporates some limited sort of capabilities in handling imprecision of information and knowledge. The use of fuzzy sets in database design (Robinson and Strahler 1984), representation of fuzzy spatial concepts, data, and relationships (Robinson 1984, Robinove 1986), query of spatial information (Robinson *et al.* 1985), and cartography (Bouille 1982, Muller 1987) have been proposed. All these studies however do not have a solid grounding of the mathematical and logical structures necessary for such a line of development.

The purpose of this paper is to outline in brief an on-going project on the fuzzy-logic foundation for constructing knowledge-based GIS (Leung 1989a, b, c, d, e). Emphasis is placed on general principles rather than specifics in this paper.

II. REPRESENTATION OF KNOWLEDGE

The mathematical representation of fuzzy propositions is instrumental to formulate formal procedures for constructing queries and making inferences in knowledge-based GIS. To make it operational, we also have to represent it in terms of a database (Leung 1989b, c).

In general, any fuzzy proposition can be represented in terms of a relation in a relational database

$$R(A_1, \ldots, A_n) \tag{1}$$

where R is a relation and A_i is an attribute (variable) taking values in U_i, i = 1,..., n.

In a GIS for climatic classification, we can have the following relations in the database:

$$\text{PIXEL(NAME, PRECIPITATION)} \tag{2}$$

$$\text{ADEQUATE(NAME, } \mu) \tag{3}$$

where μ in (3) is the grade of membership of a pixel having a specific precipitation u computed by a membership function defining *adequate*:

$$\mu_{adequate}(u) = \begin{cases} 0, & \text{if } u \le 500 \text{ m}, \\ \dfrac{u - 500}{400}, & \text{if } 500 \text{ mm} < u < 900 \text{ mm}, \\ 1, & \text{if } 900 \text{ mm} \le u \le 1100 \text{ mm}, \\ \dfrac{1500 - u}{400}, & \text{if } 1100 \text{ mm} < u < 1500 \text{ mm}, \\ 0, & \text{if } u \ge 1500 \text{ mm}. \end{cases} \tag{4}$$

Since, a fuzzy proposition can always be translated into a possibility distribution which in turn can be expressed as a relation in a relational database, propositions about spatial relationships such as *long* distance, *approximately* equilateral triangle, a *somewhat* north direction, and adjacent places can be regarded as relations with the following respective frames:

| LONG | Distance | μ |

| APPROXIMATELY EQUILATERAL TRIANGE | A | B | C | μ |

| SOMEWHAT NORTH | Bearing | μ |

| ADJACENT | Name 1 | Name 2 | μ |

Without loss of generality, let $X = (X_1, \ldots, X_n)$ be a n-ary variable in which X_i takes values in U_i, i = 1, ..., n. Then the fuzzy proposition

$$\text{p: X is A,} \tag{5}$$

where A is a fuzzy n-ary relation in the Cartesian product $U = U_1 \times \ldots \times U_n$, can be translated into

$$\Pi_{(X_1, \ldots, X_n)} = A \tag{6}$$

such that

$$\begin{aligned} &\text{Poss } (X_1 = u_1, \ldots, X_n = u_n) \\ &= \pi_{(X_1, \ldots, X_n)}(u_1, \ldots, u_n) \\ &= \mu_A(u_1, \ldots, u_n). \end{aligned} \tag{7}$$

In identifying climatic classes, we can for example have a proposition

$$\text{p: X is } \textit{subtropical}. \tag{8}$$

The term "*subtropical*" can be a fuzzy binary relation such that

$$\textit{subtropical} = \textit{warm} \times \textit{adequate precipitation} \tag{9}$$

III. FUZZY QUERIES

Adequate representation of knowledge is instrumental to do overlay, data retrieval, and inference. A mandatory feature of a knowledge-based GIS is the ability to entertain fuzzy queries and return relevant information to users. To shed some light on this matter, the basic structure of a fuzzy query is examined in the following.

To be able to represent a query, we need the concept of a relational formula (Leung 1989b)

$$R(t_1, \ldots, t_n) \tag{10}$$

where R is a relation and t_i, i = 1, ..., n, is a tagged term indicating the condition of the attribute A_i. For example, t_i can be of the following forms:

(a) A_i = variable name,

(b) A_i = numerical or linguistic value,

(c) $A_i \neq (\ge, \le, >, <)$ a numerical value.

A fuzzy query in general can be expressed as

$$\begin{aligned} &R_1 + R_2 + \ldots + R_n + F_1 + \ldots + \\ &F_r + L_1 + \ldots + L_q \end{aligned} \tag{11}$$

where R_i, i = 1, ..., n, is a relational formula; F_j, j = 1, ..., r, is a fuzzy formula, L_k; k = 1, ..., q, is a precise formula (a formula with precise built-in predicates); and + means the union.

Therefore, a fuzzy query consists of imprecise and precise conditions. Precise queries are only special cases. A fuzzy query is successfully processed if we can find and arrange all spatial units, e.g. pixels or polygons, which satisfy the relational, fuzzy, and precise formulas in a graded manner.

In the example of climatic classification, the query

> "List the precipitation and temperature of all pixels with *adequate* precipitation and a temperature greater than or equal to 23°C" (12)

can be expressed as

> PIXEL(Name = x, Precipitation = u, Temperature = v) +
> ADEQUATE(x) +
> v ≥ 23°C (13)

where PIXEL(Name = x, Precipitation = u, Temperature = v) is a relational formula; ADEQUATE(x) is a fuzzy formula; and v ≥ 23°C is a precise formula.

Under this situation, the system will find all pixels which satisfy the relational formula PIXEL(Name = x, Precipitation = u, Temperature = v) from the frame

| PIXEL | Name | Precipitation | Temperature |.

It then finds all pixels whose temperature is above 23°C and then search and arrange in order all these pixels from the frame

| ADEQUATE | Precipitation | μ |

by their degree of belonging, μ, to *adequate*.

IV. FUZZY INFERENCE

Fuzzy logic, similar to standard logic, depends on some rules of inference to deduce conclusions from a set of premises. In standard logic, the antecedent in the minor premise has to confirm precisely, or match exactly, that of the major premise before the consequence of the major premise can be regarded as the conclusion. This however is not flexible and realistic enough to model human inference in which the antecedents of the major and minor premises might approximately be equal. Under this situation, the best we can say is that the conclusion could only be approximately equal to the consequence of the major premise.

In knowledge-based GIS, the following are typical examples:

Major premise: If temperature is *high*, then the pressure is low. (14)

Minor premise: The temperature in pixel X is quite *high*. (15)

Approximate conclusion: The pressure in pixel X should be *quite low*. (16)

Here, the conclusion in (16) would not be possible under standard logic because *high* ≠ *quite high* and *low* ≠ *quite low*. However, such a reasoning process agrees with our intuition.

Unlike standard logic, FL allows the existence of fuzzy predicates in both the premises and the conclusion. We can deduce through FL possibly imprecise conclusions from a set of imprecise premises. Given a major premise, we can theoretically deduce, from minor premises with approximately equal antecedents, conclusions which are approximately equal to that of the major premise. This is a natural, flexible, and efficient way to manage inferences in knowledge-based GIS.

Among the rules of inference in FL (Leung 1989d), generalized Modus Ponens and Modus Tollens are useful schemata for making inference in GIS. The inferential procedure of the former can be expressed as

$$\begin{array}{l} p \longrightarrow q\text{: If X is A then Y is B} \\ \quad p_1\text{: X is } A_1 \\ \hline q_1\text{: Y is } B_1 \end{array} \qquad (17)$$

In semantic form, (17) becomes

$$\begin{array}{l} p \longrightarrow q \dashrightarrow \Pi_{(Y|X)} = R(A, B) \\ p_1 \longrightarrow \Pi_X^{p_1} = A_1 \\ \hline q_1 \longleftarrow \Pi_Y^{q_1} = A_1 \circ R(A, B) = B_1 \end{array} \qquad (18)$$

where X, Y, and Z take values in $\underline{U}$, $\underline{V}$, and $\underline{W}$ respectively; $\Pi_{(Y|X)}$ is a conditional possibility distribution equated with the fuzzy relation R(A, B) induced by the implication $p \longrightarrow q$.

In FL, a generalization of Modus Ponens is made possible by permitting the existence of A_1 close to A and B_1 close to B. For example,

$$\begin{array}{l} p \longrightarrow q\text{: If temperature(X) is } \textit{high}(A) \text{ then pressure(V) is } \textit{low}(B) \\ \quad p_1\text{: Temperature(X) is } \textit{very high}(A_1) \\ \hline q_1\text{: Pressure(V) is } \textit{very low}(B_1) \end{array} \qquad (19)$$

Similarly, generalized Modus Tollens can be expressed as

$$\begin{array}{l} p \longrightarrow q\text{: If X is A then Y is B} \\ \quad p_1\text{: Y is } \neg B_1 \\ \hline q_1\text{: X is } \neg A_1 \end{array} \qquad (20)$$

or in semantic form

$$\begin{array}{l} p \longrightarrow q \longrightarrow \Pi_{(Y|X)} = R(A, B) \\ p_1 \longrightarrow \Pi_Y^{p_1} = \neg B_1 \\ \hline q_1 \longleftarrow \Pi_X^{q_1} = R(A, B) \circ \neg B_1 = \neg A_1 \end{array} \qquad (21)$$

To be able to make inference which agrees with commonsense, acquisition of knowledge from users in general and experts in particular are of utmost importance. Various methods for the acquisition of fuzzy terms and rules for knowledge-based GIS are discussed in Leung (1989e) and are not elaborated here.

V. CONCLUSION

To recapitulate, existing GIS are based on Boolean logic which is inappropriate to handle imprecision. They are not intelligent geoprocessing systems with human knowledge and expertise. Their input, storage, query, analysis, and output functions are rudimentary and should be enhanced through artificial intelligence techniques in general and expert systems technologies in particular. GIS should be looked at as a basic subsystem within the general framework of a spatial decision support system which consists of a GIS, an expert system shell, and an analytical module. A natural-language interface should be constructed to integrate the three basic subsystems. The expert system serves as the brain of the whole system. It handles knowledge-bases and inference procedures. It directs the flows of data in and out of the GIS and provides rule-based inferences. The analytical-model subsystem is called upon if problems need to be solved by algorithms or existing spatial models. The GIS performs data input, storage, retrieval, and output functions. In most cases, spatial queries and analyses are handled by the expert system and analytical model subsystem. Though each subsystem has its own architecture, they however have mutual feedbacks in terms of an integrated design throughout the development process.

To be effective and viable, the design of any GIS for the future should probably follow this line of thinking. Geographic databases without intelligence has a very small chance to meet challenges of complex decisionmaking processes. Knowledge-based GIS is a must for a sustained growth in the field.

References

Bouille, F., "Actual tools for cartography today", Cartographica, Vol. 19, pp. 27-32, 1982.

Burrough, P.A., Principles of geographical information systems for land resources assessment, Oxford: Clarendon, 1986, Chap. 6.

Leung, Y., "Towards a flexible framework for regionalization", Environment and Planning A, Vol. 16, pp. 1613-1632, 1984.

Leung, Y., "A linguistically-based regional classification system", in P. Nijkamp, H. Leitner, and N. Wrigley (eds.), Measuring the Unmeasurable, Dordrecht: Martinus Nijhoff, pp. 451-486, 1985.

Leung, Y., "On the imprecision of boundaries", Geographical Analysis, Vol. 19, pp. 125-151, 1987.

Leung, Y., Spatial analysis and planning under imprecision, Amsterdam: North-Holland, 1988.

Leung, Y., "Fuzzy logic and knowledge-based geographic information systems: 1, a critique and a proposal for intelligent GIS", occasional paper No. 98, Hong Kong: Department of Geography, The Chinese University of Hong Kong, 1989a.

Leung, Y., "Fuzzy logic and knowledge-based geographic information systems: 2, knowledge representation and measure of confidence", occasional paper No. 99, Hong Kong: Department of Geography, The Chinese University of Hong Kong, 1989b.

Leung, Y., "Fuzzy logic and knowledge-based geographic information systems: 3, translation rules and truth-values of fuzzy propositions", occasional paper No. 100, Hong Kong: Department of Geography, The Chinese University of Hong Kong, 1989c.

Leung, Y., "Fuzzy logic and knowledge-based geographic information systems: 4, rules of inference", occasional paper No. 101, Hong Kong: Department of Geography, The Chinese University of Hong Kong, 1989d.

Leung, Y., "Fuzzy logic and knowledge-based geographic information systems: 5, acquisition of knowledge", occasional paper No. 102, Hong Kong: Department of Geography, The Chinese University of Hong Kong, 1989e.

McAlpine, J.R. and B.G. Cook, "Data reliability from map overlay", in Proc., Australian and New Zealand Association for the Advancement of Science, 43rd Congress, 1971.

Muller, J.C., "The concept of error in cartography", Cartographica, Vol. 24, pp. 1-15, 1987.

Robinove, C.J., Principles of logic and the use of digital geographic information systems, U.S. Geological Survey Circular 977, 1986.

Robinson, V.B., "Modeling inexactness in spatial information systems", Proc., Pittsburgh Modeling and Simulation Conference, pp. 157-161, 1984.

Robinson, V.B. and A.H. Strahler, "Issues in designing geographic information systems under conditions of inexactness", Proc., 10th International Symposium Machine Processing of Remotely Sensed Data, pp. 198-204, 1984.

Robinson, V.B., D. Thongs, and M. Blaze, "Machine acquisition and representation of natural language concepts for geographic information retrieval", Proc., Pittsburgh Modeling and Simulation Conference, pp. 161-166, 1985.

CONSIDERATION OF GEOGRAPHICAL INFORMATION SYSTEMS AS TOOLS FOR CLASSIFICATION OF REMOTELY SENSED DATA AND SPATIAL MODELLING OF LANDFORMS

K.O.NIEMANN
Department of Geography
The University of Victoria

L. NIKIFORUK
Terrain Sciences Division,
Alberta Research Council

A. BRIERLEY
Agriculture Canada

ABSTRACT

In recent years there has been an increasing interest in the use of Geographical Information Systems (GIS) to classify remotely sensed data. The primary past GIS function has been to supply digital, spatial data to enhance classifications using remotely sensed data. A second, and increasingly important, GIS function is the classification of remotely sensed data through spatial modelling.

This paper addresses these two issues with reference to two specific projects. The first issue is dealt with through the investigation of the degree of generalization which is inherent in typical cartographic data. The second issue is discussed with reference to parametric classification of landforms through the use of Digital Terrain Models.

1. INTRODUCTION

Geographical Information Systems and associated digital data bases are increasingly being utilized in the processes classifying of remotely sensed data and creating land inventory maps. One data base commonly used for these purposes contains topographic information. The data contained in these coverages are either in the form of classified polygonal maps or digital terrain models (DTM's). This paper deals with an examination of issues related to the use of both of these data forms. A problem which must be considered when using cartographic data coverages as ancillary data bases to enhance classifications, is that of generalization. The degree to which cartographic data bases is examined in this paper. A second issue addressed in this paper is related to the use of DTM's as data sources in landscape evaluation and classification schemes.

2. CARTOGRAPHIC DATA BASES AS ANCILLARY DATA FOR ENHANCEMENT OF CLASSIFICATIONS BASED ON REMOTELY SENSED DATA.

The use of digital data bases to enhance and further refine classifications based on remotely sensed data is becoming widespread One commonly accessible data base is the landform evaluation, or classification, map. This type of mapping incorporates landform shape elements, such as slope and relief, along with geomorphological and pedological variables such as soil texture, soil drainage, and parent material. Map coverages which are based on traditional mapping and cartographic practices are routinely converted from analog to digital format and integrated with remotely sensed data, using either pre or post classification sorting techniques (see for example Loveland and Johnston, 1983; Niemann et. al., 1984; Satterwhite et. al., 1984). The resulting classification is thereby stratified on some element portrayed in the land evaluation or classification map.

The introduction of these data sources into the classification process may introduce a variety of problems. The landform maps portray assemblages of geometric elements, and not individual morphological variables. These elements were defined by Speight (1974) as being '. .areas of land that resemble simple geometric surfaces without inflections, and that are typically described by altitude, slope, aspect, curvature, and a number of derived contextual parameters' (p.213). The range of information contained within the mapped landscape polygons may therefore extend beyond the simple slope, aspect, or curvature values which are desired for integration with the remotely sensed data. A second source of error is one caused by generalization. Simplification and data reduction tend to occur at a number of stages throughout the mapping process. Unless source documents allow for resolutions which are greater than, or equal to, the remotely sensed data, then a generalization in the remotely sensed data must be made to compensate for the lower level of detail.

2.1 COMPARISON OF DTM AND POLYGONAL MAP COVERAGES.

Niemann (1988) investigated the potential for using traditional cartographic products as data sources to enhance classifications based on remotely sensed data. To evaluate the utility of the interpreted polygon map a land classification map, scale 1:50,000 was digitized and the slope gradient values extracted (figure 1a). A 1:50,000 topographic map was also digitized and an elevation model was generated In the case of the interpreted map, the chosen theme consisted of polygons with simple or complex slope gradient designations. The simple class polygons were characterized by a homogeneous narrow slope gradient range. The complex class polygons had a wider range of slope values occurring within their boundaries. The two slope gradient coverages were spatially registered to coincide exactly with each other and digitally overlaid.

The distribution of DTM slopes within the simple-classed interpreted polygons are illustrated in figure 1b. For the lower classed

polygons (classes 2, 3 and 4) the distributions are similar. They are unimodal with the mode at 2% slope for all three classes. There is however an increasing positive skewness as the gradients in the interpreted polygons increase. The distribution changes for the higher slope classes (classes 5 and 6). The ranges of the slopes contained within these polygons are much greater than the lower classed units and the modes less well defined.

The histograms for the distribution of the complex polygons are shown in figure 1c. The distribution of the slope values shown in the histograms are similar to the pattern of the simple classed units. The modes of the lower classes are still low with the distributions positively skewed. The higher classed polygons with greater ranges in slope classes have a corresponding larger range in DTM values.

The comparison of these two map products emphasize the high degree of generalization associated with the interpreted map. This generalization is be the result of three factors. The first factor, as shown by the polygons representing the higher slope gradients, is that a large degree of cartographic generalization occurred during the various stages of map production. This generalization led to an apparent misclassification of areas within the polygons. For example, in the polygons with slope class 6, only 33% of the area falls within the specified range of gradient values. The generalization may be caused by several factors. The first factor is the method of survey, which may not have used slopes as the primary reason for delineating the polygon boundary. Rather, the primary reason for the delineation may have been uniform soil properties, with the slopes of only secondary importance. A second factor contributing to the generalization is that the polygon represents a landform description which zones, is that dense forest cover of varying ages, densities, and species, tend to confuse the interpretation of actual slope breaks, thereby leading to erroneous interpretations.

The rationale for the two slope coverages not coinciding is secondary to the fact that they do not match and that the spatial information content of the polygon map is substantially less than that contained within the DTM coverage. (This point was reinforced by the findings of a previous study by Nix et al. (1984), although the mismatch of their map products was attributed to differences in cartographic scale of the source coverage). The conclusion of this analysis is that, for the purposes of integration with remotely sensed data, the high degree of generalization of the polygonal map would introduce a substantial generalization in the final classified product.

3. USE OF THE DTM FOR LAND EVALUATION AND PARAMETRIC MAPPING.

DTM's are increasingly being used in land evaluation mapping. Klingebiel et. al. (1987) describe how slope gradient, aspect and elevation coverages have been used in the preparation of soil survey maps. They found that DTM map coverages could successfully be integrated into the more traditional soil polygon mapping process. Johnson (1988) describes the potential for using DTM's in the physical land classification and mapping process. In her study, DTM coverages, including slope aspect, gradient, curvatures , as well as hillshading models were created and compared with conventional land classification maps. The result of this comparison was that the problems associated with traditional mapping, such as inconsistency in the interpretation and cartographic process as well as the use of different classification schemes could be avoided (Johnson, 1988). The main shortcoming of the type of approach outlined in the studies by both Klingebiel et. al. (1987) and Johnson (1988), is that there is a tendency to analyze individual landform elements in isolation and not treat the landscape as an integrated unit described by a variety of shape attributes. This approach to classification termed parametric mapping, has been defined '...as the classification and subdivision of land on the basis of selected attributes' (Mitchell, 1973, p.31). The rationale for this approach to terrain analysis being that landform shape attributes are fundamental in influencing geomorhpological and pedological processes (Troeh, 1964; Dalrymple et. al., 1968; Mitchel, 1973; Speight, 1974; Pennock et. al., 1987). Only in a limited number of instances (see for example Franklin, 1987, Niemann, 1988), has there been any attempt to evaluate a variety of DTM products simultaneously within the context of landscape or physical processes.

3.1 THE APPLICATION OF DTM PRODUCTS TO PARAMETRIC AND LAND EVALUATION MAPPING.

The object of this component of the study was to evaluate the possibility of integrating digital terrain models with the processes of parametric mapping and landscape evaluation. A number of problems as outlined by Mitchell (1973) must be recognized prior to the classification of the data. The first problem to be addressed is that of the nature of the geometric variables to be used, and their applicability to the mapping procedure. In an analysis of geometric properties of on which to base their classification. Speight (1974) stated that the basic criteria used to classify the landscape processes included slope, curvature (both across slope and downslope) and the downslope dispersal area. Dalrymple et. al. (1968) defined a land surface model which divided the landscape into nine subunits. Each of these subunits were classified based on the slope gradient, curvature, relief, and the position within the slope catena. The choice of the appropriate properties to incorporate will depend on the nature of the map.

The second problem area identified by Mitchell (1973) was that of choosing the appropriate class boundaries. He stated that the problem of subdivision of variables into specific classes becomes acute when the specifics of the class boundaries are not known. In these situations, the natural breaks in the landscape should be used. This final point is contrary to the procedures followed by the majority of mapping surveys which tend to impose predetermined physiographic classes into their legends. In very few instances do the mapped boundaries reflect the divisions which occur in the study area in question.

With these two problems in mind, it was decided to apply a parametric mapping approach using DTM data to a portion of east central Alberta. A DTM prepared for the 1:20,000 Alberta base mapping project was used. The approach outlined by Zevenbergen and Thorne (1987) was used to derive the spatial derivatives (that is slope gradient, and curvatures). A measure of relief was derived by calculating the standard deviation of the elevation values for a moving kernel. The upslope area contributing to a grid cell was calculated for the entire elevation matrix. Finally the length of the flow path downslope of each grid cell was determined. This value represented flow dispersal as defined by Speight (1974). The flow path and the dispersal variables were described by Speight (1974) as being indicative of the moisture drainage characteristics of a particular area.

The problem of class boundaries was solved through the use of a cluster analysis The data base for the study area was systematically sampled and a k means cluster analysis (SAS,1987) performed. The appropriate number of clusters represented by the data set was determined to be eight, using the method described by Sarle (1983). The descriptive characteristics of the clusters are summarized in Table 1.

	Class 1	Class 2	Class 3	Class 4
Slope (%):	0-3	0-4	1-7	4-8
Relief (m):	0.5-1 5	1-2	1-5	9-15
Curvature (a/d):	-/-	-/-	-/-	+/-
Slope position-Flow:	0-1.5	3-5	7-1	0-1.5
Dis:	0.5-2	1-2	0.5-3.5	0.5-2

	Class 5	Class 6	Class 7	Class 8
Slope(%):	9-15	8-12	1-5	2-9
Relief	6-10	4.5-7.5	1.4-4.5	2-6
Curvature (a/d).	+/-	+/+	+/+	+/+
Slope position-Flow	3-5	12.5-15.	22-26	29-40
Dis:	0.5-1.5	0-3	0.5-3.5	2-12

TABLE 1: Summary Descriptions of the eight classes isolated through cluster analysis.

The cluster summaries for the data can initially, without the benefit of field checking, be interpreted to represent various components in the landscape. Classes 1 and 2 represent areas of low gradient and relief as well as low flow path and dispersion values. These groupings of geometric descriptors are associated with low amplitude hummocky terrain. The low flow path and dispersion values together with the negative across and downslope concavities suggest short slopes. Class 3 represents a wider range in slope gradients but also larger flow path and dispersion values, with curvatures concave in both the across and down slope directions. This would suggest a slightly greater magnitude of hummocky terrain, or a position at the bottom of the hummocks where the moisture would accumulate. These three classes have been classed as hummocky terrain as the curvatures are concave in both of the directions measured. The negative curvatures also suggest that these areas experience a concentration of flow. Classes 4 and 5 represent terrain with a greater slope gradient magnitude (4 to 8% for class 4 and 9 to 15% for class 5), moderate relief, and generally low flow path numbers. The curvatures for both of these classes are flat to slightly convex. These groupings suggest that class 4 represents landscapes which have generally short slopes, but with greater magnitude gradients. This class is prevalent in the areas which are dominated by fluvioglacial landforms (ridges) and deposits and some glacial ridges, and not the hummocky moraine of the common to the first three classes. Class 5 with its greater relief and slope magnitude is interpreted as being characteristic of valley sides and larger ridges and is commonly associated with rough broken terrain. The final three classes are characterized by generally very high flow path values and varying slope gradients. The areas which fall into these classes are along drainage lines: small gullies and channels. The areas represented by class 8 are associated with class 5 terrain while classes 6 and 7 are with the lower sloped zones.

The classes defined through this type of approach to parametric mapping are particularly useful as there are very specific relationships between the class descriptions and landscape forming processes This is particularly evident when comparing the landscape categories defined by Dalrymple et. al. (1968) where the rational for landscape division was based on geomorphic process and the parameters used to classify were geomorphometric. In the present study, the landscape classes, rather than individual landform elements, developed through the use of geomorphometric variables are tied to pedological processes related to moisture movement and degrees of drainage. Given the scale of the landforms in the study area and the resolution of the DTM (original sample spacing 120 metres) it is not felt that a realistic subdivision of the individual slope facets was possible.

4. CONCLUSIONS

This paper has emphasized two aspects of integrating geographic data bases with remotely sensed data. In the first instance the issue of using traditional cartographic data bases as ancillary data bases was addressed. It was demonstrated that a large amount of generalization occurs as a result of the mapping process. This generalization can lead to significant confusion when combined with remotely sensed data through a stratification process.

The second aspect of the paper dealt with the use of DTM's as data sources in land evaluations and landscape mapping. The traditional approach used in morphometric mapping with DTM's has been to integrate the coverages independently. It was shown that through the integration of a number of morhpometric variables that a more complete interpretation of processes is possible.

5. REFERENCES.

Dalrymple, J.B., R.J. Blong, and A.J. Conacher (1968) *A Hypothetical Nine Unit Landsurface Model.* Z. Geomorph. 12, p.60-76

Franklin, S.E. (1987) *Terrain Analysis from Digital Patterns in Geomorphometry and LANDSAT Spectral Response.* PE&RS, 53 , p.59-66

Johnson, V.A. (1988) *Digital Terrain Modelling Applied to the Alberta 1:20,000 Mapping Project.* unpub. M.Sc. thesis U. of Alberta. 96 pages

Klingebiel, A.A., E.H. Horvath, D.G. Moore, and W.U. Reybold (1987) *Use of Slope, Aspect, and Elevation Maps derived from Digital Elevation Model Data in Making Soil Maps.* in Reybold, W.U. andG W. Peterson (eds.) Soil Survey Techniques SSSA Special Publication #20 p.77-90

Loveland,T.R. and G.E. Johnson (1983) *The Role of Remotely Sensed and Other Data in Predictive Modelling: the Umatilla, Oregon Example.* PE&RS 49, p.1183-1192

Mitchell,C.W (1973) Terrain Evaluation Longman Group. 221 pages

Niemann,K.O., and G.J. Langford (1984) *The Use of LANDSAT Imagery in Wildlife Habitat Analysis.* Internal Report Prepared for the Alberta Remote Sensing Center and The Alberta Environmental Centre. 49 pages

Niemann K.O. (1988) *DEM Drainage as Ancillary Data to Enhance Digital LANDSAT Classification Accuracies.* unpub. Ph.D. dissertation 191 pages

Nix,L.E., W.A. Swain, and K.O.Kelton (1984) *Modelling Forest Biomass Accessibility in S.Carolina with Digital Terrain Data.* Proceed. 1984 Symp. on Machine Processing of Remotely Sensed Data, p. 389-394

Pennock, D.J., B.J. Zebarth, and E. DeJong (1987) *Landform Classification and Soil Distribution in Hummocky Terrain, Saskatchewan,* Canada. Geoderma 40, p. 297-315

Sarle, W.S. (1983) *Cubic Clustering Criterion* SAS Technical Report A-108 49 pages.

SAS (1987) SAS/STAT Guide for Personal Computers Version 6 Edition SAS Instutute Inc. 1028 pages.

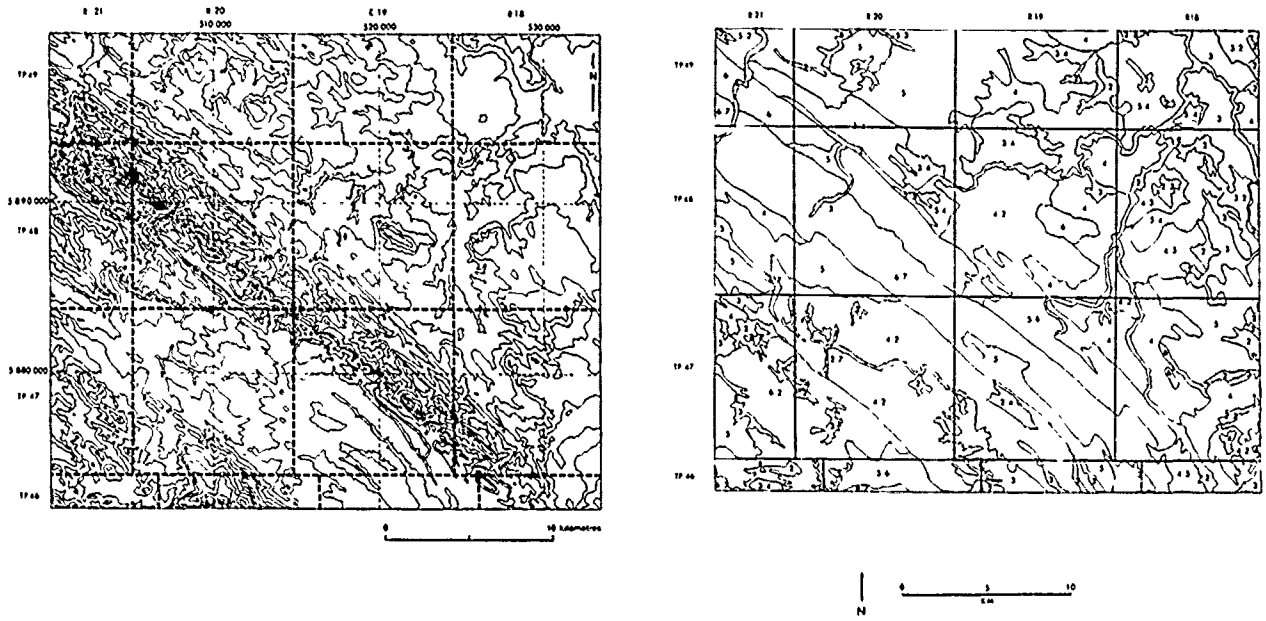
Satterwhite, M., W. Rice, and J. Shipman (1984) *Using Landform and Vegetative Factors to Improve the Interpretation of LANDSAT Imagery.* PE&RS 50, p.83-91

Speight,J.G. (1974) *A Parametric Approach to Landform Regions.* IBG Special Publication #7 p. 231-230

Troeh, F.R. (1964) *Landform Parameters Correlated to Soil Drainage.* Proceed. Soil Science of America 28, p. 801-812

Zevenbregen, L.W. and C.R.Thorne (1987) *Quantitave Analysis of Land Surface Topography.* Earth Surface Processes and Landforms 12, p. 47-56

a



b

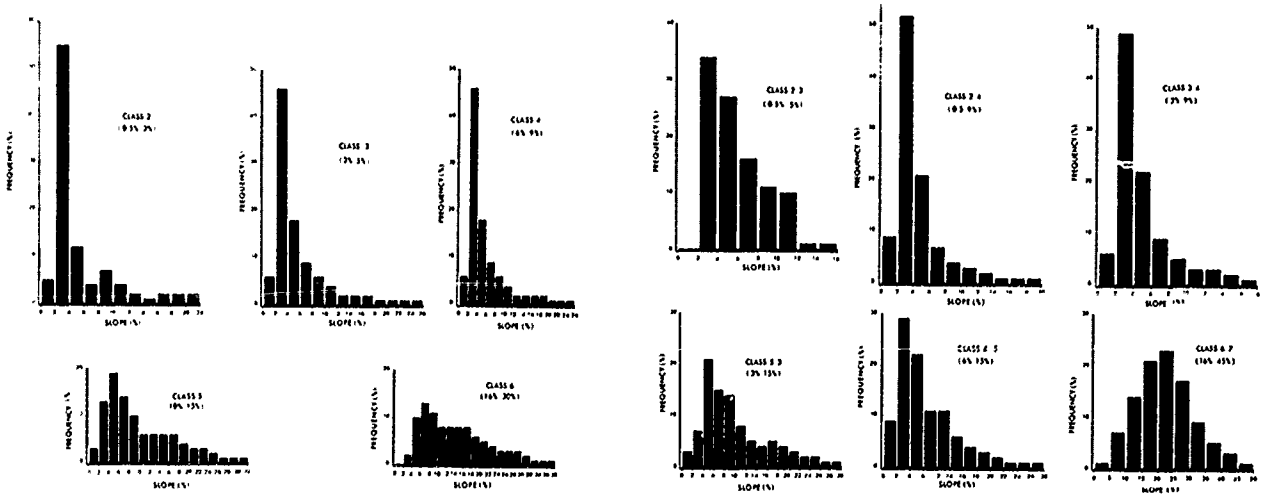


Figure 1 a) Topographic and landscape maps
b) Distribution of DTM slope gradients within interpreted slope polygons

SEGMENTATION D'IMAGE SPOT INTÉGRÉE À L'INFORMATION CARTOGRAPHIQUE EN VUE DE L'ÉTABLISSEMENT DE LA CARTE D'UTILISATION DE SOL AU MAROC

M. AIT BELAID, K.P.B. THOMSON, G. EDWARDS J.M. BEAULIEU

Département des Sciences Géodésiques et Télédétection, Centre de Géomatique

Département d'Informatique

Université Laval, Ste Foy, Québec, G1K 7P4

Tél. (418) 656-2645, Fax. (418) 656-3177

RÉSUMÉ

La présente communication s'intéresse à l'intégration de données de télédétection avec les données classiques de terrain. Elle présente une méthodologie d'intégration purement numérique d'une image multibande SPOT avec le canevas de remembrement rural. Le produit de cette opération est une autre image registrée à la carte nationale du Maroc et ayant quatre canaux dont les images ont été rééchantillonnées à 10m. Le quatrième canal contient le parcellaire agricole numérisé au moyen du sytème d'information à référence spatiale (SIRS),PAMAP.

L'algorithme de Segmentation Hiérarchique d'image basé sur l'Optimisation Séquentielle (SHOS) développé par Beaulieu est appliqué à la nouvelle "image" de quatre bandes afin de déterminer si la segmentation peut permettre d'établir la carte d'utilisation de sol et l'inventaire des cultures dans une région à microparcellaire.

ABSTRACT

This paper is concerned with the integration of remote sensing and conventional data. It presents a purely digital method of merging a multispectral SPOT image and field boundaries. This yields a product which is a new image registered to the national grid of Morocco, having four channels with images resampled to 10 m. The fourth channel contains the field boundaries which are digitized using the spatial information system PAMAP.

A Hierarchical Step-Wise Optimization (HSWO) algorithm developed by Beaulieu is applied to the new four band "image" to test the capability of the segmentation to map the land use and to provide the crop inventory in small areas of land.

1- INTRODUCTION

L'activité économique du Royaume du Maroc est intimement liée au développement et à l'exploitation des ressources naturelles; principalement dans le secteur agricole. Cela requiert la maitrise de l'utilisation de sol et l'inventaire agricole, deux éléments indispensables à la réussite de la politique de gestion et de planification agricoles (prédiction des rendements, négociation d'import/export).

Actuellement, ce problème est en bonne partie résolu par le concours de la télédétection dans les régions du monde caractérisées par une propriété de grande taille (BERNIER et Al., 1984); (BENIÉ, 1986), mais cela n'est pas le cas au Maroc et particulièrement dans la zone d'action de l'Office Régional de Mise en Valeur Agricole des Doukkala (ORMVAD) qui est notre aire d'étude.

Cette aire d'étude est la plaine fertile des Doukkala, province d'El Jadida, faisant une superficie de 5000 Km carrés dont 61000 hectares déjà équipées avec une infrastructure développée d'irrigation. Toute la plaine est caractérisée par un microparcellaire auquel se juxtapose la pratique multiculturale. En effet, la mise en valeur agricole et l'irrigation de cette plaine sont enteprises dans le cadre juridique défini par le Code des Investissements Agricoles promulgé le 25 Juillet 1969 qui rend obligatoire le remembrement et la pratique de l'assolement.

Le remembrement regroupe les parcelles d'un même propriétaire tandis que l'assolement subdivise la propriété ainsi constituée en plusieurs parcelles ou soles occupées par un groupe de cultures en rotation. De plus, certaines parcelles supportent plus d'une récolte à l'intérieur d'une même campagne agricole. Cela rend difficile le suivi régulier

de l'assolement et l'inventaire des cultures par les méthodes ponctuelles de terrain.

Une scène multibande du capteur Haute Résolution Visible HRV/SPOT de Mai 1988 couvrant la majeur partie du périmètre a été acquise. De plus, nous avons fait l'inventaire des cultures en place (pendant la période de prise de cette image) à l'intérieur d'un seul bloc (no.322), en effet, le bloc constitue dans la région l'unité d'exploitation agricole. L'assolement en place montre qu'il y a 5 cultures (figure 2-1) qui sont le blé dur (D) et le blé tendre (T) comme céréales, le bersim (B) et la luzerne (Z) comme cultures fourragères et la betterave à sucre (S) comme culture industrielle.

Nous retrouvons parmi les objectifs de ce projet le développement d'une procédure de segmentation optimale pour la région, l'identification et la cartographie des cultures ainsi que l'estimation de leurs superficies à partir de l'analyse de l'image SPOT.

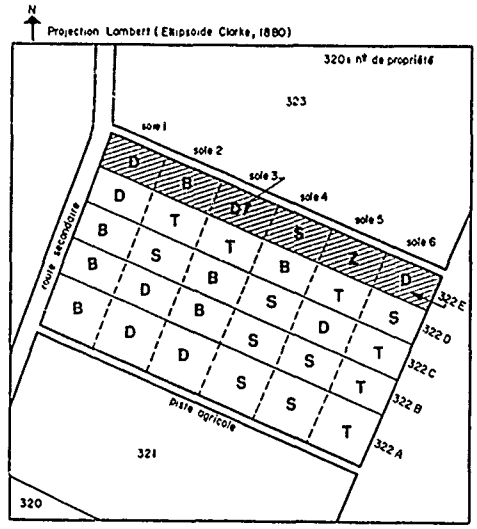


Fig 2-1 Plan parcellaire du 1:2000 réduit au 1:11260

2- INTÉGRATION DE L'IMAGE À L'INFORMATION CARTOGRAPHIQUE

L'interprétation de l'image de télédétection s'avère d'un intérêt vital pour diverses applications surtout en agriculture. Cependant, pour extraire le maximum d'information de façon précise à partir de ces images nous avons besoin d'un genre d'algorithmes "intelligents". À ce sujet, la segmentation et la classification d'image, l'utilisation de données auxiliaires (SIRS) ainsi que leur intégration sont autant d'atouts prometteurs, mais cela présente encore certaines difficultés. Il faut donc déployer des efforts de recherche pour développer des méthodes automatiques d'analyse d'image qui soient intégrées et entièrement opérationnelles.

Pour cela, la présente étude a été orientée de façon à pouvoir combiner au moyen d'un SIRS l'information cartographique avec l'image satellite comme support à la segmentation (EDWARDS et Al., 1989). La réussite d'une telle combinaison dépend de plusieurs facteurs à savoir les informations cartographique et spectrale ainsi que le modèle d'intégration et le SIRS utilisés.

Les premiers essais d'intégration existants ont tenté de combiner des images provenant de capteurs différents (WELCH et Al., 1987) puis on a envisagé la combinaison de l'image avec la photographie aérienne digitalisée (CHAVEZ, 1986) et enfin, la combinaison de l'image avec la carte vectorielle (PEDLEY, 1986).

Nous remarquons que ces essais d'intégration ont été développés surtout comme base à l'interprétation visuelle. Or du fait que l'on ne peut afficher à l'écran à la fois que trois canaux au maximum, on s'aperçoit tout de suite des limites de ces essais d'intégration utilisant généralement des opérateurs arithmétiques (+, - ou %) comme modèles d'intégration. Pour cela, nous suggérons une méthode d'intégration numérique que nous appliquons au cas de l'ORMVAD.

2-1 La méthode d'intégration proposée

La méthode d'intégration proposée est purement numérique; elle combine une image satellite à (n) canaux avec la carte numérisée et le produit de cette opération est une autre "image" à (n+1) canaux tout à fait prête pour des analyses aussi bien visuelle (rehaussement) qu'automatique (segmentation, classification). Le canal supplémentaire contient l'information cartographique.

Le canevas de remembrement est numérisé au moyen d'un SIRS à partir du plan parcellaire au 1/2000 de l'ORMVAD (figure 2-1) après avoir créé un fond cartographique avec les paramètres de la projection Lambert (Ellipsoide Clarke 1880) adoptée au Maroc.

Ensuite, la carte numérisée est convertie du mode vectoriel en trame (mode raster); la taille d'un pixel étant de 10m par 10m.

L'image utilisée est déjà corrigée au niveau (1B). Donc pour la rendre supperposable à la carte topographique du Maroc, nous avons effectué la correction géometrique en modélisant les déformations géometriques dans l'image au moyen de polynômes et en utilisant un certain nombre de points de contrôle. Ceci a donné lieu à une erreur moyenne quadratique résultante inferieure à 0.5 pixel. Ensuite, nous avons procédé à un rééchantillonnage des pixels en utilisant la méthode du plus proche voisin.

Dans la méthode générale d'intégration proposée, on distingue deux variantes différentes pour tenir compte du type d'application et d'analyse envisagées.

Première variante: MEthode des COntours (MECO): dans cette variante, on fait correspondre aux contours de propriétés une valeur thématique au moyen du SIRS et une autre valeur différente en dehors des contours. Une fois le parcellaire raster exporté vers le système de traitement d'image, ces valeurs thématiques seront interprétées comme des niveaux de gris. En d'autres termes ces limites de champs auront une largeur de 1 pixel.

Deuxième variante: MEthode des ETiquettes(MEET): dans ce deuxième cas, on assigne à chaque propriété une étiquette distincte. Donc chaque pixel reçoit une valeur correspondant à cette étiquette; et si deux pixels voisins ont des valeurs différentes, alors il existe une frontière entre les deux.

Le SIRS permet de former des polygones à partir du parcellaire numérisé et de calculer en même temps la superficie et le périmètre de chaque polygone (champ). Selon le concept de base de données à référence spatiale, chaque propriété constitue alors une entité avec un identifiant (numéro de propriété) et plusieurs attributs (superficie, matériel, thème, nom du propriétaire, type de culture, commune rurale).

3- SEGMENTATION

À ce fait, l'algorithme SHOS (version D, rédigé en langage C) de segmentation hiérarchique de l'image basé sur une optimisation séquentielle (Beaulieu, 1984) a été utilisé après avoir été adapté pour tenir compte de l'information cartographique.

L'algorithme débute par une partition initiale où chaque pixel représente un segment et réduit séquentiellement le nombre de segments en les fusionnant selon le critère de (WARD, 1963). Ceci produit une décomposition hiérarchique de l'image en fonction des valeurs des pixels avec comme contrainte sur la forme de la propriété l'information cartographique. L'algorithme devrait reconnaitre automatiquement les parcelles à l'intérieur de chaque propriété.

Cet algorithme exploite le fait que la fusion de deux segments affecte seulement leurs voisins. De plus, le critère d'optimisation séquentielle est déduit du critère d'approximation globale de l'image. On a donc deux niveaux d'optimisation.

La version utilisée de SHOS approxime l'image par un modèle simple en considérant celle-ci comme composée de régions homogènes ayant chacune un niveau de gris uniforme. Cette approximation produit le critère suivant (Beaulieu, 1984):

$$C_{i,j} = \frac{Ni \cdot Nj}{Ni + Nj} \sum_{k=1}^{k} W_k \, (Mi - Mj)^2 \qquad (3.1)$$

où Ni et Nj sont les nombres de pixels dans les deux segments voisins Si et Sj,
Mi et Mj les valeurs moyennes des niveaux de gris de chacun des segments,
Wb le facteur de poids utilisé pour combiner les valeurs de niveaux de gris des différentes bandes, et
k le nombre de bandes dans l'image une fois intégrée à la carte

Les résultats de la segmentation d'une fenêtre de 100 par 100 pixels de l'image avant et après l'intégration à la carte sont illustrés respectivement par les figures (3-2) et (3-3).

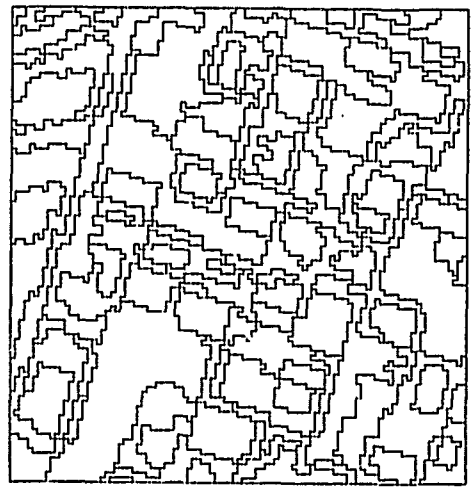


FIGURE 3-2: RÉSULTAT DE LA SEGMENTATION DE L'IMAGE SPOT CORRIGÉE GÉOMETRIQUEMENT, 120 Segments

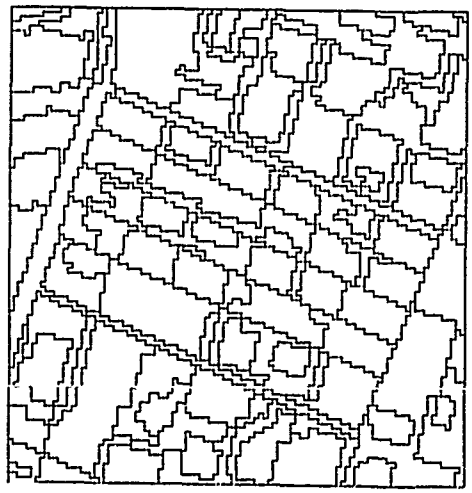


FIGURE 3-3: RÉSULTAT DE LA SEGMENTATION DE L'IMAGE SPOT INTÉGRÉE À LA CARTE PAR LA MÉTHODE MEET, 120 Segments

4- CLASSIFICATION PAR SEGMENT

La classification par segment a été retenue, en effet, la classification par pixel donne une faible précision et cela est prévisible étant donnée la taille très petite de la parcelle moyenne; environ un hectare.

Les résultats de la classification par segment seront disponibles sous peu. Cette classification peut être exécutée en deux étapes: D'abord, générer par calcul les statistiques relatives à chacun des segments de l'image segmentée et avoir une meilleur connaissance des différentes classes de culture existantes, puis, exécuter la classification segment par segment en tenant compte de ces informations dans notre regle de decision.

5- CONCLUSIONS

La comparaison des figures (3-2) et (3-3) montre les avantages de l'intégration du parcellaire à l'image. De plus, la figure (3-3) montre la haute fidélité (précision) de la carte produite par la segmentation comparée à la vérité terrain (figure 2-1). On remarque aussi une légère sursegmentation de l'image, mais cela peut être filtré au niveau de la classification.

En conclusion, la segmentation hiérarchique de l'image SPOT intégrée à la carte peut permettre d'établir la carte d'utilisation de sol et l'inventaire des cultures dans une région à microparcellaire à une échelle raisonable de 1/25000.

Comme implication directe, cette carte d'utilisation de sol, permet dans un premier temps de contrôler le respect de l'assolement par les agriculteurs.

Quant à l'inventaire agricole, il sera communiqué au Ministère de l'Agriculture et de la Réforme Agraire et à l'industrie agricole. Les sucreries par exemple seront intéréssée par la superficie occupée par la betterave à sucre pour planifier l'arrachage et le traitement.

En ce qui concerne l'ORMVAD, il lui revient l'exploitation de cette inventaire et de la carte pour les fins de gestion de sa zone d'action (besoin en eau des cultures, tarification des redevances d'eau, traitement des parasites et plus tard la prévision des récoltes).

Il ressort de la présente recherche que l'intégration des données image avec l'information cartographique dans un milieu caractérisé par la micropropriété est vivement recommandée afin de supporter la segmentation d'image à haute résolution et assurer une continuité dans les méthodes de gestion utilisant des données classiques representées par une carte par exemple

6- RÉFÉRENCES

BEAULIEU, J.M., "Hierarchical picture segmentation by step-wise optimization", Thèse de Doctorat, Dépt. de Génie Electrique, Université d'Ottawa, Ottawa, Canada, 1984.

BÉNIÉ, G.B., "Segmentation d'images de hautes limites de résolution pour fin d'inventaire agricole au Canada: Adaptation et développement d'algorithmes", Thèse de Doctorat, Dépt. des Sci. Géod. et Télédétection, Université Laval, Québec, Canada, 1986.

BERNIER, M., M. THERRIEN, K.P.B. THOMSON et C. GOSSELIN, "Potentiel d'applications des données du Thematic Mapper en agriculture au Québec", 9ème Symp. Canad. de Téléd., Terre neuve, 1984.

CHAVEZ, P.S., "Digital merging of Landsat TM and digitized NHAP data for 1:24000 scale image mapping", PE&RS, vol.LII, no.10, pp.1573-1700, 1986.

EDWARDS G., Y. BÉDARD et M. EHLERS, "On the integration of remote sensing with geographic information systems", Conférence Nationale sur les SIG, 28 Fev.-2 mars, Ottawa, Canada, 1989.

PEDLEY M.I., "Combined remotely sensed and map data as an aid to image interpretation and analysis", Internat. Jour. of remote sensing, vol.7, no.3, pp.305-308, 1986.

WARD J.H., "Hierarchical grouping to optimize an objective function", J. Amer. Stat. Ass., vol.58, pp.236-245, 1963.

WELCH R. and M. EHLERS, "Merging multiresolution SPOT/HRV and Landsat/TM data", image processing brief, PE&RS, vol.53, no.3, pp.301-303, March 1987.

REMERCIEMENTS

Nous adressons nos remerciements au Conseil de Recherche en Science Naturelle et Génie du Canada (CRSNG) et au directeur de l'ORMVAD pour leur contribution financière à ce projet. L'un des auteurs M. AIT BELAID, dénéficie d'une bourse dans le cadre des actions structurantes du Ministère de l'Éducation Supérieur et de la Science du Québec.

INTEGRATION OF DIGITAL ELEVATION MODEL PARAMETERS WITH LANDSAT AND SPOT IMAGERY FOR MAPPING MOUNTAIN GLACIERS: BUILDING A GLACIOLOGICAL INFORMATION SYSTEM

Denis J. Gratton and Philip J. Howarth

Department of Geography, University of Waterloo
Waterloo, Ontario N2L 3G1, Canada

Abstract

This paper deals with the building of a glaciological information system as an answer to the management of multi-source data required for the study of mountain glacier variations. A prime data source comes from remote sensing documents including satellite imagery. Topographic information is derived from the currently available 1:250,000 scale DEM of the Surveys, Mapping and Remote Sensing Branch of Energy Mines and Resources, Canada.

The SIF vector DEM file is read with the ARC/INFO GIS package from which an 8-bit integer elevation value grid file is produced. An 8-direction weighted-distance interpolation is performed to obtain a DEM in a raster format from which slope and aspect coverages are extracted. This information is used in the radiometric correction of Landsat MSS and TM images and in the characterization of glacier basins into physiographic units. The output coverage becomes the basic information to establish multidate climate and glacier behaviour patterns.

Keywords: Digital elevation model, Glacier mapping, Glaciological information system, Satellite imagery

Introduction

Mapping and monitoring glaciers stimulates scientific interest for two major reasons. First, glaciers in mountainous areas, particularly in Europe, have been recognized as important water storage systems for municipal/industrial and hydroelectric power purposes. The surface area and volume of individual glaciers are monitored, along with climatic parameters, so as to estimate future water availability. Second, glacier fluctuations are, and have been used for some time, as a surrogate measure for climatic change and variability. The use of mountain glaciers as a climatic indicator requires an assessment of the complex parameters of mass balance and glacier dynamics. Most historical data sets focus on the local and/or short-term aspects of these parameters. To obtain a more global perspective on this topic, it is important to be able to generalize and study the data over both regional and temporal scales. A glaciological information system offers an up-to-date solution to the management of multi-source data required to meet these objectives.

Methodology

The prime data source comes from remote sensing documents including satellite imagery which provide continuous information about glacial cover types at different spatial scales. A derivative of these documents, the digital elevation model (DEM), permits the extraction of topographic parameters, such as slope and aspect essential for an accurate automatic classification of imagery and to correlate glacier and climate variations. Other sources of data can be incorporated such as terrestrial triangulation assessments, *in situ* ablation stake and local stream flow measures according to the study scale.

In this paper, the steps required for the integration of a DEM with a satellite imagery base and the production of slope and aspect coverages to generate a glacial cover-type physiography map are described (Figure 1). Studies are being conducted in the Columbia Icefield region of Alberta using Landsat (MSS and TM) and SPOT panchromatic imagery. The DEM used in the analysis is the currently-available 1:250,000 scale Intergraph Standard Interchange Format (SIF) model available from the Surveys Mapping and Remote Sensing Branch of Energy Mines and Resources, Canada. The use of this regional-scale DEM for the radiometric correction of similar scale imagery revives an interest in MSS data for glacier cover-mapping.

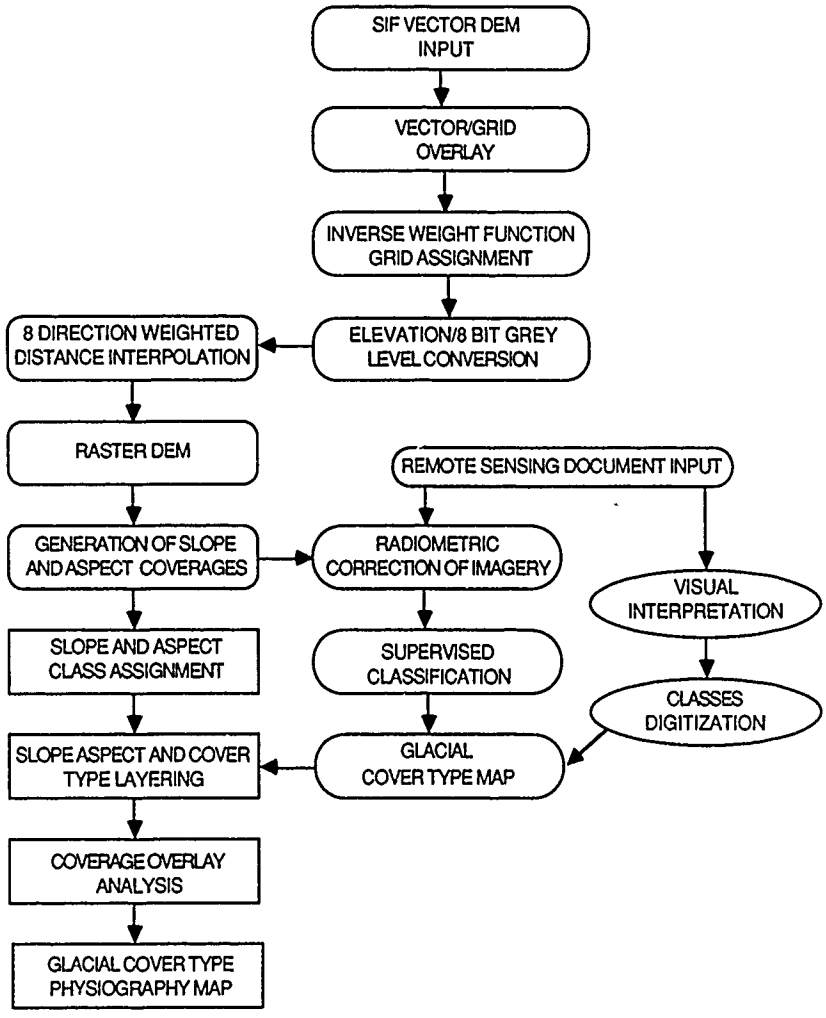


Fig. 1 - **Methodology for the integration of vector-based DEM parameters with remote sensing documents.**

Discussion

The use of the SIF for Canada's DEM format poses certain problems. Some GIS packages have the capability to read only the x and y coordinates of the SIF file. Even with ARC/INFO's SIFARC module that allows the reading of 3D SIF files, extra data conversion is required. The large file size of this ASCII DEM restricts the editing process and necessitates transformation into a grid file to permit sub-area selection. This represents unnecessary computer time and a loss of precision from the original data.

The vector/grid conversion is done using an inverse weight function (1/elevation) applied to the elevation values to preserve accuracy in the DEM. The values of this grid/vector DEM are then converted into an 8-bit integer file (0-255).

An interpolation is performed using an 8-direction weighted-distance algorithm (Personal Communication, Jinfei Wang, 1989). This step is required to fill in the null cells found in the grid. The slope and aspect coverages are then generated by the computation of the x and y distances between pixel values from a 3 x 3 window moved across the raster elevation image.

The input of glacial cover type can come from two sources. A regional study requires the analysis of a large number of glacier basins, so automatic classification of synoptic imagery such as MSS is appropriate. For local analysis, the visual interpretation of finer-resolution documents, such as SPOT panchromatic images and aerial photographs, and subsequent digitization maximizes the information input.

Using the real values of slope and aspect, the Landsat MSS and TM images are corrected for radiometric variations with the cosine of the incidence angle equation. This conventional approach has the advantage of reducing the illumination variability on a sloped terrain with a limited number of variables. A supervised classification of six cover types (snow, ice, debris on ice, surficial deposits, rock and vegetation covers) is performed on the corrected MSS and TM bands.

The multi-scale cover-type maps resulting from the visual and automated classification procedure are inputted, with the slope and aspect maps, to TYDAC'S SPANS GIS. The conversion into a quadtree grid structure in SPANS assists the introduction of raster and vector format data making the overlay analysis more precise. This analysis provides us with a characterization of glacier basins into different slope and aspect classes, the output product being the glacial cover-type physiography map.

Overall the DEM proves to be useful for the radiometric correction of imagery in relation to shadowing effects, as well as for producing slope and aspect maps that are a primary source of information in a glaciological information system. Work is presently in progress assessing the accuracy of these map products.

REMOTE SENSING AND GEOGRAPHIC INFORMATION SYSTEMS: TOWARDS INTEGRATED SPATIAL INFORMATION PROCESSING

Manfred Ehlers
National Center for Geographic Information and Analysis (NCGIA)
Department of Surveying Engineering
University of Maine
Orono, ME 04469, USA
Phone: (207) 581-2164
Fax: (207) 581-2206
E-Mail: Ehlers@Mecan1.bitnet

ABSTRACT

In recent years the demand for timely and accurate information has increased for a number of disciplines, including urban and regional planning, natural resources management, agricultural studies, topographic mapping, and geological exploration. A much discussed approach to meeting these demands is the development of Geographic Information Systems (GIS) which allow efficient data storage, incorporate processing and display functions, and permit the integration of digital remote sensing data.

Current and future satellite and shuttle programs such as Landsat, SPOT, MOS, ERS, JERS, SIR-C/X-SAR, RADARSAT, MOMS, and the space platform's Earth Observing System (Eos) are based on a variety of imaging sensors that will provide timely and repetitive multisensor earth observation data on a global scale. Visible, infrared and microwave images of high spatial and spectral resolution will eventually be available for all parts of the earth. It is, however, essential that efficient synergistic processing techniques be developed to cope with the large multisensor data volumes and to allow efficient GIS integration.

This paper discusses techniques for image fusion and GIS integration that have proved successful for synergistic evaluation of SPOT HRV, Landsat TM and SIR-B images. Examples of integrated remote sensing/GIS applications include topographic mapping/automated DEM extraction, cartographic feature extraction, spatial resolution improvement, and urban and regional planning.

Once incorporated in a digital database, the extracted information can be used to create cartographic products, analyze spatial distribution of growth patterns, or update existing transportation layers. Other GIS layers such as soils or zoning maps may then be overlayed and compared with the actual land use/cover information. Discrepancies can be quickly identified and analyzed. The data can also be used to update GIS files. It can be shown, for example, that merged SPOT multispectral (MSS) and panchromatic data can be effectively used in a GIS environment to routinely map and monitor land use change at scales of 1:24,000 and smaller.

SATELLITE SENSOR REMOTE SENSING

Earth observing systems of the future such as the proposed polar orbiting space platforms of NASA, ESA and Japan will likely bring another dimension to remote sensing. A variety of imaging (and non-imaging) sensors will be employed to cover the full range of the electromagnetic spectrum available for remote sensing of the earth (Butler et al., 1986). For example, a 30-m resolution imaging spectrometer will provide image data with a spectral coverage of 0.4 to 2.5 mm and a spectral resolution of 9.4 to 11.7 nm (Goetz et al., 1987). This amounts to 196 simultaneously recorded spectral bands. In addition, other sensors will provide information in different spectral bands (e.g., thermal infrared and microwave) and/or at different spatial resolutions yielding data volumes and spectral band combinations for which efficient processing methods are yet to be developed.

This multisensor, multispectral, multiresolution, multitemporal information will eventually be available for all parts of the earth and presents a data processing challenge to the remote sensing society that has to be addressed. Integrative processing techniques have to fuse the multi-image information to make it useful for a user community that is concerned with mapping, monitoring and modeling the earth's components. The potential of a multisensor dataset, however, can be readily assessed using current satellite imagery (Table 1). Satellite missions of the near future include a number of radar satellites (ERS-1, JERS-1, Radarsat and SIR-C/X-SAR) and follow-up missions such as SPOT-2, MOS-2, Stereo-MOMS and Landsat-6.

Table 1 Current Satellite Remote Sensing Mission Suitable for Mapping Applications.

Platform/ Year	Sensor	# Spectral Bands	Spectral Range *)	Ground Resolution	Country
Shuttle	SIR-B	1	Radar	17 - 58 m	U.S.
1983 -	MOMS	2	VIS/NIR	20 m	Germany
Landsat-4/-5	MSS	4	VIS/NIR	80 m	U.S.
1982/84 -	TM	7	VIS/NIR/ MIR/TIR	30 m (TIR:120m)	
SPOT	HRV-P	1	Pan(VIS)	10 m	France
1986-	HRV-XS	3	VIS/NIR	20 m	
MOS	MESSR	4	VIS/NIR	50 m	Japan
1987-					
IRS-1A	LISS	4	VIS/NIR	36.5 m	India
1988-					

*) VIS = Visible
NIR = Near Infrared
MIR = Middle Infrared
TIR = Thermal Infrared
Pan = Panchromatic

REMOTE SENSING AND GIS

Until recently, remote sensing was viewed as a science that provided definitive end products usually maps, statistics, and/or reports. Today, however, these end products are increasingly being produced in digital formats to be used with Geographic Information Systems (GIS). Several factors have accelerated this trend, including: 1) recent developments in hardware and software for GIS systems (Dangermond, 1988; Frank, 1988; Croswell and Clark, 1988); 2) the availability of high resolution satellite data in digital format such as SPOT's High Resolution Visible (HRV) and Landsat's Thematic Mapper (TM); and 3) progress in automated information extraction, especially the application of image matching techniques for Digital Elevation Model (DEM) generation (Ehlers and Welch, 1987; Day and Muller, 1988; Swann et al., 1988).

It can be demonstrated that image data, especially that from Landsat TM and SPOT HRV, are suitable for base map production and map revision tasks at scales which were previously deemed to be impossible for remote sensing applications (Dowman, 1987; Gugan, 1988).

Remotely sensed data, on the other hand, can be put to their best use if they are incorporated in Geographic Information Systems (GIS) which are "designed to accept large volumes of spatial data, derived from a variety of sources, including remote sensors, and to efficiently store, retrieve, manipulate, analyze and display these data according to user-defined specifications." (Simonett et al., 1983). A GIS therefore, when combined with up to date data from a remote sensing system, can assist in the automation of interpretation, change detection, map compilation, and map revision functions (Ehlers, 1989). One major part of a GIS is the ability to overlay various layers of spatially referenced data. This allows the user to determine graphically and analytically, how structures and objects (e.g. roads, water distribution, commercial zoning) interact with each other.

Until recently, GIS databases were almost solely produced by digitizing maps. The usual result of this process, however, is that the original data and all related information, such as levels of abstraction and generalization or information about the digitizing procedure (e.g. scanning or manual) and its reliability, are no longer available for the GIS user.

In addition, the land mass is a dynamically changing entity. For example, rivers change location, forests are cut, roads and houses are built. Consequently, our information stored in GIS is only a static model of the real world and has to be updated on a regular basis. Satellite data offer repetitive and global coverage of the earth's surface and, as such, offer the potential to monitor these dynamic changes within a GIS. The potential for timely updates of geographic datasets will become even more important if GIS technologies are being extended to marine areas.

APPLICATIONS OF REMOTE SENSING/GIS INTEGRATION

For remote sensing and GIS systems to be truly integrated, several technical problems still need to be overcome. The most straight forward of these problems is the raster/vector dichotomy. The major problem is caused by the difference in the structures used to acquire and store the data. Remote sensing systems are almost universal in the use of the raster, or "pixel" format. In this format, all information is stored as a collection of picture elements (pixels), each holding the information at specific spatial coordinates. GIS systems, on the other hand, typically use the vector, or "line" format. This format holds the data as a collection of lines and polygons, where each structure holds the information for a specific region of the image. While both systems have their advantages, neither is ideal. In addition to these, other data structures are being explored as alternatives (Edwards et al., 1989).

Various efforts have been made to research the problems of integrating remote sensing data into a GIS. The most promising areas that have emerged to-date can be associated with automated DEM generation, change detection and database/base map production (Ehlers and Welch, 1985; Estes, 1985; Logan and Bryant, 1987; Barker, 1988).

DEM Generation

Ehlers and Welch (1987), in one of the first stereocorrelation experiments undertaken with real satellite data, achieved a root-mean-square error (RMSE) in Z of ± 42 m with sidelapping Landsat TM images. At the very weak base-to-height (B/H) ratio of 0.18, this Z error is equivalent to a planimetric correlation error of ±0.3 pixel, confirming that correlations to better than ±0.5 pixel are feasible with real data. Similar results in correlation accuracy have recently been recorded for SPOT stereo data (Simard et al., 1988; Day and Muller, 1988). With more favorable base-to-height ratios

and 10-m pixels for SPOT, however, RMSEz values range from ±6 m to ±18 m. These accuracies allow the generation of DEMs and ortho-images at scales of 1:50,000 from which further topographic information for GIS applications can be derived.

Change Detection

Using timely remote sensing data to identify and quantify changes of the earth's surface has been one of major application areas for GIS analyses. Recently, Ehlers et al. (1989) demonstrated that SPOT data could be used in a GIS environment for regional growth analysis and local planning. For the first time, satellite data were operationally used at a scale of 1:24,000 which is suitable for many regional and local planning tasks. The steps involved were: (1) georeferencing the SPOT panchromatic and multispectral data; (2) merging panchromatic and multispectral data to form multispectral images of enhanced spatial resolution (Ehlers, 1988); (3) semi-automated classification of land cover and land use; (4) integration with GIS database; and (5) rigorous statistical error analysis based on extensive field checking.

It can be shown that SPOT image data yield accuracies for growth detection as high as 93%. Once incorporated in a GIS database, the spatial growth pattern can be readily analyzed. The data can also be used to quickly update GIS files at scales of 1:24,000 and smaller.

Database/Base Map Production

A very important factor of high resolution satellite data is that they can serve as a map basis when no other reliable data is available. Welch et al. (1985) demonstrated that Landsat TM image data are suitable for digital map production at scales of 1:100,000 and smaller. MacDonald Dettwiler prepared topographic, thematic and perspective map products from TM data at a scale of 1:50,000 (Rose et al., 1986). All information, except the cartographic names, were derived from TM image data. More recently, Konecny et al. (1988) concluded that 1:50,000 scale map products can be compiled from SPOT stereo data. In a related study, TM image data were successfully used as a digital database for multisensor integration (Welch and Ehlers, 1988b). Overall rectification accuracy of SIR-B image data could be improved by about 50% when the data were registered to a georeferenced TM image rather than rectified using map ground control points. The amount of detail of cartographic information, on the other hand, that could be depicted from the merged dataset was significantly higher than from the single TM and SIR-B image layers (Ehlers, 1988).

TRENDS IN REMOTE SENSING/GIS INTEGRATION

Although the potential for GIS/remote sensing integration is readily evident from these and other examples, they do not yet represent an integrated GIS/remote sensing processing concept. Rather, they can be seen as case studies for information extraction from remotely sensed data for integration with a GIS which usually results in data transfer between independent image processing systems and GIS. Recently, however, manufacturers of GIS and remote sensing/image processing systems have started to offer integrated solutions and/or standard interfaces between different systems to facilitate data integration (Welch and Ehlers, 1988a). Still, the ultimate potential of remote sensing/GIS integration can only be realized if this integration can be fully automated. At this time, this automation is not yet possible, since human interaction, assistance, and editing is still required. Recent developments, based on artificial intelligence and expert system technology, may accelerate the automated integration process (McKeown, 1987; Goodenough et al., 1987). Other factors that may accelerate the pace toward integrated systems include the progress in display systems, storage devices, and advanced computer architecture (workstations, parallel computing, networking). These factors may eventually shift the focus from vector-based systems to raster and/or hybrid GIS/remote sensing systems.

CONCLUSIONS

The potential of multisensor satellite remote sensing for GIS can be demonstrated using current satellite data from the Landsat, Shuttle, and SPOT programs. These remote sensing data have been successfully used in a GIS environment for base map production, automated DEM extraction, terrain visualization, and map revision/update at scales from 1:24,000 to 1:100,000. To realize the full potential of multisensor remote sensing for GIS, however, automated integration techniques have to be developed. It is anticipated that advances in hardware, software, and AI research will ultimately lead to fully integrated GIS/remote sensing processing.

REFERENCES

1. Barker, G.R. "Remote Sensing: The Unheralded Component of Geographic Information Systems", Photogrammetric Engineering and Remote Sensing, Vol. 54, No. 2, pp. 195-199, 1988.
2. Butler, D.M. et al. From Pattern to Process: The Strategy of the Earth Observing System. NASA Eos Steering Committee Report, Vol. II, 140 pp., 1986.
3. Croswell, P.L. and S.R. Clark. "Trends in Automated Mapping and Geographic Information System Hardware", Photogrammetric Engineering and Remote Sensing, Vol. 54, No. 11, pp. 1571-1576, 1988.
4. Dangermond, J. "A Technical Architecture for GIS", Proceedings GIS/LIS '88, San Antonio, Texas, Vol. 2, pp. 561-570, 1988.
5. Day, T. and J.-P. Muller. "Quality Assessment of Digital Elevation Models Produced by Automatic Stereomatchers from SPOT Image Pairs", Photogrammetric Record, Vol. 12, No. 72, pp. 797-808, 1988.

6. Dowman, I.J. "Prospects for Topographic Mapping Using SPOT Data", Proceedings, SPOT 1: Image Utilization, Assessments, Results, Paris, France, pp. 1163-1172, 1987.
7. Edwards, G., Y. Bedard, and M. Ehlers. "On the Integration of Remote Sensing with Geographic Information System", Proceedings, GIS National Conference, Ottawa, Canada (in press), 1989.
8. Ehlers, M. "Multisensor Image Fusion Techniques in Remote Sensing", Proceedings, XVIth Congress of ISPRS, Kyoto, Japan, Vol. B7, pp. 152-162, 1988.
9. Ehlers, M. "The Potential of Multisensor Satellite Remote Sensing for Geographic Information Systems", Technical Papers, 1989 ASPRS-ACSM Annual Convention, Baltimore, MD, Vol. 4, pp. 40-45, 1989.
10. Ehlers, M. and R. Welch. "Satellite Remote Sensing Data as Input to Geographic Information Systems" Proceedings of the 11th International Symposium on Machine Processing of Remotely Sensed Data, Purdue University, West Lafayette, Indiana, p. 219, 1985.
11. Ehlers, M. and R. Welch. "Stereocorrelation of Landsat TM Images", Photogrammetric Engineering and Remote Sensing, Vol. 53, No. 9, pp. 1231-1237, 1987.
12. Ehlers, M., M.A. Jadkowski, R.R. Howard and D.E. Brostuen. "Application of SPOT Data for Regional Growth Analysis and Local Planning", Photogrammetric Engineering and Remote Sensing (in press), 1989.
13. Estes, J.E. "Geographic Applications of Remotely Sensed Data", Proceedings of the IEEE, Vol. 73, No. 6, pp. 1097-1107, 1985.
14. Frank, A.U. "Requirements for a Database Management System for a GIS", Photogrammetric Engineering and Remote Sensing, Vol. 54, No. 11, pp. 1557-1564, 1988.
15. Goetz, A.F.H., et al. "HIRIS", High-Resolution Imaging Spectrometer: Science Opportunities for the 1990's. NASA Eos Instrument Panel Report, Vol. IIc, 74 pp., 1987.
16. Goodenough, D.G., M. Goldberg, G. Plunkett and J. Zelek. "An Expert System for Remote Sensing", IEEE Transactions on Geoscience and Remote Sensing, Vol. GE-25, No. 3, pp. 349-359.
17. Gugan, D.J. "Satellite Imagery as an Integrated GIS Component", Proceedings, GIS/LIS '88, San Antonio, Texas, Vol. 1, pp. 174-180, 1988.
18. Konecny, G., K. Jacobsen, P. Lohmann and W. Müller. "Comparison of High Resolution Satellite Imagery", Proceedings, XVIth Congress of ISPRS, Vol. B9/IV, pp. 226-237, 1988.
19. Logan, T.L. and N.A. Bryant. "Spatial Data Software Integration: Merging CAD/CAM/Mapping with GIS and Image Processing", Photogrammetric Engineering and Remote Sensing, Vol. 53, No. 10, pp. 1391-1395, 1987.
20. McKeown, D.M. "The Role of Artificial Intelligence in the Integration of Remotely Sensed Data with Geographic Information Systems", IEEE Transactions on Geoscience and Remote Sensing, Vol. GE-25, No. 3, pp. 330-348, 1987.
21. Rose, D.R., I. Laverty and M. Sondheim. "Base Map Production form Geocoded Imagery", Proceedings of the ISPRS Commission VII Symposium, Enschede, The Netherlands, pp. 67-71, 1986.
22. Simard, R., G. Rochon and A. Leclerc. "Mapping with SPOT Imagery and Integrated Data Sets", Proceedings, XVIth Congress of ISPRS, Kyoto, Japan, Vol. B11/IV, pp. 440-449, 1988.
23. Simonett, D.S., R.G. Reeves, J.E. Estes, S.E. Bertke, C.T. Sailer. "The Development and Principles of Remote Sensing", in Colwell, R.N. (Ed) Manual of Remote Sensing, Second Edition, Vol. I, American Society of Photogrammetry, pp. 1-35, 1983.
24. Swann, R., D. Hawkins, A. Westwell-Roper and W. Johnstone. "The Potential for Automated Mapping from Geocoded Digital Image Data", Photogrammetric Engineering and Remote Sensing, Vol. 54, No. 2, pp. 187-193, 1988.
25. Welch, R., T.R. Jordan and M. Ehlers. "Comparative Evaluations of the Geodetic Accuracy and Cartographic Potential of Landsat-4/-5 TM Image Data", Photogrammetric Engineering and Remote Sensing, Vol. 51, No. 9, pp. 1249-1262, 1985.
26. Welch, R. and M. Ehlers. "LIS/GIS Products and Issues: A Manufacturer's Forum", Photogrammetric Engineering and Remote Sensing, Vol. 54, No. 2, pp. 207-210, 1988a.
27. Welch, R. and M. Ehlers. "Cartographic Feature Extraction from Integrated SIR-B and Landsat TM Images", International Journal of Remote Sensing, Vol. 9, No. 5, pp. 873-889, 1988b.

INTERACTIVE QUERY OF CANADA - WIDE MINERAL DEPOSITS DATA, USING A GEOGRAPHIC INFORMATION SYSTEM

T.L. Webster

TYDAC Technologies Inc.
1600 Carling Avenue
Suite 310
Ottawa, Ontario K1Z 8R7

G.F. Bonham-Carter, D.F. Garson, R.M. Laramee

Mineral Resources Division
Geological Survey of Canada
601 booth Street
Ottawa, Ontario K1A 0E8

ABSTRACT

This paper describes a demonstration project, showing how to interrogate large point files using a micro-computer based GIS.

For evaluation, an 18,500 record subset of a Canada-wide mineral occurrence index was written to an ASCII file, one record per occurrence. The fields in each record comprise decimal latitude, longitude, status (0-6) of six commodities (Zn, Pb, Cu, Au, Sn, Ag), overall status, name and pointer index of the occurrence.

A basemap of Canada was created from a raster image showing land and water using a Lambert Conformal projection. Canada-wide gridded data, showing Bouguer Anomaly, Isostatic Anomaly, Vertical Gradient, Gravity, Digital Elevation and Total Field Magnetics were imported into the system. Geological regions and political provinces were digitized from hardcopy maps and imported into the same database.

Once in the GIS the point file can be displayed over any reference map and/or vector files. The Nearest Point Query Operation is then executed. For any window the cursor is moved, the nearest point is highlighted and the attributes associated with that point are instantaneously displayed on the monochrome screen.

Further details about the occurrence may be obtained by temporarily exiting the GIS to DOS and executing a routine to display additional information, such as deposit type, NTS#, and associated mineralogy. Only one key stroke is required to re-enter the GIS.

Introduction

A Geographic Information System (TYDAC, 1989) is being evaluated by the Geological Survey of Canada (GSC) for interactive interrogation of large point files. For this study, 18,500 Canada-wide mineral occurrences were written to an ASCII file, one record per occurrence. The fields in each record of the file comprise decimal latitude, longitude, status (0-6) of six commodities (Zn, Pb, Cu, Au, Sn, Ag), overall status of the location, name and pointer index of the occurrence. Another point file representing deposits typical of the major deposit types (Eckstrand, 1984) was also imported into the system. This file consists of about 900 records containing fields similar to the mineral occurrence file, but with the addition of a "deposit type" attribute code.

Several Canada-wide gridded geophysical files were obtained from the Geophysics Division, GSC. The files were transformed to image format and downloaded from the VAX mainframe to an image analysis system operating on a 386 micro-computer. The image files were then further transformed into an interchange format using a raster interface module of the GIS. The geophysical data included: Bouguer Gravity Anomaly, Isostatic Gravity Anomaly, Vertical Gradient Gravity, Digital Elevation (5km pixels), and Total Field Magnetics (2km pixels).

In addition to the raster imagery, digitized hardcopy maps were also input as topologically-labelled vector data. These maps included: geological regions and political provinces of Canada. Provincial capitals and the Trans-Canada Highway were added as cultural features. The vector files were created using the GIS digitizing module.

Establishing The Database

A "Universe" was established with a digitized basemap obtained from Environment Canada (Environmental Information Systems), using a Lambert Conformal projection, standard parallels 49 N and 77 N, origin at 40 N and central meridian 95 W. The ASCII point files were imported directly into the system and projected into the Lambert universe.

The 5km gravity and digital elevation raster data were converted into quadtree map files. A quad level of 11 (15 is the maximum) was chosen to maintain the 5km resolution of the original raster. The level defines the size of the smallest quad cell to be used for data storage, i.e. level 11 is equivalent to $2^{11} = 2048$ cells wide. The magnetic data was also converted into a quadtree map, using level 13, i.e. with $2^{13} = 8192$ cells wide.

The graphics board used for the study was a Number Nine Corporation, PRO 1280 card, with output to a Mitsubishi colour monitor. The graphics card supplies 1280 by 1024 pixels and supports a colour range of 0-255. The 8 bit (0-255) raster imagery was therefore left unclassified when converted into quadtrees. A second series of quadtree maps was created based on classification break points to give a histogram equalization. The vector map data, geological regions and political provinces, were also transformed into quadtree maps.

A vector file was generated from the quadtree basemap showing the shoreline and major lakes. The geological regions and provincial boundary vectors can be superimposed over any map for reference purposes.

Query Operations

With the maps and point files in the GIS the information can then be interactively queried. A typical operation includes the following steps:

1. Selection of a suitable window, e.g. Maritime provinces;
2. Display a reference map, e.g. Bouguer Anomaly;
3. Superimpose vectors with selected colour or line style, e.g. provincial boundaries;
4. Selection of the point file and display of all copper occurrences in the file;
5. Enter point query mode

The arrow keys or a mouse is used to move the cursor. The nearest point is highlighted and the attributes associated with that point are instantaneously displayed on the monochrome screen. For example, the following attribute record

45.2056 -64.0403 Zn 6 PB 6 Cu 6 Au 0 Sn 0 Ag 6 6 701 Walton

includes latitude, longitude, status code (0-6) for Zinc, Lead, Copper, Gold, Tin and Silver, an overall status of the occurrence, a unique pointer index and occurrence name. The status codes have been simplified to 0, 2, 4, 6 corresponding to:

0 - does not occur
2 - occurrence
4 - prospect
6 - deposit

The user can temporarily leave the GIS and execute a routine to access further details including: NTS number, geological province and subprovince, deposit type, description of mineralogy, and a list of all commodities, which occur and their status.

The major deposits point file can also be displayed and interrogated in a similar manner. Points can be displayed based on the status of one of the same six commodities (Zn, Pb, Cu, Au, Sn, Ag) or, based on a "deposit type" code assigned in Eckstrand (1984). Selected deposit types can be highlighted (e.g. Mississippi Valley Pb-Zn and about 80 others), in addition to the point query.

The map files in the database can be queried in a similar operation. Up to 10 maps can be interrogated simultaneously by cursor position, latitude, longitude or by projection XY coordinates. The cursor is moved, the values of the selected maps at that position are instantaneously displayed on the monochrome screen. Thus, for a particular mineral occurrence, the various geophysical responses, the elevation and geological provinces can be determined at that site or at a succession of point locations in the same neighbourhood.

Conclusions

1. Point-query operations with the GIS are fast enough for interactive spatial query of large point databases, even though the system operates on a microcomputer.
2. Flexible display of large raster images; such as 2 km gridded airborne magnetic data for the whole of Canada, on colour display monitors up to 1280 x 1024 resolution, provides the graphical background against which mineral deposit data can be displayed and interrogated, aiding interpretation.
3. The interactive query operations are an important function that complement the modelling capabilities of geographic information systems (Bonham-Carter et al, 1988).

Acknowledgements

Geophysical data and associated projection parameters were obtained with the assistance of J. Halpenny and D. Dods of the Geophysics Division, Geological Survey of Canada. The basemap was kindly supplied by Environmental Information Systems, Environment Canada.

References

Bonham-Carter, G.F., Agterberg, F.P., and Wright, D.F., 1988, Integration of Geological Datasets for Gold Exploration in Nova Scotia, Photogrammetric Engineering and Remote Sensing, V 54, No U, p. 1585 - 1592

Eckstrand, O.R., (editor), 1984, Canadian Mineral Deposit Types: a geological synopsis, Geological Survey of Canada, Economic Geology Report 36, 86p.

TYDAC, 1989, SPANS Spatial Analysis System, Version 4.0, TYDAC Technologies Inc., 1600 Carling Ave., Ottawa, Ontario.

SYNERGISM OF REMOTELY SENSED AND CONTEXTUAL DATA TO MONITOR CHANGES IN LAND USE

H. Xu & J.A.T. Young

University of Edinburgh
Department of Geography
Drummond Street
Edinburgh EH8 9XP

ABSTRACT

This paper examines the feasibility of LANDSAT image data in land use studies for a test location in South-East Scotland. Integration of remotely sensed data and contextual data is achieved by using an interface between GEMSTONE-35 image processing system and ARC/INFO Geographic Information System (GIS). Included are: classification of land use/land cover, accuracy assessment, transformation of classified image to GIS and comparison of results for different dates. Analysis on the TM image of September 14, 1986 has been completed. The overall accuracy of the image classification was assessed as 88%. Along with a discussion on the research results and difficulties in image analyses, preliminary conclusions are drawn that the synergism of remotely sensed and contextual data is of substantial significance in land use studies.

KEY WORDS: Remote Sensing, GIS, Interface, Overlay, Classification, Land Use.

1. INTRODUCTION

Major changes in land use will take place in Europe in the immediate years ahead as the effects of how policies change on agricultural support become felt. Satellite imagery affords a means of monitoring changes in land use and updating information on the distribution and dynamics of land use. Using satellite imagery successfully as a regular input for land use monitoring in Scotland requires a measure of luck in being able to obtain usable imagery on a routine basis and substantial contextual knowledge of the ground situation to counter this and the subtlety of the changes in land cover. One of the characteristics of land use in Scotland is the complexity within a short distance. This leads to difficulties in analysing satellite images. Therefore incorporation of ancillary data and a prior knowledge are of special value.

This paper examines the application and the feasibility of LANDSAT image data in land use studies for a test location in South-East Scotland and seeks to establish a feasible approach to monitoring recent changes in land use. Integration of the data is achieved by using an interface, which has been developed in the Department of Geography, University of Edinburgh, between the GEMSTONE-35 image processing system and ARC/INFO-ORACLE. Image analysis is the principal thrust of the study and the major procedure includes: definition of a classification system, classification of land use/land cover, post-classification filtering, accuracy assessment, transformation of classification results to GIS and comparison of results for different dates. Finally it is followed by a critical discussion based on the research results, concerning the classification accuracy and the usefulness of the LANDSAT imagery and its potential synergism with contextual data.

2. METHODOLOGY

The adopted methodology involves several steps. First is the selection of the study area. The Lammermuir Hills area in South-East Scotland covers different land use zones from lowland to upland and has experienced such changes as afforestation of former moorland, conversion of rough pasture to improved grassland and expansion of arable land towards the upland over the past four decades (Eadie, 1984). Availability of a range of data and easy access to the area for ground truth work also influenced this choice of a study area.

The datasets for the study included: the yearly Farm Cropping Programme by the Edinburgh Centre of Rural Economy which provides valuable reference information on ground truth data; images for two dates with one being MSS data for April 24, 1984 and the other TM data for September 14, 1986; and additional cartographic data from topographic maps, maps of land capability for agriculture,

and soil maps, which provide a basis for improving classification of the LANDSAT imagery.

GEMSTONE-35 is a 1024 x 1024 workstation with its own integral 68020-based processor and is linked to a MICROVAX computer for access to the satellite data and file management. With up to 27 image stores, the GEMSTONE system can perform a wide variety of processing operations. ARC/INFO-ORACLE was developed in the Department of Geography, University of Edinburgh and is interfaced with the GEMSTONE-35 system via a series of programs. Classified raster image files can be transformed into ARC/INFO vector format and recoded into new class groupings whilst ARC/INFO polygon and arc coverages can be converted into GEMSTONE image format for manipulation in the image processing environment.

Prior to the image classification, geometric correction of LANDSAT imagery helps to eliminate spatial distortions of the image, stemming from many factors which affect the collection of remotely sensed data. Mergence with other datasets and transference between vector and raster data also require that both image and the digital cartographic data layers are geometrically corrected to the same coordinate system. In this study, one TM image and one MSS image have been corrected for the study area. The output images have been registered to the British National Grid coordinate system. The pixel sizes have been resampled to 50m x 50m in order to overcome problems resulting from the different spatial resolutions of the two images.

A second step prior to classification was to gather ground truth data for training the classifier and generating the statistics. The yearly Farm Cropping Programme by the Edinburgh Centre of Rural Economy provided a means to assist discrimination between fields with different crops. The aspatial data of agricultural statistics were mapped into spatial form. The resultant map was used as ground truth data for defining training areas. In addition, aerial photographs and Ordnance Survey maps were also used to assist in defining areas. The statistics generated from the training areas were then saved and used for the image classification. The integrated land cover classification was carried out on a pixel by pixel basis using a Maximum Likelihood Classifier.

LANDSAT image classification has enabled rapid generation of land cover maps. But user acceptance has lagged behind due to the difficulties in the specification and statistical testing of accuracy. In this study, a post-classification median filter was applied to remove isolated pixels. The proportion of mixed pixels remaining unclassified was also reduced to a large extent after the filtering. Accuracy assessment of the classification performance was then made by random sampling of individual areas for each class to examine inclusive and exclusive classification errors. Finally, satisfactory classification results for different dates can be compared to show changes in land use. This can be achieved by transferring the digital classification results to ARC/INFO-ORACLE after recoding each class to a new group.

3. CURRENT RESEARCH STATUS

Imagery analysis on the TM data of September 14, 1986 has been completed.

After contrast stretching the image bands 3,4 and 5, ground control points were chosen by identifying recognisable points from both the image and the Ordnance Survey 1:50000 map sheets. Then using GEMSTONE modules, the image was geometrically corrected and the pixel size was resampled from its original 30m x 30m to 50m x 50m. The cubic convolution method was chosen for the interpolation of the output pixel intensity. The relationship between image coordinates and ground coordinates was also established for the area delimited by British National Grid Coordinates.

Then twenty-eight areas were defined with the help of the aforementioned Farm Cropping Programme, aerial photographs and other ancillary data incorporated through ARC/INFO. A prior knowledge of the ground situation and visual interpretation of the image also played a part in defining the training areas.

The statistics generated from those twenty-eight areas were saved and used to classify subscenes of the image. The segmentation of the image to subscenes for classification helped to reduce certain obvious errors. For example, pixels classified as built-up areas on the moors and built-up areas classified as moors can be readily excluded. Classification results on subscenes were then copied into one image with the same pixel size and location as prior to the classification. This combined image shows the image classification results for the whole study area.

The image containing classification results was enhanced by applying a median filter to exclude isolated pixels. The image file was further recoded into new groupings for rapid combination of classes prior to transference to ARC/INFO or for redisplay on the GEMSTONE workstation. The number of class after the recoding was reduced from 28 to 9, namely, water surface, grassland, pasture, arable land, moorland, forest land, built-up area, cloud/shadow and unclassified area.

An accuracy assessment was then carried out on seven of those nine classes. Cloud/shadow and the unclassified areas were ignored. The random sampling of the individual pixels was applied to each class with a sample size determined by its weight in the whole image and the assumed prior accuracy (Hay, 1979; Rosenfield, 1982). Each sample point was then checked with updated Ordnance Survey 1:25000 map sheets and supplemented by field checking. A prior knowledge also permits a decision to accept or reject the classification of certain pixels. A contingency table was thus presented (Table.1). As van Genderen (1977) demonstrated, this table shows the following aspects:

** The frequency that any one land use type (on the ground) is erroneously attributed to another class, as those values in Row F1 of Table.1.

** The frequency that the wrong land use (as observed on the ground) is erroneously included in any one class, e.g., those in column F2 of table.1.

** The proportion of all sampled pixels which are misclassified. In this case, 42/363 of all attributions are incorrect.

** The determination of whether the errors are random or subject to a persistent bias.

Thus the overall accuracy was assessed as 88 per cent with an accuracy for each class better than 77 per cent. In general, the classification performance was satisfactory, although certain classes such as grassland could not satisfy the general criteria proposed by Loelkes et al (1983), where the minimum level of interpretation accuracy in identifying land use/land cover categories from remote sensing data should be at least 85 per cent.

Using the integration between GEMSTONE image processing system and ARC/INFO GIS, the results of classification were transferred to ARC/INFO as a new thematic overlay for further use.

A number of important points about using remote sensing image and the classification performance need to be made as follows.

First of all, the effect of clouds and shadows seemed a problem which precluded a complete study of the image. In the TM image, the classified clouds and shadows took 1.24 per cent of the pixels in the whole scene within the study area. Apart from the fact that areas under clouds or shadows will not give any information on land use, the existence of clouds and shadows may affect the classification performance. For instance, some of the dark shadow areas had very low spectral reflectance values and therefore were

LAND USE (image classification)	LAND USE (on the ground)								
	water	grass	pasture	arable	moorland	forest	built-up	total	F2
water	26							26	0
grass		37	2	8		1		48	11/48
pasture		4	45		1	2		52	7/52
arable		5		108			6		11/119
moorland			3		29	4		36	7/36
forest						51		51	0
built-up				5	1		25	31	6/31
total	26	46	50	121	31	58	31	363	
F1	0	9/46	5/50	13/121	2/31	7/58	6/31		42/363

Table 1. Numbers of Sampled Pixels in Actual and Classified Land-use Categories

misclassified as water whereas they were actually crop lands. As aforementioned, it is quite difficult to get cloud free images in Scotland. Thus the incorporation of ancillary data such as land use survey from other sources is crucial. Work by Gurney (1983) has shown that contextual methods can be used in separation of cloud and cloud shadow from the remainder of a satellite scene. To achieve this, the synergism of remote sensing and contextual data is of particular value and significance.

Another problem arose from the definition of a category of land use. One example can be the definitions of grassland, pastures and arable land. By broad definition, grasslands include heath, bracken and other rough grassland and pastures are mainly the improved grassland while arable lands comprise crop land and fallows. On the image, these three land uses may have a very similar spectral reflectance at the time when the image was taken. Therefore classification errors may occur easily. This would affect the overall classification performance. From Table 1, it can be seen that 8 out of 11 pixels which were misclassified as grasslands were attributed to arable lands and 4 out of 7 errors for pastures were attributed to grasslands. To overcome those difficulties, more detailed ground truth data and certain image enhancements prior to the classification may be helpful.

Finally, whilst post-classification filtering helped to remove isolated pixels and reduce the proportion of the unclassified pixels, it affected the classification to some extent. Small areas may be merged into their neighbouring land uses, e.g., a farm house or a road next to croplands may be represented as arable land on the classified image after the filtering. Considering the characteristics of land use in Scotland, where subtle variations occur within a short distance, the effect of merging should not be ignored completely.

4. PRELIMINARY CONCLUSIONS

Results so far have shown the great potential of LANDSAT imagery in area measurement and mapping spatial distribution of land cover types. The integration of remote sensing and GIS techniques is of substantial significance in land use studies. Remote sensing generates a wide variety of data as an input to update GIS data planes while GIS provides an efficient use of the ancillary data required by remote sensing analysis and enables relationships between datasets to be explored and tested.

BIBLIOGRAPHY

1. Eadie,J., "Trends in agricultural land use: the hills and uplands", Agriculture and the Environment, edited by D. Jenkins pp13-20,1984.

2.Genderen, J. L. van, "Testing land-use map accuracy", Photog. Eng. Rem. Sensing, Vol.43, No.9, pp1135-1137, 1977.

3.Gurney, C.M., "The use of contextual information in the classification of remotely sensed data", Photog. Eng. Rem. Sensing, Vol.49, No.1,pp55-64, 1983.

4. Hay,A.M., "Sampling design to test land use map accuracy", Photog. Eng. Rem. Sensing, Vol.45,No.4,pp529-533, 1979.

5.Loelkes, G.L. et al, "Land use/land cover and environmental photointerpretation keys", USGS Bulletin 1600, US Government Printing Office: 1983.

6.Rosenfield, G.H. et al, "Sampling for Thematic Map accuracy testing", Photog. Eng. Remote Sensing, Vol.48, No.1,pp131-137, 1982.

7. Young,J.A.T.,"A U.K. Geographic Information System for Environmental Monitoring, Resource planning and using Satellite remotely sensed data", Remote sensing society monograph No.1, 1986.

GESTION DES RESSOURCES AGRICOLES À L'AIDE D'UN SYSTÈME D'INFORMATION GÉOGRAPHIQUE: APPLICATION À L'ÉVALUATION QUANTITATIVE DE L'ÉROSION HYDRIQUE.

*BARIL, D., *PERRAS, S., **PESANT, A. et *BONN, F.

* Centre d'applications et de recherches en télédétection (CARTEL), Université de Sherbrooke, Sherbrooke, Québec, Canada. J1K 2R1. tél. (819) 821-7180

** Agriculture Canada. Station de recherches de Lennoxville, Lennoxville, Québec, Canada. J1M 1Z3. tél. (819) 565-9171

Résumé

On a intégré à un système d'information géographique (S.I.G.) l'équation universelle des pertes de sol ce qui nous a permis de calculer l'érosion hydrique d'un secteur agricole de 500 km^2 des Cantons de l'Est (Québec).

Une image HRV de SPOT nous a permis d'en cartographier l'utilisation du sol (facteur C). Le modèle numérique d'élévation (M.N.É.) a été intégré au S.I.G. et il nous permet d'en dériver le facteur de topographie (facteur LS). Finalement, à la carte pédologique numérisée, nous associons l'érodabilité de chacun des types de sol (facteur K).

En appliquant cette équation, nous obtenons les pertes de sol pour l'ensemble du territoire d'étude. Les pertes totales sont de 791 tonnes métriques/ha/an réparties comme suit. 24 t/ha/an pour la forêt, 95 t/ha/an pour les cultures et 672 t/ha/an pour les secteurs de sol nu.

Abstract

We integrated to a geographical information system (G.I.S.) the Universal Soil Loss Equation (U.S.L.E.) to estimate soil erosion losses due to water erosion in an agricultural region of 500 km^2 in the Eastern Townships (Québec).

A HRV SPOT image was used to determine the land use (C factor). The digital elevation model (D.E.M.) was also integrated to a G.I.S. to derive the topographic factor (LS factor). Finally, we added the digitized soil map units with their soil erodibility to the G.I.S. (K factor).

When applying this equation, we obtained the total soil loss for our study area. The total soil loss is 791 metric tons/ha/year divided as follows: 24 t/ha/an for forest, 95 t/ha/an for crops and 672 t/ha/an for bare soils.

Mots-clés: Érosion hydrique, télédétection, U.S.L.E. (équation universelle des pertes de sol), S.I.G. (système d'information géographique), imagerie HRV de SPOT, M.N.É. (modèle numérique d'élévation).

Problématique

La baisse de fertilité des sols agraires est en partie due aux pertes de sols causées par l'érosion hydrique. L'augmentation de la superficie des cultures sarclées (maïs) dans des secteurs à topographie accidentée tels que les Cantons de l'Est accroit les risques d'érosion hydrique (Pesant et Boivin, 1985). A la station de recherches de Lennoxville, on a mesuré, sur des pentes de 10%, des taux d'érosion hydrique pouvant atteindre 15 700 kg/ha/an (Pesant, 1983).

De plus, des études ont démontré que l'érosion hydrique a déjà érodé de 25 à 75% de l'horizon A des sols de cette région (Dubé, 1975). Il s'en suit une perte de fertilité qui doit être compensée par des apports suplémentaires en fertilisants. Ceci entraine une augmentation des coûts de production et leur dispersion contribue à la pollution des cours d'eau car ils sont eux aussi érodés et transportés par l'eau de ruisellement.

L'érosion hydrique des sols entraine aussi d'autres problèmes comme la sédimentation dans les cours d'eau, les réseaux de drainage et les réservoirs artificiels. Aux États-Unis, on a évalué des coûts annuels de 144 $ à 197 $ millions pour les pertes de capacité d'emmagasinage en eau (Crowder, 1987) et de 1 $ à 1.2 $ milliards pour les coûts associés au dragage des matériaux provenant de l'érosion des sols agraires seulement (Clark II, 1985). De plus, les sédiments érodés, transportés par les cours d'eau, entraînent des coûts supplémentaires lors des opérations de filtrage des eaux de consommation. On a évalué, pour les communautés de l'état d'Ohio, (États-Unis), qu'une réduction de 25% de l'érosion hydrique des sols se traduirait par une économie de 2.7 $ millions en frais de traitements des eaux. (Forster *et al.*, 1987).

Malgré tous ces impacts économiques et environnementaux, très peu d'études globales des problèmes d'érosion ont été faites à ce jour. L'équation universelle des pertes de sol (U.S.L.E) (Wischmeier and Smith, 1958) est une méthode qui permet d'évaluer les zones à haut risque d'érosion hydrique. D'une façon traditionnelle, elle est utilisée avec des outils conventionnels d'investigation (photographies aériennes, cartes topographiques et pédologiques et mesures sur le terrain). Nous nous proposons d'en améliorer la rapidité de calcul en employant des outils modernes tels la télédétection et les S.I.G. afin de couvrir de plus vastes territoires tout en réduisant les coûts de réalisation. Nous pourrons ainsi déterminer plus rapidement et adéquatement des moyens pour corriger la situation avant que les sols les plus sensibles ne deviennent infertiles.

Hypothèse

L'intégration des données satellitaires, de la carte pédologique numérisée et d'un modèle numérique d'élévation (M.N.É.) à un (S.I.G.) augmente la rapidité de mesure de l'érosion hydrique. Nous assurons ainsi un suivi d'un territoire plus vaste qu'avec les méthodes traditionnelles.

Objectif

Le but principal de cette recherche est d'intégrer des documents numérisés provenant de sources diverses à un (S.I.G.) dans le but d'augmenter la rapidité de mesures des pertes de sol dues à l'érosion hydrique à l'aide de l'U.S.L.E.. Les documents utilisés sont les suivants:

-Image HRV de SPOT (utilisation du sol: facteur C)
-Carte pédologique numérisée (indices d'érodabilité: K)
-M.N.E. (facteur de topographie: à partir de S (pente) et L (longueur de pente))

Les S.I.G permettent de gérer, de synthétiser et d'analyser la masse de données disponibles sur un territoire. Il s'avère donc l'outil le plus adéquat pour la gestion agricole et plus particulièrement le suivi de l'érosion hydrique.

Méthodologie

a) Secteur d'étude:

Le secteur d'étude est situé à 5 km au sud-est de Sherbrooke et couvre 500 km^2 (Figure 1). Il est limité au nord par la rivière Saint-François, au sud par Cooaticook, à l'ouest par le lac Massawippi et à l'est par Sawyerville. On y retrouve principalement des fermes laitières et, associées à celles-ci, la culture du maïs ensilage, des céréales et des plantes fourragères.

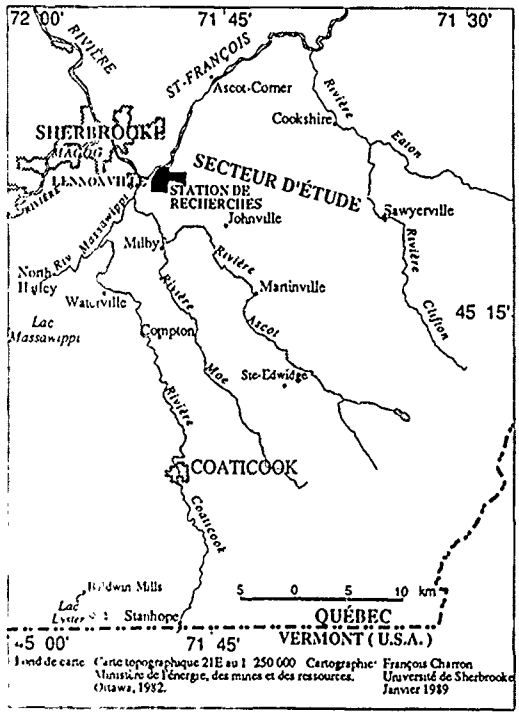


Figure 1: Secteur d'étude

b) Équation universelle des pertes de sol (U.S.L.E.):

Cette équation, mise au point par Wischmeier et Smith (1958), nous permet de calculer les pertes annuelles de sol causées par l'érosion hydrique. Elle s'exprime comme suit:

$$A = R*K*LS*C*P$$

où: A=pertes annuelles de sol (kg/ha/an)
R=données pluviométriques: énergie cinétique des pluies (MJ*mm/ha*heure*an)
K=indice d'érodabilité: dérivé de tests utilisant le simulateur de pluie. (varie de 0 à 1) (t*ha*heure/ha*Mj*mm)
LS=facteur de topographie (calculé à partir de la pente (S) et de la longueur de la pente (L) tous deux dérivés du M.N.É.
C=Évolution des cultures: données moyennes (à partir de tableaux de Wischmeier (1978) (varie de 0 à 1)
P=Pratiques culturales: données moyennes (à partir de tableaux de Wischmeier (1978) (varie de 0 à 1).

Cette équation nous servira à déterminer les pertes de sol dues à l'érosion hydrique sur l'ensemble du territoire.

c) Traitements numériques des images:

Une image HRV du satellite SPOT acquise le 25 octobre 1986 a été utilisée pour déterminer l'utilisation du sol.

-Étapes de traitements:

i) Corrections radiométriques: Pour obtenir des valeurs de réflectance au capteur dans les trois bandes HRV de SPOT, nous devons transformer les niveaux de gris en réflectance apparente au capteur en nous servant des valeurs de calibration fournies avec l'image.

ii) Corrections géométriques. Tous les documents sont corrigés géométriquement à un système de référence spatial (grille U.T.M.) pour pouvoir être intégrés à un S.I.G.. (cartes topographiques (1: 50 000) de Sherbrooke et Coaticook (21E/5 et 21E/4)).

iii) Classification supervisée: Des sites d'entraînement sont choisis sur l'image pour chacune des quatre classes d'utilisation du sol à cartographier (hydrographie, forêt, cultures et pâturages et sols nus). L'algorithme de classification du maximum de vraisemblance est alors appliqué à l'image en utilisant les trois bandes spectrales (verte, rouge et infra-rouge). (tableau 1).

Tableau 1: Superficies obtenues par la classification supervisée; image HRV de SPOT du 25 octobre 1986 avec indices C et P associés à chacune des classes.

Classes	Superficies		Indices	
	(%)	(km^2)	C	P
Eau	0,07	0,351	0,000	0,0
Forêt	66,09	331,575	0,003	0,5
Culture	26,49	132,931	0,030	1,0
Sol nu	7,35	36,873	1,000	1,0
Total	100,00	501,730		

d) Carte pédologique:

La numérisation des 15 types de sol de la carte pédologique s'est effectuée sur le système de traitements d'images DIPIX ARIES III à l'aide de la caméra Eikonix. Nous effectuons ensuite une correction géométrique pour superposer cette image à l'image HRV de SPOT. (tableau 2).

Tableau 2: Superficies de chacun des types de sol et indices K correspondants.

	Types de sol	Superficies		Indices
		%	(km^2)	K
D	Terre franche sablonneuse de Dufferin	13,80	73,199	* 0,21
Sh	Terre franche sablonneuse de Sherbrooke	15,26	80,952	* 0,11
Gl	Terre franche de Greensboro	26,18	138,822	* 0,09
Ca	Terre franche de Calais	7,35	38,992	0,07
T	Terre marécageuse (tourbe)	1,38	7,316	0,00
CL	Terre franche argileuse de Coaticook	6,72	35,663	* 0,21
Ml	Terre franche rocheuse de Magog	4,94	26,173	0,24
Sf	Terre franche sablonneuse de Sheldon	12,36	65,526	* 0,30
M	Sable fin de Milby	3,63	19,245	0,04
Cn	Terre franche sablonneuse fine de Colton	3,41	18,088	0,05
Ua	Sol alluvionnaire non classifié	0,33	1,771	0,06
Dg	Terre franche graveleuse de Danby	0,32	1,706	0,31
L	Terre franche argileuse de Lennoxville	0,12	0,629	0,21
Rs	Roche	0,53	2,817	0,00
Bm	Terre franche de Berkshire	3,67	19,442	0,06
Total		100,00	** 530,341	

* Indices K mesurés à l'aide du simulateur de pluie (été 1988).
** Les superficies totales varient entre certaines cartes du fait que les documents ne proviennent pas des mêmes sources.

e) Mesure des indices d'érodabilité (K) des différents types de sols de la région à l'aide du simulateur de pluie.

Les indices d'erodabilité de cinq types de sols de la région ont été mesurés à l'aide du simulateur de pluie de la station de recherches de Lennoxville (Agriculture Canada). Ce simulateur de pluie est dérivé de celui mis au point par Meyer et McCune (1958). Nous effectuons deux simulations sur chacun des cinq types de sols et ce dans trois classes de pente (environ 4, 8 et 12%). Nous en dérivons le facteur K (érodabilité) de l'U.S.L.E. (tableau 2).

f) M.N.É.

Ce modèle numérique d'élévation provient de la *U S. Geological Survey (USGS)* Il a été réalisé à partir de la carte topographique au 1 250 000 et sa résolution spatiale a été rééchantillonnée à 50 m pour les besoins de cette étude Sa résolution altimétrique est de 30,3 m Le M N É est corrigé géométriquement dans le but de le superposer à l'image HRV de SPOT. Nous en dérivons les facteurs S (pente) et L (longueur de la pente) , nécessaires lors du calcul du facteur de topographie (LS). Les pentes sont exprimées en quatre classes: 0-5%, 5-10%, 10%-15% et >15%, (les bornes supérieures étant exclusives). Les longueurs de pentes sont comprises entre 0 et 63 (*50m par niveau).

L'équation pour le calcul de LS facteur de topographie (Wischmeier and Smith, 1978) intègre les valeurs de pente et de longueur de pente. Elle s'exprime comme suit:

$$LS = (L/22,15)^{m} * (65,41 * (\sin(S))^{2} + 4,56 * \sin(S) + 0,065)$$

où m= indice variant selon la pente

Tableau 3: Superficies de la carte des pentes, indices S (radians) et m associés.

Pente	Superficies		Indices	
%	(%)	(km²)	S (radians)	m
0-5	71,97	381,682	0,024993	0,3
5-10	21,45	113,774	0,074857	0,5
10-15	4,23	22,415	0,124355	0,5
>15	2,35	12,470	0,197397	0,5
Total	100,00	530,341		

g) Données pluviométriques:

Le facteur R de l'U.S.L.E. représente l'énergie cinétique annuelle des pluies. Pour cette étude, nous le considérons comme constant sur l'ensemble du territoire. La valeur de 1275 MJ•mm/ha/heure/an lui a été attribué (Lagacé, 1980).

h) S.I.G.:

Les images correspondant aux différents facteurs de l'U.S.L.E. sont ensuite converties en format matriciel Quadtree avec le logiciel SPANS de TYDAC. Le format Quadtree condense l'information là où il n'est pas nécessaire d'avoir une précision plus grande. Les zones uniformes sont constituées d'unités de plus grandes dimensions tandis que les zones hétérogènes sont décimées jusqu'à uniformité du pixel. La figure 2 nous montre un exemple de décimation Quadtree à un niveau 3. Nous pouvons ainsi obtenir jusqu'à 16 niveaux de décimation de l'image et ce processus s'arrête lorsque les pixels sont uniformes. Ceci permet d'augmenter la rapidité d'exécution des calculs car, avec ce format, il y a généralisation de l'information.

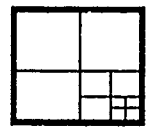


Figure 2: Format Quadtree (niveau 3)

La prochaine étape consiste à créer des fichiers de variables associés à chacun des documents (facteurs K, P, C, R). Pour chacun des thèmes, un chiffrier contenant les valeurs de chaque facteur est créé.

Nous écrivons l'équation des pertes de sols dans un fichier et ce programme fait appel à chacune des valeurs préalablement incluses dans les différents chiffriers pour calculer les pertes de sols. La carte résultante est aussi automatiquement obtenue selon les classes de pertes de sols désirées.

Tableau 4: Superficies de pertes de sol et pertes totales de sol pour chaque utilisation du sol

	Pertes de sol (kg/ha/an)											
Utilisation du sol	0	0-10	10-50	50-100	100-500	500-1 K	1-2 K	2-5 K	5-10 K	>10000		Pertes Totales (t/ha/an)
Eau	0,351	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	km²	0
	0,07	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	%	
Forêt	8,461	0,172	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	km²	24
	1,69	64,37	0,03	0,00	0,00	0,00	0,00	0,00	0,00	0,00	%	
Cultures	0,945	113,444	16,259	1,813	0,471	0,000	0,000	0,000	0,000	0,000	km²	95
	0,19	22,61	3,24	0,36	0,09	0,00	0,00	0,00	0,00	0,00	%	
Sols nus	0,405	0,000	14,278	9,042	10,292	1,650	0,882	0,296	0,027	0,001	km²	672
	0,08	0,00	2,85	1,80	2,05	0,33	0,18	0,06	0,01	0,00	%	
Total	10,162	436,387	30,709	10,855	10,763	1,650	0,882	0,296	0,027	0,001	km²	791
	2,03	86,98	6,12	2,16	2,15	0,33	0,18	0,06	0,01	0,00	%	

Résultats

Les pertes de sol dues à l'érosion hydrique ont été cartographiées en 10 classes. Le tableau 4 nous présente les superficies pour chaque thème de la carte de l'utilisation du sol en plus des pertes totales pour chaque thème. Les pertes totales sur l'ensemble du secteur d'étude sont de 791 tonnes métriques/ha/an réparties comme suit. 24 t/ha/an pour la forêt, 95 t/ha/an pour les cultures et 672 t/ha/an pour les secteurs de sol nu. Ceci nous permet de mettre en lumière le fait que l'érosion hydrique est en forte relation avec l'utilisation du sol.

Les pertes moyennes de chacun des thèmes de la carte d'utilisation du sol se répartissent comme suit. 0,71 kg/ha/an pour le couvert forestier, 7,16 kg/ha/an pour les cultures et 182,26 kg/ha/an pour les zones de sol nu.

Pour la carte des pentes, les pertes moyennes de sol se répartissent de la façon suivante. 5,74 kg/ha/an pour les pentes de 0-5%, 29,41 kg/ha/an pour les pentes de 5-10%, 56,38 kg/ha/an pour les pentes de 10-15% et 88,77 kg/ha/an pour les pentes de plus de 15% et ce, indépendamment de l'utilisation du sol.

Conclusion

La cartographie quantitative de l'érosion est donc beaucoup simplifiée par l'utilisation d'un S.I.G.. En effet, les paramètres du modèle peuvent être modifiés à volontée ce qui nous permet de visualiser indépendamment les effets et l'importance de chacune des variables du modèle. Cette étude est donc une première démarche en vue de l'optimisation du suivi de l'érosion hydrique des sols. La méthodologie ainsi développée nous permettra de calculer systématiquement les pertes de sol sur de vastes terriroires, et ce, à moindres frais. Cette méthode est d'ailleurs la seule pouvant atteindre un tel but pour des superficies aussi grandes.

Par contre la méthode pourrait s'avérer beaucoup plus précise si tous les facteurs nécessaires au calcul de l'érosion hydrique étaient plus précis.

Le facteur de topographie (dérivé de S et L) serait plus exact si le M.N.É. avait une meilleure résolution altimétrique et spatiale.

Le facteur C serait quant à lui beaucoup mieux évalué si plusieurs images étaient acquises pendant la saison végétative d'une même année et ce, pour quelques années successives. Nous pourrions ainsi avoir une meilleure idée de l'évolution des cultures au cours d'une même année ainsi que de la rotation des cultures au fil des ans. La classification pourait comporter plus de thèmes ce qui permettrait d'augmenter la précision du facteur C.

Le facteur de pluviométrie (R) pourrait quand à lui être calculé à chaque année. Ceci nous permetterait de tenir compte des valeurs extrèmes parfois observables d'une année à l'autre.

Cette étude étant une première approche, nous tenions à vérifier la validité de la méthode. L'amélioration de la précision de certains documents utilisés aurait pour effet d'améliorer de beaucoup la précision du calcul des pertes de sol. De plus, l'ajout de modèles de diffusion des particules solides et de la pollution diffuse (produits dissouts) nous permetterait de tendre vers un système expert de gestion agricole.

Remerciements

Nous tenons à remercier les gouvernements fédéral et provincial pour leur aide financière dans le cadre de l'entente auxiliaire fédérale provinciale en agro-alimentation (dossier # 3B1-68400260-002), le centre des recherches en services agricoles du Québec (Ministère de l'agriculture, des pêcheries et de l'alimentation: subvention #2124) ainsi que M. François Charron pour la réalisation de certaines figures.

Références

1. Clark II, E. H., "The off-site costs of soil erosion", J. Soil Water Conserv., Vol. 40, pp 19-22, 1985.

2. Crowder, B. M. "Economic costs of reservoir sedimentation: a regional approach to estimating cropland erosion damages", J. Soil Water Conserv., Vol. 42, pp 194-197, 1987.

3. Dubé, A., "L'eau et l'érosion des sols", Ressources, Vol. 6, Ministère des richesses naturelles du Québec, 1975.

4. Forster, D. L., Bardos, C. P. and Southgate, D. D., "Soil erosion and water treatment costs", J. Soil Water Conserv., Vol. 42, pp 349-352, 1987.

5. Lagacé, R., "L'équation universelle de pertes de sol. un outil", Érosion et conservation des sols, 8ième colloque de génie rural, Unv. Laval, pp 1-35, 1980.

6. Meyer, L. D. and McCune, D. L., "Rainfall simulator for runoff plots", Agricultural Engineering, Vol. 39, pp 644-648, 1958.

7. Pesant, A., "Les dangers de la culture du maïs sur sols en pente", Faits saillants, Agriculture Canada, Station de recherche de Lennoxville, Québec, 4 p., 1983.

8. Pesant, A. et Boivin, F., "Photo interprétation de la pente, de la superficie et du type de dépôt des sols cultivés en maïs fourrager dans les comptés de Richmond, Sherbrooke et Stanstead", Faits saillants, Agriculture Canada, Station de recherche de Lennoxville, Québec, 4 p., 1985.

9. Pesant, A., Dionne, J. L. and Genest, J., "Soil and nutrient losses in surface runoff from conventional and no-till corn systems", Canadian J. Soil Sci., Vol. 56, pp 443-451, 1987.

10. Wischmeier, W.M., and Smith, D.D., "Predicting rainfall erosion losses - A guide to conservation planning", Agr. Handbook, No. 537, U.S. Dept. Agr., Washington D.C. 58 p., 1978.

11. Wischmeier, W.M., and Smith, D.D., "Rainfall energy and its relationship to soil loss", Trans. Am. Geophys. Union, Vol. 39, pp 285-291, 1978.

INTEGRATED IMAGE-MAP PRODUCTS FOR TOWNSHIP LEVEL LAND INFORMATION

M.J. Manore
Canada Centre for Remote Sensing

R. Protz, B.A. Proud
Land Resource Science,
University of Guelph

ABSTRACT

As part of a project to introduce remotely sensed data into an existing Land Related Information System (LRIS) for local level land-use planning, a series of digital map and image products were developed and evaluated for their potential to provide useful information to local land managers. The test site was Oxford Co., Ontario, where a comprehensive LRIS is used operationally by the county planning office. Using Thematic Mapper, SPOT and MOS-1 image data, a series of enhancements and ground cover classifications were produced. These image products were combined with data from the LRIS in two ways; in image (raster) form as either themes or enhanced images, and as vector map layers extracted from the boundaries of image classifications. While the visual impact of enhanced images overlayed with existing map files was very strong, the integration of map and image data on a single, interactive display remains a technical problem for many GIS/Image Analysis systems. On the other hand, image classifications converted to vector map layers offer to potential users information on land use/land cover in an easily retrieved and manipulable form. The technical issues of integration are fewer in this approach and disruption of existing data structures and system management are kept to a minimum.

Keywords: Remote Sensing/GIS Integration, Soil Conservation

INTRODUCTION

The integrated use of remotely sensed data and map information from Geographic Information Systems (GIS) has received much attention recently. While many technical issues of data exchange between specific vendors' systems have been resolved, the use of integrated image/map products is still low. One reason for this is the scarcity of digital GIS databases used operationally within Canada. For several years, Oxford County, Ontario has been the site of an operational Land Related Information System (LRIS) for land use and municipal planning (Ottaway, 1988). Previous unpublished work within CCRS used GIS layers within an image analysis system to evaluate the use of integrated products for local level land management. While the Oxford Co. LRIS is primarily focused on urban planning, this work concentrated on rural land use information and soil conservation, a lower resolution application better suited to the capabilities of remote sensors (Cihlar *et al.*, 1987; Sauchyn, 1989). This paper describes a follow-on project which had two specific objectives;

- to identify the specific information needs of the township level managers which may be satisfied by combined image/map products; and

- to develop sample products for evaluation by the local users.

The intent was to develop digital products which could be manipulated within a GIS in order to take advantage of the network for data distribution which already existed within the county. However, restrictions in the current display hardware of the Oxford Co. system would prevent the full colour display of image data, so some of the initial products have been created within the image analysis environment where the viewing of images is superior.

At the time of writing, the work was not as advanced as had been expected. This paper, therefore, will report only on the progress to date and will discuss only preliminary results. Example products and user reaction will be available for discussion at the time of presentation in July.

DATA

A variety of image and map data sources were available for the project (Table 1). The focus of the data was Norwich Township, a subset of the Oxford County data set. All map data were supplied in topologically structured ARC/INFO coverages (line and/or polygon) except for the NTS 1:50000 map sheet which contained feature-coded spaghetti vectors in

Intergraph IGDS format. These data were converted to ARC/INFO format through in-house systems and vendor supplied translation packages. The NTS 1:50000 files have been used to date, instead of the OBM 1:20000 files, because of the smaller data volume and their compatibility with image data at 25 metre pixel spacing. Future work with high resolution SPOT PLA data may require the higher positional accuracy of the OBM data to achieve acceptable overlays. Significant misregistration of features was observed between map layers from different sources (sometimes > 50 metres) with the worst registrations observed in the agricultural overlays.

To date, the Thematic Mapper and MOS-1 MESSR image data have been UTM geocoded at 25 metre pixel spacing through manual ground control point selection. The SPOT data will be UTM geocoded at 10 metre spacing in the near future. Only the May 28, 1988 scene has be used in the generation of products so far because of the significance of spring time vegetation and crop residue cover in soil conservation.

The exchange of data between the IAS and GIS was performed using a combination of vendor supplied and custom software. The systems used in this work were a Dipix ARIES-III image analysis system (IAS) and an ESRI ARC/INFO geographic information system, both located at the Canada Centre for Remote Sensing. The map coverages from ARC/INFO were transferred to the ARIES and rasterized via exchange software supplied by Dipix (Landriau and Manore, 1989). Raster theme files which resulted from image classification were transferred to ARC/INFO grid files where they were vectorized by standard ARC/INFO routines. The translation of the raster data to ARC/INFO grid format was accomplished with in-house programs.

TRIAL PRODUCTS

Map Overlays on Image Data: Using the map and pre-processed image data, a series of trial products was generated. The first were products for viewing on systems with good image display capabilities. These were three-band colour composites comprising contrast stretched TM 4, TM 5, and TM 3 (RGB) onto which selected, rasterized map layers were overlayed. The road, natural drainage, and property overlays were found to improve the visual quality of the imagery more than any others (Figure 1). They provided a good geographic frame of reference which assisted the visual interpretation and location of ground features, but which did not obscure the within field ground detail. It is anticipated that users of this type of presentation will appreciate the familiarity of simple image products (because of their similarity to aerial photographs) and will be impressed with the level of detail visible within individual fields. In particular, micro-drainage patterns, variations in soil moisture, and the distribution of vegetation and crop residue were easily interpretable. This type of product (in either digital or photographic form) could be used as a site reference when dealing with relatively small areas such as a single farm.

The remaining map layers, such as soils, tile drainage, and contours were considered to be of less value in this type of presentation because their content was less readily related to what could be visually interpreted from the scene, and because they sometimes obscured detail from the underlying image. It was reasoned that the value of these layers lies more in their distinctive attributes (soil texture, elevation, etc.) than in their location alone, and that these are more appropriately treated in the GIS.

Image Classification: Rapid parallepided classifications were performed on the May TM image to extract three visually evident classes; vegetated fields, woodlot, and fields with high harvest residue. Using only TM 3, TM 4, and TM 5, classifications for each class were created individually based on visually identified training sites. A post classification filter with a threshold of 25 pixels was applied to 'clean' the results of spurious pixels.

When displayed over the image data, the classified theme layers were found to be disruptive to visual interpretation rather than beneficial. Although the areas of the chosen class were highlighted in the presentation, the opaque themes obscured all underlying detail.

The three classifications were then exported in raster format to the GIS where they were subsequently vectorized into the same map coordinates as the original map coverages. In this format, the results of the classification become as displayable and manageable as any existing map layer. Several options for display were examined; the most agreeable of which was the classification displayed in solid shades underlying vector linework and/or attribute text from the other coverages. Because the classified data are solid themes with no interior detail, the overlying layers do not significantly hinder the visibility of the underlying information.

Although the visual map products appear to very useful, the real value of these transferred classification polygons is that they may be used in analytical map operations with the other covers. Example products that will be derived in the near future include;

- the calculation of total vegetated, bare, and high residue surface cover by region or operator's property;
- the identification of bare fields on high slopes of light soil textures to target areas of high soil erosion potential;
- the identification of land cover (vegetated, residue, bare) on fields with artificial drainage (tiles).

DISCUSSION

The preliminary results of this work have identified a small number of potential products derived from only a sub-set of the available image and map data. What has emerged, however, are two distinct classes of products which reflect the display and analysis capabilities of IAS and GIS systems.

The high level of local detail interpretable from enhanced, multi-band imagery is readily usable because of its similarity to familiar aerial photographs. This detail cannot be extracted or reduced to data volumes manageable by vector based GIS systems, so it is best left in image format for visual analysis. Only small areas at a time can be effectively analyzed this way, so this type of product would be appropriate for consultations with individual farm operators on soil conservation practices, or for the assessment of land cover/soil condition in known problem areas. It was found that only a minimum of ancillary map layers could be overlayed on these image data before they obscured too much of the underlying detail.

Ground cover maps that could be extracted from the imagery by classification were observed to be well suited to the analysis environment of the GIS. These data could be extracted for large areas (i.e. whole townships or counties), but at a lower level of detail than available through visual interpretation. By definition, a classification is a reduction of the data. In this instance, the data are reduced to a form ideally suited to thematic mapping and spatial analysis in a GIS.

A significant technical issue which influences the development of integrated image/map products is the limited display hardware supported by IAS and GIS vendors. Traditionally, IAS vendors have used raster mapped 24 or 32 bit colour displays, while GIS vendors have used high resolution vector graphics devices, often with restricted colour palettes. The best display system for integrated products permits image display with high radiometric resolution (24 bits) while maintaining good vector graphics capabilities for map information and text. Only a limited number of vendors support these types of displays, which to date has restricted the easy integration of the two data types and continues to be a barrier to the adoption of remotely sensed data to GIS users.

ACKNOWLEDGEMENTS

The authors wish to thank Dr. Brian Brisco of the Canada Centre for Remote Sensing for his critical review of this paper and Mrs. G. Ferguson for producing the paper in its final form.

REFERENCES

Cihlar, J., R. Protz, and C. Prevost. "Soil Erosion Assessment Using Remotely Sensed Data", Proceedings, 11th Canadian Symposium on Remote Sensing, Waterloo, Ontario, June 22-25, 1987.

Landriau, P., and M.J. Manore, 1989. "Integrated GIS and Image Analysis Software: An Application in Vegetation Monitoring", submitted to IGARSS'89/12th Canadian Symposium on Remote Sensing, Vancouver, B.C., July 10-14, 1989.

Ottaway, E., 1988. "A Working, County/Municipal GIS", Proceedings, Geographic Information Systems Seminar, Ontario Ministry of Natural Resources, Toronto, October 3-5, 1988.

Sauchyn, D.J., 1989. "Mapping and Analysis of Soil Degradation in Southern Saskatchewan", presented at the National GIS Conference, Ottawa, February 27 - March 3, 1989

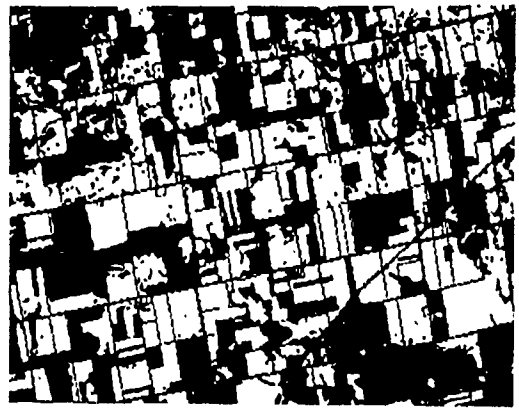


Figure 1: Thematic Mapper Channel 3 overlayed with property boundaries for a portion of Norwich Township, Oxford Co., Ontario. The property lines were found to reinforce the patterns in the landscape and assist visual interpretation.



Figure 2: The results of a classification of spring vegetated fields overlayed by property boundaries (dark lines) and soil units (light lines). Annotation indicating soil type is also displayed.

TABLE 1: Data Available for Norwich Township, 1988

a) Digital Map Data

Data	Source	Format	Content
NTS 1:50000	Canada Centre for Mapping	IGDS	line features from NTS topographic maps
Ontario Base Map 1:20000	Oxford Co. Planning Office	ARC/INFO	line features from OBM topographic maps, DTM
Polaris	Oxford Co. Planning Office	ARC/INFO	property boundaries from Ontario land registry
Agricultural Thematic Maps	Ontario Ministry of Agriculture and Food (OMAF)	ARC/INFO	general land use, tile drainage, soils

b) Satellite Image Data

Sensor	Date	Scene Reference
Landsat TM	07-5-88	18-30 Quad 3
	24-6-88	18-30 Quad 3
SPOT PLA	28-5-88	GRS 612-263
MOS-1 MESSR-1	29-9-88	MRS 167-59

THE USE OF MICROCOMPUTER GEOGRAPHICAL INFORMATION SYSTEM FOR LAND EVALUATION

Li Gang

(Institute of Geography, Chinese Academy of Sciences)
Building 917, Da Tun Road, Outside An Din-men
Beijing, China

ABSTRACT

The present paper describes a land evaluation process by using the powerful microcomputer based Geographical Information System (GIS). It proposes a new model for land quality assessment, the revising coefficient, and an approach to identifying limitations. It shows that the GIS, based on the microcomputer, is practical and easy to use. The paper also suggests the importance of developing a data analysis software package concurrently with the perfection of the GIS.

Key words: microcomputer GIS, land evaluation, evaluation model.

INTRODUCTION

Land is the natural complex on the earth's surface consisting of natural elements such as relief, rock, vegetation, soil and hydrology (C.S.Christian and G.A. Stewart,1953). One of the basic characteristics of land is having production capability. Land evaluation embraces both assessment of its natural characteristics and estimation of its agro-economic ones. The former is called land quality assessment including specific purpose assessment and general purpose assessment. The latter is discussed here including its capability and suitability for farming, forestry and animal husbandry.

Conventionally, land evaluation was made by personnel with practical experience through field surveying and investigation. On the one hand, it requires persons to evaluate and make a map of a large area. On the other hand, because the results depended on the evaluator's intentions to a great extent, it is not easy to compare the conclusions made by different evaluators. The first problem has been solved, along with the development of remote sensing techniques and their applications in land research. But the second one has remained unsolved.

The Geographical Information System has developed rapidly in recent years. It has been applied to many fields. Its four main functions--data collection, data managment, data analysis and information output--supplied an advanced and strong mean for land resources research in land information managment, analysis and automated cartography (P.A.Burrough, 1982). To date, the GIS has been running mostly on expensive minicomputers and mainframe computers (P.F.Fisher et al,1988). They are not suitable for wide use in developing countries. The functions of GIS based on the microcomputer, although its capacities are smaller than the mainframe computer, were quite perfect (G.Schultink,1987). We established a GIS in IBM-XT computer and used it to evaluate the land quality successfully in Huang Tu Liang area (50 km^2), Ping Quan county.

The approaches of land evaluation

are classified as categoric system and parametric system. Both of them have either advantages or drawbacks. We combined both in our work dividing the suitability class using categoric system, then calculating the subclass and limitations using parametric system. The results were quite satisfactory.

METHODS

Grid-digitizing the landtype units and inputting into GIS to store as evaluation units; establishing the mathematical evaluation model and taking it as one of the analysis packages of GIS; comprehensing all land resources information to assess land quality; outputting maps and tables of land evaluation and limitations.

EVALUATION UNITS

Researches of landtype are the basis of land quality assessment. We mapped the landtype in the research area at the scale of 1:50000 by interpretation of Landsat TM and infrared airphoto and registrated it with terrain maps. The land site (it is the lowest grade among land class, subclass and site) was regarded as land evaluation object. Grid-digitizing it, according to the evaluation accuracy , the size of grid cell was defined as 100*100 meter square on earth. Coding and keyboarding it into GIS to store being the unit of evaluation.

EVALUATION FACTORS

Using categoric system, the landtype was divided into four groups: agriculture -suited, forestry-suited, animal husbandry-suited and inadvisable for all of them. We selected several evaluating factors for the first three groups according to such respects steadily influencing land quality. These were geomorphic conditions, soil conditions, fertility, irrigation and limitations. Becaues the growth of crops, trees and herbages need different natural conditions, the factors selected for each group were not at all the same.

The data and information of each factor were obtained by remote sensing interpretation, DTM derivation and sorting out soil survey information. Some of this information was quantitative or semi-quantitative and the others were qualitative. Parts are shown in table1. They all were inputted into GIS, creating the attribute document of evaluation units respectively.

EVALUATION MODEL

There are two main methods of land evaluation in a parametric system which, having been universally accepted, are Californian Storie Index Rating (Storie 1933, 1976)and German Bodenschatzung and its derivatives (Rothkegel 1935; Strzemski 1972). Because the latter requires very detailed knowlege of the land being evaluated, their application is restricted to those countries where such information is available. The Storie Index is widely applied but has certain inherent limitations, especially when the results are compared with a farmer's experience of the land. The author of this paper therefore suggests a new method of evaluation.

According to comprehensive principle (Land quality is formed by comprehensive action of each natural element where characteristics of any single element can not substitute for land quality), differentia principle (Because the effect of each element on land production capability is not same as others, they shouldn't be treated equally during the evaluation process) and limitation principle (The element whose attributes are least suitable for land production capability compared with other elements is the limitation factor), a weighted addition system is proposed:

$$QI=\sum_{i=1}^{n} Wi*CFi \quad n\text{--number of factors}$$

where QI__index of land quality, land will classified into different class of suitability by dividing its scores (see to table2).

Wi__weightness of each factor, it is calculated by Saaty's method.

Table1. Factors and its attributs of agriculture suitable land

ty-\fac-pes \tors	soillay-er(cm)	text -ure	slope	organ. matt.%	P (ppm)	PH	irrig-ation
1	33.8	sandy	0.9	1.01	7.2	7.5	wat.log
2	52.7	lig.loam	0.7	1.13	1.4	7.5	wat.log
3	70.0	san.lig.	2.4	1.09	4.3	7.0	wat.log
4	90.0	loamy	0.8	1.16	4.8	7.8	suffic.
5	61.3	lig.loa.	3.5	1.13	4.2	7.3	suffic.
6	77.0	loamy	1.6	1.07	3.5	7.6	having
7	63.0	sandy	1.8	1.04	3.5	7.3	having
8	96.0	loa/clay	8.0	1.09	3.7	7.4	notgood
9	60.0	lig.loa.	11.8	1.01	3.2	7.5	notgood
10	62.0	loa/clay	13.6	1.11	2.7	6.4	suffic.
11	88.0	san.-lig	10.2	1.05	3.5	7.5	notgood
12	120.0	loa/clay	13.4	0.85	4.0	7.5	nothave
13	60.5	lig.-loa	12.9	1.43	2.0	7.3	suffic.

1__Flat land with sandy soil
2__Shrub flat land with sandy meadow
3__Shrub flat land with meadow soil
4__Mixed grass flat land with meadow soil
5__Leached flat land with drab soil
6__Leached flat land with meadow soil
7__Leached flat land with sandy drab soil
8__Gentle slope land with leached drab soil
9__Gentle slope land with slightly eroded drab soil
10__Valley land with leached drab soil
11__Tableland with leached drab soil
12__Low mountain slope land with leached drab soil
13__Valley land with brown forest soil

CFi__single element evaluating function of each factor. It is constructed with the relationship between factors and land production. That is to say, the most suitable value of each factor's attribute is defined as CF 100, the most unsuitable value is defined as CF 0, between them, the function is lineally related with the factor's attributes.

Table2. Division of land capability class and subclass in agriculture suitable land

suitable classes	suitable subclasses	scores of index
	1	96-100
I	2	93-96
	3	90-93
	1	85-90
II	2	80-85
	3	75-80
	1	70-75
III	2	65-70
	3	60-65
	1	55-60
IV	2	50-55
	3	<50

In the past,it was difficult to compare the land quality belonging to different climate regions. In this paper, a revising coefficient (δ) is proposed:

$$\delta = \frac{K1}{K2}$$

where K1,K2__the effective growth coefficient belonging to climate region1 and 2 respectively.

$$K = \sum_{X_1}^{X_2} Pt * Pw$$

X1,X2__the starting and ending date of effective vegetal period.
Pt__ the effective temperature coefficient (Ref.5).
Pw__the effective precipitation coefficient (Ref.6).

There are different climates in mountains and on plains because of the variation of altitude. The effective vegetal period in mountains in King Quan county starts in the last ten days of May and ends in the second ten days of September whereas that in plain starts in the first ten days of May and ends in the last ten days of September. The coefficient Pt and Pw are shown in table3. The revising coefficient is 0.899.

According to the differentia principle, the limitations can not be pointed out directly through factor characteristics. The "quotient weightness" is then pro-

Table3. Coefficient Pt and Pw within growth period

	May			June			July			August			September		
	FIR	SEC	LAS	FIR	SEC	LAS	FIR	SEC	LAS	FIR	SEC	LAS	FIR	SEC	LAS
Pt	.11	.49	.55	.58	.63	.66	.68	.68	.69	.68	.63	.59	.49	.43	.40
Pw	.25	.25	.30	.35	.40	.50	.75	.85	1	1	1	1	.65	.75	.70

posed to indentify limitations. (Dividing the single element evaluating functions by square roots of their weight respectively, the minimum value of the calculation is limitation.)

$$L=\min(CFi/\sqrt{Wi})$$

Translating the evaluation model into applied subprogramme of GIS, transferring the attribute documents of evaluation units to evaluation subprogramme, operating and outputting various maps of land evaluation and tables of evaluation classes and limitations, the land evaluation map is shown with a scale of 1:50000.

CONCLUSION

1. There are many more functions of the Geographical Information System, for example, beside the comprehensive evaluating and mapping, the single factor evaluating and mapping, multi-factors' overlay will be done if necessary. It shows a wide perspective for deeper research of land.

2. GIS based on microcomputer is handy and easy to extend or rewrite softwares. Because of the limitation of microcomputer RAM, it is just suitable for applying to the works on small area recently.

3. The crucial factor of whether the results of land evaluation using microcomputer GIS are satisfactory is the success of mathematical model of evaluation and the reliability of information obtained. It is even more important to develop the data analysis package than to perfect the whole system.

REFERENCES

1. Burrough,P.A., Principle of Geographical Information System for land resources assessment. Oxford University Press, 1986.

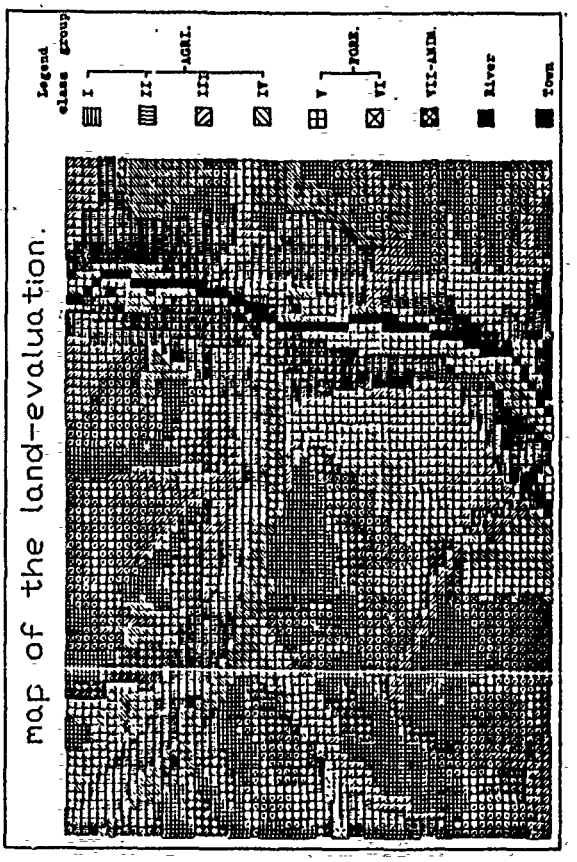
2. Schultink,G. et al, User's Guide to CRIES Geographical Information System, Resource Development Occasional Paper, CRIES Project, Michgan State University, East Lansing, 1986.

3. Fisher,P.F. et al, "The Use of Micro computers in Soil Survey", Soil Survey and Land Evaluation, No.1, 1988.

4. Saaty,T.L., The Analytic Hierachy Process, McGrawHill, Inc, N.Y.,1980.

5. Deng Genyun, "The Resources of Light and Temperature and the Climate Potiential in China", Natural Resource, No.4, 1980. (in Chinese)

6. Lui Hongshun, "The Climate Potiential and the Map of Light Resources Distribution in China", Agricultural Climate, No.1,1985. (in Chinese)



AUTOMATIC DETECTION OF LINEAMENTS FROM LANDSAT DATA

ASHOK KUMAR JOSHI

Heroy Geology Laboratory
Department of Geology
Syracuse University
Syracuse, NY 13244-1070

ABSTRACT

Lineaments are features that represent faults, fractures, joints, contacts, topographic linear ridges, valleys and tonal contrasts, and have a degree of linearity. They are important in understanding the structure and tectonics of a region. They often represent a major structural discontinuity that controls the emplacement of igneous intrusions and may be related to the mineralization or entrapment of oil and gas reservoirs. Lineaments are of primary importance to geologists and can be studied using remotely sensed data. The interpretation and mapping of lineaments is often a subjective matter and varies with individuals' ability and experience.

An automatic lineament detection technique has been developed that is useful in mapping lineaments by image processing Landsat data. The technique involves the enhancement of linear features followed the delineation of these enhanced lineaments through pattern recognition techniques. Lineament maps can be produced using this technique, which eliminates subjective decisions and proves very effective in extracting structural and tectonic information.

Key words: Remote sensing, automatic lineament detection, pattern recognition.

I. INTRODUCTION

A lineament is a mappable, simple or composite linear feature of a surface that is aligned in a rectilinear or slightly curvilinear relationship (O'Leary, et al., 1976). Lineaments are either due to physiographic features (positive or negative) or tonal changes: physiographic features are linear ridges, valleys, faults, fractures, joints and tectonic trends in the rocks: and tonal features mainly reflect changes in lithology representing the contact of two formations. Lineaments may illustrate the crustal structure and are useful in the study of the tectonics of a region. Explorationists are interested in mapping lineaments, since lineaments may be loci of the deposition of economic mineral resources. A number of metallic and nonmetallic mineral deposits have been found along lineaments, where they represent fractures, faults or shear zones and which control the emplacement of metasomatic deposits. Structural traps of the oil and gas can also be interpreted from discontinuities of the surface.

Lineament mapping has applications in the fields of structural geology, tectonics, geophysics, engineering and economic geology. Lineaments have been mapped manually from aerial photographs and satellite imageries for subjective analysis, but during the last decade, mapping has been aided by the digital processing of remotely sensed data. Deconvolution filtering of satellite images in both the spatial and the frequency domains has been done to enhance the linear features, improve image quality and help in recognition and interpretation. However, lineament recognition is subjective and varies from interpreter to interpreter. Lineament mapping depends on visual perception and the ability to recognize patterns in images; thus, experience and knowledge of the area under study are key factors in the identification of lineaments.

Studies on line detection, line extraction and pattern recognition from aerial photographs have been done in the field of urban environment. Pavlidis (1980) proposed a thinning algorithm for discrete binary images to obtain a continuous skeleton of the features in the image. Nevatia and Basu (1980) have described the combination of edge detection and line features extraction to discern highways and runways. Gurney (1980) suggested mapping linear features from Landsat MSS data less

than one pixel wide by a careful selection of thresholds. Wang and Newkirk (1988) designed an expert system to map highways from Landsat -Thematic Mapper data. Holyer and Peckinpaugh (1989) proposed a new edge detection algorithm based on the gray level co-occurrence (GLC) matrix for automatic delineation of mesoscale structure in digital NOAA -7 AVHHR images of the ocean.

The use of remote sensing for geologic studies presents problems that complicate the lineament detection process. Lineaments may be seen as changes in tones of gray, mainly due to the ground cover that controls the spectral characteristics of the lineaments. Cultural features such as highways and field boundaries also interfere with lineament recognition. Shadows may enhance or mask lineaments, and lineaments are often irregular or curvilinear. Thresholding the image at different levels will give different results due to reduction of the gray level range that segregates information in to selected levels. Lineaments may be enhanced using high frequency filters; these, however, sometimes increase the noise in the image, and so a method of enhancing the high frequencies selectively can be useful. A GLC matrix approach helps to control the noise level in the image but is sensitive to very small gradations in the image, thus causing the image quality to deteriorate. Also, this method does not give good results with Landsat MSS data. Variation in the spectral characteristics of the lineaments restrict the use of multispectral classification techniques to detect most features, although they can be applied to highway extraction.

Photo-interpretation of images for lineament analysis is generally based on color, tone, texture, size, pattern and the nature of the surroundings, and these characteristics, along with knowledge and experience, are part of the procedure to recognize lineaments in images. Consequently, an approach based on a set of logical rules can be formulated to interpret the lineaments quantitatively from satellite images.

A methodology for automatic line recognition that takes into consideration the noise, threshold, size and orientation of the lineaments of interest has been developed. Because lineament mapping is a form of pattern recognition, a technique of comparing the lines of known orientation and size with the image is suggested to extract the linear features from the remotely sensed data.

II. ENHANCEMENT OF THE IMAGE

The line detection technique can be considered in three steps. (1) the enhancement of the image, then setting a threshold for the image; (2) conversion of the image to a binary image; and (3) thinning and line linking to permit the tracing of the linear features.

Lineament display is generally enhanced by the application of a high pass filter or by an edge detector. The most common of the edge detectors used are 3x3 convolution kernels, which are sensitive to noise. Common operators are Sobel, Laplacian, Prewitt and Roberts. After the lineaments have been enhanced using high pass filters, a gray level threshold is selected before converting the image to a binary image. Selection of a suitable threshold is critical. A technique proposed by Gurney (1980) may be used to decide the threshold level. The GLC matrix technique (Holyer and Peckinpaugh, 1989) is also effective for suppressing the noise and thresholding the image to detect edges from infrared ocean images. The third step in the procedure has been dealt in different ways by many workers. The detection of lines is followed by thinning and line linking operations, which join the lines and shrink the thick lines to a skeletal structure.

III. DETECTION OF THE LINEAMENTS

Lineaments can be delineated from the binary image by thinning and line joining to obtain thin and regular lines. Such processing extracts lines that may be straight or curved. Most of the curved lines may not be due to the geologic features, unless the region is folded or thrust faulted. The size of the lineament to be detected in the image will be controlled by the resolution of the image (about 79 meters in case of Landsat -MSS data) as well as by the spatial filter that is used to enhance the lineaments in the image. It is not practical to detect lineaments smaller than 2 to 4 pixels in length because that will result in a large number of lineaments that can be neither analyzed nor interpreted. However, regional information about structural style can be extracted using lineaments 5 pixels and more in length.

The proposed method of lineament detection, namely to trace the lineaments from the binary image, takes into consideration the length of the lineament to be detected and its orientation. In this procedure the size and orientation of the lineaments are set up as a matrix and moved to every pixel location as a convolution kernel. The pattern in the kernel is compared with the binary image within the window. When the pattern matches the binary image, the central pixel is assigned to class 1, showing the presence of the lineament, and 0, indicating its absence. The process is repeated for each pixel, and lineaments in the form of lines are traced as an output image. The decision about the presence of a lineament can be modified by reducing the matching from 1:1 to any desired minimum, depending upon the regularity of lines in the binary image.



Figure 1. A subset (24x24km) of Landsat MSS, band 7 image, Salem area, Tamilnadu state, India.

An area (24 x 24 km) was selected from a Landsat-1, MSS, band-7 image of the south Indian craton (Figure 1). The band-7 image was chosen because lineaments are more apparent than in other bands of MSS. The terrain is deformed and eroded, high-grade metamorpi ic Archean crust.

Image enhancement is important because it determines the final outcome of the linear feature extraction. First derivative filters were used to enhance the edges in two orthogonal directions and in two diagonal directions. The enhanced images were merged using a linear combination technique to a single band image (Figure 2). The histogram of the filtered image showed the distribution of the gray levels to be at a few levels, and most of the edge information appeared to be confined to five gray levels.

The enhanced image was classified into five classes using a statistical clustering algorithm based on spectral distances. The five classes in the classified image were similar to the five thresholds in the filtered image. Good results were obtained by selecting a match of one to one. The process was repeated for other directions, and a sequence of passes in the north, northeast, east and southeast directions produced a good lineament map of the whole region. After the image was scanned for lines in different orientations, the results were combined, producing a lineament map of the area under study.

Good results have been obtained by starting with a 5 to 7 pixel lineament length to determine the dominant directions present in the image (Figure 3). This information then could be utilized to detect larger lineaments. In the present image, lineaments about 7 pixels long were detected first and then the size was extended up to 15 pixels (1200 meters); these were summed to produce the final lineament map (Figure 4).

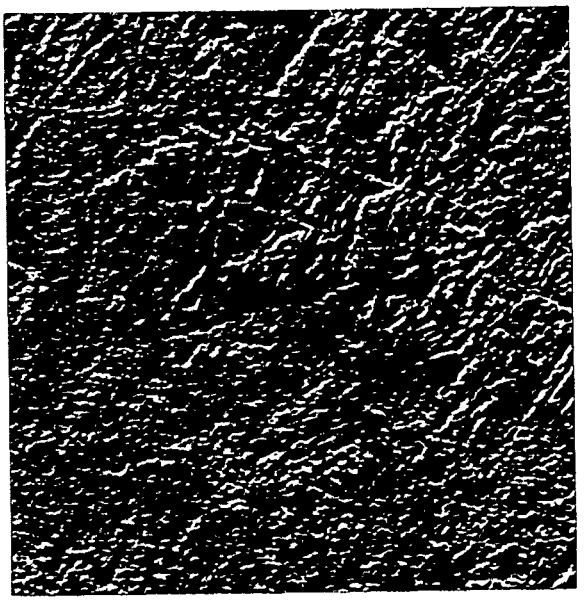


Figure 2. Edge enhanced image of the subset of figure 1. The lineaments were enhanced in two orthogonal and two diagonal directions, using 3x3 edge detectors.

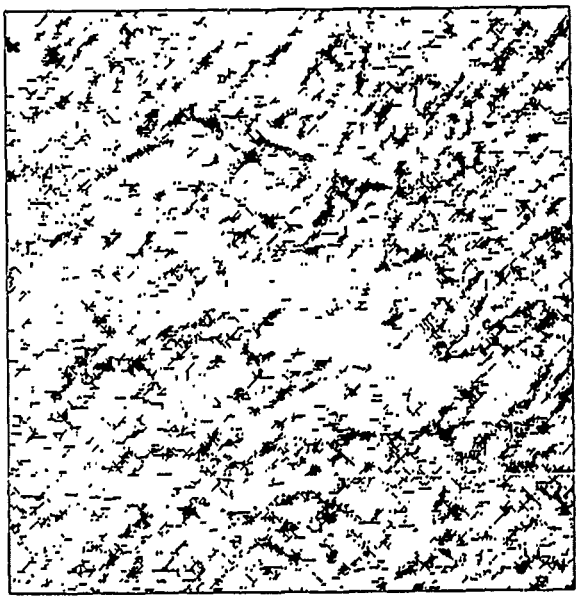


Figure 3. Lineaments detected from the enhanced image of figure 2. Lineaments 7 pixels in length were mapped by four passes in north, northeast, east and southeast directions.

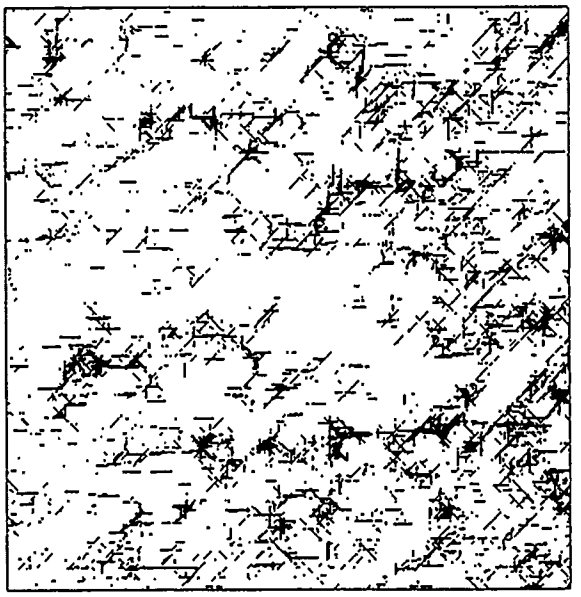


Figure 4. Lineaments detected from the enhanced image of figure 2. Lineaments up to 15 pixels in length were mapped in four passes in north, northeast, east and southeast directions.

IV. RESULTS AND DISCUSSIONS

Automatic detection of lineaments is effective for producing a lineament map from remotely sensed data. This procedure allows considerable flexibility in tracing the direction and size of the lineaments of interest; all lineaments detected were continuous and regular.

Enhancing the lineaments and considering more than one threshold for lineament tracing is useful for extracting significant information. The lineaments may be checked with topographic maps to eliminate the lines that are from cultural features.

The plotted lineaments can be analyzed for density mapping, frequency diagrams and domain outlining. The automatic analysis program may be used with knowledge-based systems, which can directly interact with lineament maps and produce a preliminary interpretation of the area of interest. Because the lineament map for each direction is obtained separately, the use of a program for counting the number of lineaments in each direction is useful for studying the dominant directions and for determining regional stress. Lineament maps obtained from this technique are already in digital form and could be combined with other types of geologic data for decision making.

V. ACKNOWLEDGEMENTS

The author would like to thank Dr. J.E. Robinson, Dept. of Geology, Syracuse University for useful discussions. Additionally thanks to Advanced Graphics Research Laboratory, Syracuse University for computer facilities and the Dept. of Geology, Syracuse University for research facilities. This research was supported by the Ministry of Education, Government of India.

VI. REFERENCES

Gurney, Charlotte M., "Threshold selection for line detection algorithms", IEEE Trans. Geosci. & Remote Sensing, vol. GE-18, pp 204-211, 1980.

Holyer, Ronald J. & Peckinpaugh, S.H., "Edge detection applied to satellite imagery of the oceans", IEEE Trans. Geosci. & Remote Sensing, vol. GE-27, pp 46-56, 1989.

Nevatia, Ramakant & Basu, K. Ramesh, "Linear feature extraction and description", Computer Graphics and Image Processing, vol. 13, pp 257-269, 1980.

O'Leary, D.W., Friedman, J.D. & Pohn, H.A., "Lineaments, linear, lineation: Some proposed new standards for old terms", Geol. Soci. America Bull. vol. 87, pp 1463-1469, 1976.

Pavlidis, Theo, "A thinning algorithm for discrete binary images", Computer Graphics and Image Processing, vol. 13, pp 142-157, 1980.

Wang, Fangju & Newkirk, Ross, "A knowledge based system for highway network extraction", IEEE Trans. Geosci. & Remote Sensing, vol. GE-26, pp 525-531, 1988.

THREE-DIMENSIONAL COMPUTER ANALYSIS AND MODELLING OF REMOTE SENSING-STRUCTURAL GEOLOGIC PROBLEMS

Richard L. Thiessen, Jay R. Eliason*, and Eric R. Rieken.

Geology Department, Washington State University,Pullman, WA 99164-2812, USA (509) 335-4947, 335-8850;
*-also at Geologic Analysis and Consulting Services, P.O. Box 315, Deary, ID 83823.

ABSTRACT

Fault and fracture plane orientations and locations are vital for energy resource exploration and assessment, site stability evaluations, mineral exploration, earthquake hazard assessment, groundwater modelling, and basic geologic mapping. A package of geologic spatial analysis (GSA) computer programs designed to determine crustal fracture geometries is being generated. Crustal structures are determined from analysis of digital data bases developed from topography, lineaments from satellite imagery and radar data bases, faults and joints observed in the field, earthquake foci, gravity, magnetics, and borehole data. The unique power of the GSA techniques being developed are their three-dimensionality; older techniques looked at orientations of structures or their surface traces, but rarely both.

Keywords: Fracture, Lineament, Seismic foci, Spatial Analysis, Digital Elevation Model

GEOLOGIC SPATIAL ANALYSIS

Topographic alignments related to crustal fracture planes were observed and documented by Hobbs (1903) who coined the term lineament. Subsequent studies of lineaments observed in both topography and imagery have proliferated with over 2,000 studies having been published to date. These qualitative lineament analyses are strongly influenced by the illumination direction which creates strong orientation biases (Wise, 1969, 1981; Sawatzky and Lee, 1974; Siegal, 1977; Eliason, 1984). Numerous studies (Podwysocki and others, 1975; Siegal, 1977; Werner, 1979; Bradley, 1983; Rosenfield, 1986; Thiessen and others, 1987, 1989A) have shown that when different operators interpret the same image, significantly different lineament maps are produced. Distinct maps are also produced when the same interpreter works with several image types (Siegal, 1977; Bradley, 1983; Thiessen and others, 1987, 1989A). Finally, previous lineament analyses provide the 2D surface trace of the structure, and not its full 3D orientation in the earth.

These problems stimulated development of the Geologic Spatial Analysis (GSA) research project in order to provide researchers with tools that are not influenced by illumination effects, will yield reproducible results, and provide true 3D spatially referenced analyses of structures. The GSA project is being jointly developed by Washington State University, Geologic Analysis Consulting Services, and Battelle PNL under grants from the U.S. Department of Energy's Office of Basic Energy Sciences. It is focused on conducting comprehensive analyses of regions using geologic data sets that can be referenced by latitude, longitude and elevation/depth (Table 1). Based on these data, the GSA techniques determine the location and orientation of faults and fractures in 3D space in order to develop a regional geologic model for the user.

The GSA routines require input of data in 3D space which will be vector, point, and point/orientation data. Vector data are defined as end points of a vector in 3D space (DEM defined valley bottoms, lineaments, mapped faults, fractures observed on acoustic images, geophysical anomalies, etc.). These end points imply that the intervening data points along the vector are related, i.e. the linear surface trace of a fault or fracture. Point data are local events or measurements that can be located in 3D space, i.e. producing oil or gas wells, seismic hypocenters, mines, or geochemical anomalies. Point/orientation data are measurements of the strike and dip of planar features at a single location. These might include faults and fractures observed in the field or in a borehole.

Vectors are analyzed in 3D space with the coplanar analysis method. A cross product algorithm is used to determine which vectors are coplanar (i.e., lie within the same plane) and therefore are likely to be controlled by a fault, fracture, or bedding plane. The dominant coplanar detections are selected

Input Data Sources:	Output Products:	Analysis Techniques:
- Topography - Remote Sensing - Seismic Foci - Geophysical Data - Underground Acoustic Imaging - Field Studies - Borehole Data - Laboratory Analyses	- 3-D Crustal Fracture Models - 3-D Active Seismic Surface Models - Assessment of Acoustic Images - Stress Field Analyses - Comprehensive 3-D Geologic Model - Maps/Cross Sections - Statistical Summaries/Analysis Trails	- Coplanar - Seismic - Lineament

Table 1. Major components of the GSA analysis system.

automatically or with an interactive program. The dominant planar features can then be displayed in stereo with a 3D output analysis model. Intersection zones of the dominant fracture sets can be identified in order to assess their control on petroleum production, mine locations, and geochemical anomalies. These initial versions of the COPLANEV (V for vector) analysis programs were developed using fracture controlled valley bottoms, which were defined automatically from digital elevation models (DEMs).

Two separate approaches to the determination of alignments of point data are being developed for seismic foci and will be discussed later in this report. Extensions of the present COPLANEV programs to incorporate point/orientation data are currently being developed. The new programs would accept, as input, fracture plane orientations measured in the field as well as in the a borehole. The data will be incorporated into the analysis as a synthesized set of vectors radiating at a 120 degree spacing from the data point's location and within the measured plane's orientation.

Five test areas are initially being used to evaluate and guide the development of the GSA analysis techniques. These include study areas in northeastern Washington state (Eliason, 1984; Eliason and Eliason, 1985; Beaver and others, 1989; Thiessen and others, 1989B) where mapped faults, seismic reflection profiles, and economic mineralization have been related to fractures based upon DEMs. In southern California, seismic foci and topographically defined fractures are being compared to ones observed with a televiewer (Ader and Springer, 1987; Springer and Ader, 1987) in the Cajon Pass DOSECC borehole. At the Big Sandy Gas Field in southwestern West Virginia (Foley and others, 1988), enhanced gas production from Devonian shales has been related to fracture controlled increases in permeabilities. Lineaments and mapped faults have been correlated to DEM fractures in the Paiute Ridge quadrangle at the Nevada Test Site (Thiessen and others, 1987). In south-central Washington, faults and lineaments have been compared to DEM defined fractures (Thiessen and others, 1989A) and earthquake foci (Rieken, 1985).

DIGITAL ELEVATION MODELS

The initial coplanar analysis concept was based on the geologic observation that planes of weakness, such as faults, fractures, and bedding planes, localize erosion and are often followed by topographic lows. The technique involved searching for these lows in a digital elevation model (DEM) (Figure 1A) and then fitting vectors to all straight valley bottoms. The vectors are analyzed in 3D space using the cross product algorithm in order to determine which vectors are coplanar (i.e., lie within the same plane) and therefore are likely to be controlled by a fault, fracture, or bedding plane. The dominant coplanar detections are selected automatically or by using an interactive program and are then be displayed in stereo with a 3D output analysis model (Figure 1B).

Noise in DEMs can bias the detection of valley bottoms. At some sites, DEMs are not available. An alternate program (DVV) has been developed for digitizing valley bottoms directly from topographic maps and inputting the resulting vectors into the COPLANEV program series.

SEISMICALLY ACTIVE PLANES

The identification of seismically active fractures is crucial for understanding the tectonics and seismic risk for a region. Examination of seismic alignments has been mainly 2D with locations projected to a cross section or a map where alignments are visually or automatically picked. Problems arise due to operator bias and choice of the projection plane. We are developing two separate 3D approaches for determination of alignments of seismic foci. The first (Eliason and others, 1989) utilizes the basic coplanar algorithm, which has been expanded to analyze point data. Randomly chosen earthquakes are used in sets of four events. Each set defines pairs of vectors which are input into the coplanar analysis algorithm. The results can be displayed using the same programs developed for the analysis of DEMs. The other approach (Rieken, 1985) uses a 3D counting box algorithm. A series of variable width boxes with a specified strike and dip are placed over a subset of the seismic foci. Boxes which contain relatively high concentrations of earthquakes may represent seismically active surfaces. The strike and dip of the boxes are incremented, creating computer contoured tables which display the frequency of events for each combination of strike, dip, and location. The surface traces of the boxes that contain a higher number of foci can be correlated to mapped structures (Figure 2). The earthquakes which define the maxima in the frequency tables can be statistically examined to determine planar or complex fault surfaces utilizing a multiple linear regression technique. An ANOVA table is calculated to determine what order fault surface best fits the seismic events. Finally, a 3D stereoscopic view of the structure is generated.

LINEAMENTS

A comprehensive lineament analysis system has been written at WSU. The analysis package allows the user to produce various lineament maps, rose diagrams, and maps of rose diagrams. The total amounts of lineaments for each location on the map can be contoured, or the amounts within an operator specified orientation range can be tallied. One of the unique capabilities of our LINEAMENT system is the ability to compare multiple sets of lineaments (Johnson, 1988; Thiessen and others, 1989A). As previously stated, different interpreters working with multiple image types will produce distinct lineament maps. Several studies (Isachsen, 1976; Wise, 1981; Thiessen and others, 1987) have shown that only a portion of lineaments in an analysis actually correlate to mappable geologic structures. Using our comparison routines, we can automatically focus in on those specific structures that are observed the greatest number of times by different interpreters on different images and data types. These are the features most likely to correspond to major structures.

The LINEAMENT program can be used to analyze mapped traces from any data base, such as linear magnetic or gravity anomalies, radar lineaments, and mapped faults. Registration of these to a DEM create vectors in 3D space that can be used as input into the COPLANE analysis sequence. With modifications, it will be able to examine structures found on acoustic imaging data bases as well.

CONCLUSIONS

The GSA tools provide geologists with quantitative 3D structural analyses relevant to exploration and development of fracture-controlled petroleum, groundwater, mineral, and geothermal resources. Seismically active GSA defined fractures are crucial for assessment of seismic hazards and understanding the neotectonics of a region.

One of the major goals of the GSA system is to integrate into an analysis of fault and fracture orientations all data bases and data types that may contain indicators of these structures. In this scheme, a single fault may be simultaneously defined by alignments observed on acoustic images, valleys detected in a DEM, field mapped faults, seismic foci, lineaments from remote sensing images, geothermal

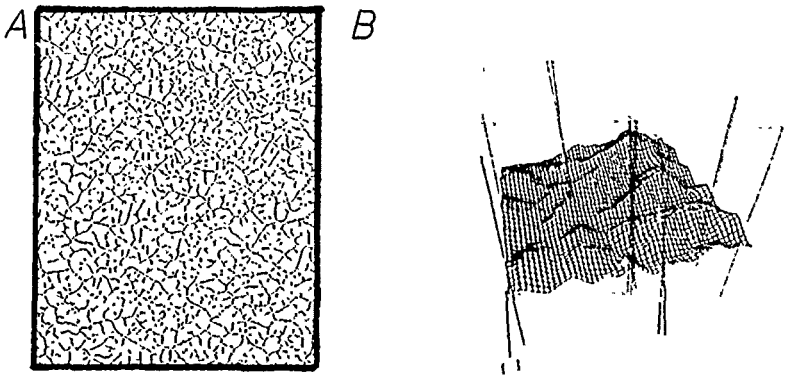


Figure 1. Coplanar output analysis model for a single 1:24,000 scale quadrangle in the Big Sandy Gas Field in southwestern West Virginia. A) Topographic low points used in valley vector analysis. B) Three-dimensional view of the central portion of the quadrangle, showing selected fracture planes identified with COPLANEV. Note the extremely irregular topography. These GSA defined high angle fracture sets are currently being evaluated by DOE's Morgantown Energy Technology Center (METC) staff for correlation to variations of gas seepage from soils and production data from existing gas wells.



Figure 2. Preliminary run of SEISPLN on a seismic foci data set in central Washington. The area is 74 km on a side. A) Earthquake foci locations projected to the earth's surface, B) Previously mapped faults and folds in the test area. C) Surface traces of seismically active planes detected with the counting box routine.

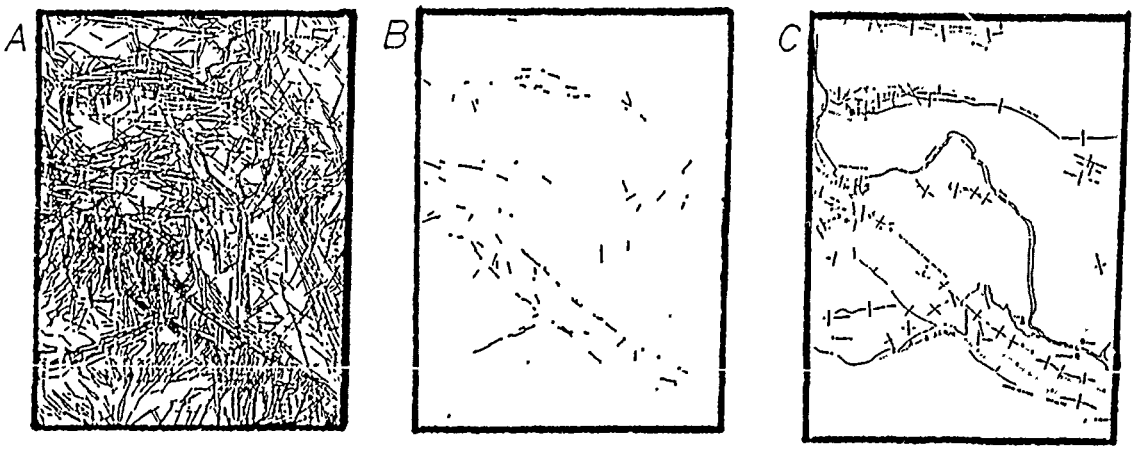


Figure 3. Synopsis diagram for the Hanford, Washington lineament study area. See Thiessen and others (1989B) for more details. A) All linear features seen on a radar mosaic, DEM image, and aeromagnetic maps. The latter includes previous interpretations by other groups as well as WSU interpreters. B) Computer generated compilation map of features within five degrees and 0.8 km of each other. To be included on this map, a single lineament had to be observed a multiple number of times on at least two of the original data bases (radar, DEM, aeromagnetic). C) Mapped faults and folds in the study area.

springs, aeromagnetic and gravity map patterns, geochemical anomalies, mine sites, and locations of high oil and gas production.

REFERENCES

Ader, M.J. and Springer, J.E., "In Situ Stress Project Technical Report Number 1: Televiewer Data Report for the Test Interval, 6000-6250 ft., Cajon Pass Well, California," U.S. Geological Survey Open-File Report 87-289, 1987.

Beaver, D.E., Eliason, J.R., and Thiessen, R.L., "Regional Digital Analysis of Major Crustal Structures in Washington State", Proceedings of the Seventh International Conference on Basement Tectonics, in press, 1989.

Bradley, J.S., "Multi-scale Analysis of Lineaments Glen Canyon area, Utah and Arizona: Part 1 - Lineament Analysis", Proceedings of the Fourth International Conference on Basement Tectonics, pp 245-274, 1983.

Eliason, J.R., "A Technique for Structural Geologic Analysis of Topography", Ph.D. thesis, Washington State University, Pullman, WA, 166 p, 1984.

Eliason, J.R. and Eliason, V.E., "A Comparative Study of Fracture Planes Computed from Topography and Lineaments from Imagery with Structures and Mineralization in the Magnesite Belt of Washington State", International Symposium on Remote Sensing of the Environment, Fourth Thematic Conference, Remote Sensing for Exploration Geology, pp 655-664, 1985.

Eliason, J.R. and Thiessen, R.L., "Geologic Spatial Analysis, a New Multiple Data Source Exploration Tool", International Symposium on Remote Sensing of the Environment, Fifth Thematic Conference, Remote Sensing for Exploration Geology, pp 763-774, 1987.

Eliason, J.R., Thiessen, R.L., and Foote, H.P., "Quantitative Three-Dimensional Structural Techniques: Short Course S36B Notes", Twenty-eighth International Congress, Washington, D.C., in press, 1989.

Foley, M.G., Beaver, D.E., Glennon, M.A., Mroz, T.H., Eliason, J.R., and Thiessen, R.L., "Application of Remote Geologic Analysis to Gas Exploration in West Virginia", Geological Society of America Abstracts with Programs, Vol. 20, p A229, 1988.

Hobbs, W.H., "Lineaments of the Atlantic Border Region", Geological Society of America Bulletin, Vol. 22, pp 123-176, 1903.

Isachsen, Y.W., "Fracture Analysis of New York State Using Multi-Stage Remote Sensor Data and Ground Study", Proceedings of the First International Conference on the New Basement Tectonics, pp 200-217, 1976.

Johnson, L.K., "Computer Analysis of Remote Sensing and Geologic Datasets", M.S. thesis, Washington State University, Pullman, WA, 215 p, 1988.

Podwysocki, M.H., Moik, J.G., and Shoup, W.C., "Quantification of Geologic Lineaments by Manual and Machine Processing Techniques", Proceedings of NASA Earth Resources Survey Symposium, NASA TMX-58168, 1B, pp 885-903, 1975.

Rieken, E.R., "Computer Generated Fault Surface Determinations from Earthquake Foci, Washington State, 1969-1983", M.S. thesis, Washington State University, Pullman, WA, 126 p, 1985.

Rosenfield, G.H., "Determination of Total, Commonality, and Uniqueness of Interpreted Structural Elements from Remotely Sensed data in Alaska", Mathematical Geology, Vol. 18, pp 161-179, 1986.

Sawatzky, D.L. and Lee, K., "New Uses of Shadow Enhancement", Remote Sensing of Earth Resources, Vol. 3, pp 1-18, 1974.

Siegal, B.S., "Significance of Operator Variation and the Angle of Illumination in Lineament Analysis on Synoptic Images", Modern Geology, Vol. 6, pp 75-85, 1977.

Springer, J.E. and Ader, M.J., "In Situ Stress Project Technical Report Number 2: Televiewer Data Report for the Test Interval, 6250-6935 ft., Cajon Pass Well, California", U.S. Geological Survey Open-File Report No. 87-290, 1987.

Thiessen, R.L., Eliason, J.R., Beaver, D.E., and Watkinson, A.J., "Determination of Crustal Fractures in Northeastern Washington Using Geologic Spatial Analysis and Their Correlation to Mineralization", Geologic Society of America Abstracts with Programs, Vol. 21, p 151, 1989B.

Thiessen, R.L., Eliason, J.R., Johnson, L.K., Brougher, C.W., Foley, M., and Beaver, D.E., "Structural Analysis of the Central Columbia Plateau Utilizing Radar, Landsat, Digital Topography, and Magnetic Data Bases", Eighth International Conference on Basement Tectonics, in review, 1989A.

Thiessen, R.L., Johnson, L.K., Foote, H.P., and Eliason, J.R., "Surface Reflectance Correction and Stereo Enhancement of Landsat Thematic Mapper Imagery for Structural Geologic Exploration", International Symposium on Remote Sensing of the Environment, Fifth Thematic Conference, Remote Sensing for Exploration Geology, pp 763-774, 1987.

Werner, E., "Photolineament Mapping: Intra- and Inter-operator Variation", Geologic Society of America Abstracts with Programs, Vol. 11, p 58, 1979.

Wise, D.U., "Pseudo-Radar Topographic Shadowing for Detection of Sub-Continental Sized Fracture Systems", Proceedings of the Sixth International Symposium on Remote Sensing of Environment, Vol. 1, pp 603-615, 1969.

Wise, D.U., "Fault, Fracture and Lineament Data for Western Massachusetts and Western Connecticut", U.S. Nuclear Regulatory Commission NUREG/CR-2292, 253 p, 1981.

EDGE FOLLOWING AS GRAPH SEARCHING AND HOUGH TRANSFORM ALGORITHMS FOR LINEAMENT DETECTION

Jinfei Wang and Philip J. Howarth

Earth-Observations Laboratory, Institute for Space and Terrestrial Science
Department of Geography, University of Waterloo
Waterloo, Ontario N2L 3G1, Canada

Abstract

Most lineament mapping is done visually using enhanced images. To partially automate this procedure, algorithms have been developed to imitate some of the visual rules employed by geologists when mapping lineaments. In this paper, the capabilities of Edge Following as Graph Searching (EFGS) and the Hough Transform algorithms for automated lineament analysis are demonstrated. The EFGS method can be used to extract edge magnitudes and directions to produce an edge image. The Hough Transform identifies straight lines which represent edges.

The extraction procedures are demonstrated using part of a Landsat TM image of the Canadian Shield near Sudbury, Ontario. Results of the automated analyses are compared with the major faults shown on a geologic map of the area. They show that automated interpretation identifies more of the faults than visual interpretation.

Keywords: Lineament detection, Hough transform, Edge following as graph searching, LINDA system.

Introduction

A lineament is a linear topographical or tonal feature on the terrain representing a zone of structural weakness (Williams, 1983). It may be recognized on and interpreted from images and maps. Identification and mapping of lineaments from satellite images is an important use of remote sensing data in the exploration for mineral deposits, since the lineaments may be related to ore deposits.

On satellite images, lineaments usually appear as straight edges, but frequently a lineament may have gaps in it due to occlusions, poor contrast of the lineament with its surroundings or coverage of surface materials. In visual interpretation and mapping of lineaments, geologists use their knowledge and experience to connect edges which are collinear and broken into a series of segments. It is suggested that some of the "rules" that are used by geologists in their image interpretation can be applied in automated lineament extraction from digital imagery. In this paper, the capabilities of Edge Following as Graph Searching (EFGS) and the Hough Transform algorithms for automated lineament analysis are demonstrated.

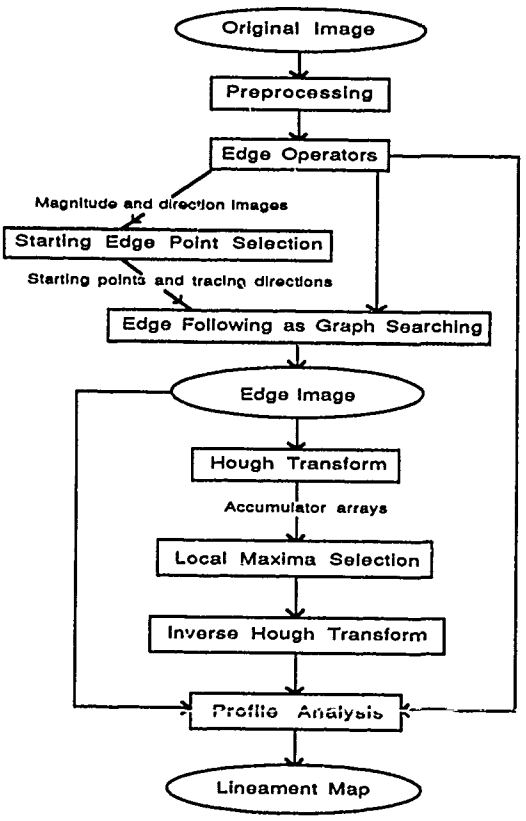


Figure 1 Procedures for lineament extraction and analysis

Methodology

The procedures involved in lineament extraction and analysis are illustrated in Figure 1. This forms part of a LInear-feature Network Detection and Analysis (LINDA) System developed by Wang (1988). The major steps are as follows:

Preprocessing

Median filtering using a filter size of 3 x 3 pixels is used to smooth the image and remove some of the noise. If this were not done, problems of isolating individual lineaments would occur later in the analysis.

Edge Following as Graph Searching (EFGS)

In the second stage, three major steps are involved. First, an "edge operator" is used to obtain the magnitude and direction of the edgedness. This is based upon determining the locations where maximum changes in digital values occur and what the directions of these changes are. The procedure can be applied to detect edges as well as light lines or dark lines, depending upon the appearance of the lineaments in the image. Second, the starting points for the edges are identified. The starting edge point selection algorithm identifies the most prominent edges and it is found that a large number of these correspond to parts of lineaments. Finally, edge following as graph searching is used to trace all the edges on the image. A graph can be formed from each starting edge point. Each arc in the searching graph is associated with a cost. The cost is a function of edge magnitude and direction, as well as the tracing direction. Following is the edge following algorithm (Wang and Howarth, 1987):

[1] Accept all starting points as edge elements.
[2] If there are no more starting points, stop. Otherwise, assign the next starting point as current node.
[3] If there is no neighbour in front of the tracing direction of the current node, goto Step [2].
[4] Compute the cost for the arc connecting the current node to each of its neighbours in the tracing direction. Accept the minimum cost neighbour as an edge element. If no neighbour is accepted, go to Step [2]. Otherwise, assign this neighbour as the next current node and the direction of the arc from the previous current node to this node as its tracing direction. Go to Step [4].

Hough Transform

Hough (1962) proposed an interesting and computationally-efficient procedure for detecting lines in images. It is known as the Hough Transform. Rosenfeld (1982) described a method for replacing the original problem of finding collinear points by a mathematically-equivalent problem of finding concurrent lines. This method involves transforming each of the figure points into a straight line in parameter space. Hough chose to use the familiar slope-intercept parameters, and thus his parameter space was the two-dimensional slope-intercept plane. However, both the slope and the intercept are unbounded, which complicates the application of the technique. Duda and Hart (1972) pointed out that use of the angle-radius rather than the slope-intercept parameters simplifies the computation further.

In this study, the method described by Duda and Hart (1972) was used and modified for geologic use. The procedure involves use of the Hough Transform, the finding of local maxima, application of an inverse Hough Transform and straight line profile analysis. Two parameters are controlled by the analyst. First, a threshold must be set for local maxima selection. This value controls the minimum length of the lines to be detected. In this way, short lines and noise in the image can be eliminated. Second, the analyst can decide on the size of gap that will be closed up if two line segments are identified which lie on the same line. The values for these parameters will vary depending upon the type of image being analysed and the characteristics of the study area.

Mapping and Analysis

Lineament maps are finally produced. Geometric characteristics of these maps can be analysed and measurements of line lengths, line densities, etc. can be made, as well as rose diagram plots to show preferred orientations of lineaments in specific areas.

Application

Study Area and Data

To demonstrate the capabilities of the algorithms for lineament detection and analysis, a study area of part of the Canadian Shield near Sudbury, Ontario has been selected. In Figure 2, a subscene of a Landsat 5 Thematic Mapper (TM) Band 4 image recorded on June 4, 1985 is displayed. This shows part of the exposed Grenville Province. The dominant rocks in this area are middle Precambrian metasediments and an anorthosite suite of intrusive rocks (Ontario Geological Survey Map 2361). Structural control is suggested by the preponderance of elongate lakes confined to a few orientations. Northeasterly-trending structures dominate in the subarea and it was observed that the TM Band 4 image (near infrared) displays the lineaments most clearly.

Analysis

After median filtering of the image, EGFS was applied to trace all the edges on the image. A large number of the starting points correspond to parts of lineaments. A few lie on the hydroline corridor in the northeastern part of the image, but almost no starting points are found on the major highways and railroads. This is because roads and railroads are not very prominent features on the TM Band 4 image when compared with the high contrast that occurs at water/land boundaries.

The edge image is shown in Figure 3. Comparing this image with the original image shown in Figure 2, it can be seen that the edge image contains all the edges at water/land boundaries. One edge, however, is



Figure 2 A subscene of a Landsat 5 TM-Band 4 image (512 by 512 pixels) recorded on June 4, 1985, near Sudbury, Ontario. The area shown is approximately 15 Km by 15 Km.



Figure 3 Edge image after Edge Following as Graph Searching has been applied to the image shown in Figure 2.

the hydroline corridor in the northeast of the area. Other abrupt changes of soil, vegetation or rock type, which could represent lineaments, have also been detected in the edge image.

In Figure 4, an example of one of the lineament maps is shown. In this case, a threshold for local maxima of 75 and a gap of 50 were chosen for the calculations.

Results and Discussion

The lineament analysis procedures described above were applied using four different combinations for the threshold for local maxima selection and for determining the maximum length over which a gap could be closed up (HT-1, HT-2, HT-3 and HT-4 as shown in Table 1).

It is obviously impossible to field check the validity of all the lineaments shown on the resultant plots. However, to determine the capabilities of the algorithms to detect major faults, it is possible to compare the lineament plots with the faults shown on the geologic map of the area. The most detailed map available, however, is at a scale of 1:253,440. In addition, a lineament map produced from visual interpretation of a colour transparency by two geologists was also included in the comparison.

A quantitative comparison was done using the effectiveness index E:

$$E = \frac{L_{TM}}{L_{MAP}} \times 100\%$$

suggested by Strong (1986) where L_{TM} represents the total length of lineaments (cm) common to the lineament map and the geologic map, and L_{MAP} represents the total length of faults (cm) on the geologic map.

Table 1 Effectiveness of the Lineament Maps Compared with the Geologic Map

Imagery	Gap Allowed	Thresh. local max.	Length(cm) on TM	Length(cm) on map	E(%)
Geologic map			1009	1009	100.00
Visual interp.			506	1009	50.15
EFGS			548	1009	54.31
HT-1	40	80	542	1009	53.72
HT-2	40	75	660	1009	65.41
HT-3	40	70	695	1009	68.88
HT-4	50	75	698	1009	69.18

The results are shown in Table 1. As can be seen, visual interpretation identified approximately 50% of the faults shown on the geologic map. The EFGS resulted in an effectiveness index of just over 54%. Best results were obtained with the automated lineament maps where effectiveness indices ranged from 53.72% to 69.18% depending on the threshold values used. It should be pointed out, however, that the lineament maps also indicate many lineaments which are not depicted on the geologic map.

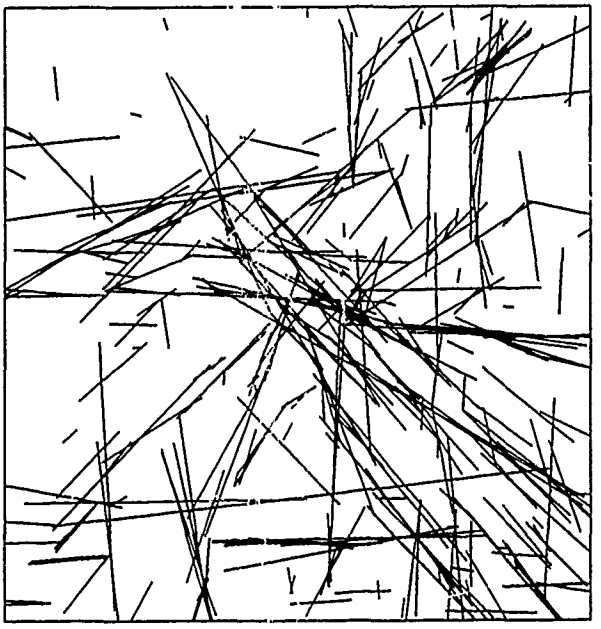


Figure 4 An example of a lineament map after applying Hough Transform to the image shown in Figure 3. A threshold for local maxima of 75 and a gap of 50 were chosen.

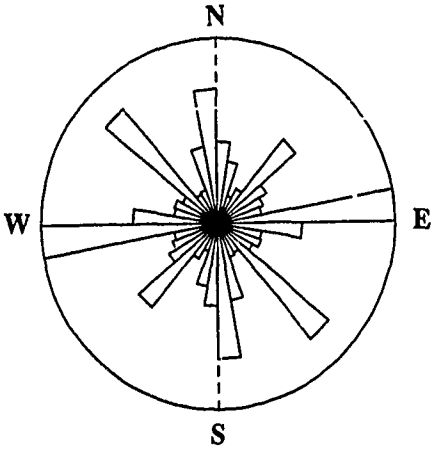


Figure 5 A rose diagram for the subarea, showing major trends for geological lineaments.

In terms of orientation, Figure 5 shows a rose diagram for the subarea. The major orientation from the automated analysis is shown to be east-west, and several faults are shown on the geologic map with this orientation. Other orientations are northwest - southeast, north-south and northeast-southwest. In all cases there is good correlation between the lineament maps and the geologic map.

Conclusions

In this paper, it has been demonstrated that it is possible to develop algorithms to extract lineament information from Landsat TM imagery of part of the Canadian Shield. In a comparison with the geologic map of the area, preliminary results show that automated lineament identification is able to detect more of the faults shown on the geologic map than geologists using visual interpretation.

Acknowledgements

Financial support was provided by NSERC Operating Grant A0766 awarded to Dr. Howarth. Dr. Wang's studies were supported by IDRC, Ottawa.

References

1. Duda, R.O. and P.E. Hart, "Use of the Hough Transform to detect lines and curves in pictures." Communications ACM, Vol. 15, pp. 11-15, 1972.

2. Hough, P.V.C., "Methods and Means for Recognizing Complex Patterns", U.S. Patent 3,069,654, 1962.

3. Rosenfeld, A. and Kak, A.C., "Digital Picture Processing", (Second Ed.), Academic Press Inc., New York, 1982.

4. Strong, W.L., "Remote sensing of seismic lines and other disturbance features associated with oil and gas development: A comparison of Landsat Thematic Mapper and multispectral scanner imagery", Report to the Alberta Remote Sensing Centre, 1986.

5. Wang, J.F., "A New Automated LInear-feature Network Detection and Analysis (LINDA) System and its Applications", Unpublished Ph.D. Thesis, University of Waterloo, p. , 1988.

6. Wang, J.F. and Howarth, P.J., "Methodology for automated road network extraction from Landsat TM data", ASPRS/ACSM Ann. Convention Tech. Papers 1, pp. 429-438, 1987.

7. Williams, R.S., Jr., "Chapter 31 Geological Applications" in R.N. Colwell, Ed. Manual of Remote Sensing, 2nd Edition, Amer. Soc. for Photogrammetry. and Remote Sensing, 1983.

GEOMORPHIC PATTERNS PRODUCED BY THE LAST CANADIAN ICE SHEET: MATCHING THE SCALES OF REMOTE SENSING WITH THE FREQUENCY OF NATURAL VARIATION.

C.D.CLARK & G.S.BOULTON

Grant Institute of Geology,
Edinburgh University,
West Mains Road,
Edinburgh, U.K.

ABSTRACT

Utilisation of six scales of remote sensing in the mapping of ice flow landforms in the area covered by the last Canadian Ice Sheet has produced new insights into its dynamics. The spatial frequency range of natural variation within streamlined glacial landforms is illustrated as a schematic distribution curve. Each remote sensing scale filters these frequencies into discrete ranges, determined by resolution and image size. The relationship between these six overlapping frequency ranges and the frequency distribution curve for ice flow landforms is illustrated in an adapted Venn diagram. This allows visualisation of the proportion of the natural variation that is sampled by each remote sensing scale.

It is shown that incomplete sampling of the frequencies of natural variation can lead to misleading interpretations, and therefore as many scales as possible should be examined. This clearly has implications for remote sensing investigations of many other natural phenomena.

KEY WORDS: Spatial frequency distribution, glacially streamlined landforms, Venn diagram, LANDSAT, AVHRR, air photographs.

1. INTRODUCTION

Utilisation of the full range of remote sensing scales for the mapping of ice flow landforms in the area covered by the last Canadian (Laurentide) Ice Sheet has produced new insights into the dynamics of the Ice Sheet.

This illustrates the importance of using techniques which explore as fully as possible the complete spatial frequency range of natural variation.

2. GEOMORPHIC PATTERNS AND ICE SHEET DYNAMICS

An Ice Sheet flowing across a sediment bed produces streamlined landforms parallel to the flow direction on a number of scales (drumlins, flutes, megaflutes etc.). Such landforms were mapped from air photographs and asserted to be of the Laurentide Ice Sheet at its maximum extent (Prest *et al*, 1968). Two main radial flow patterns were taken to indicate a stable configuration of ice domes.

Our examination of LANDSAT images shows that there are streamlined glacial landforms on a scale larger than can be readily observed on conventional air photographs. This mega-scale glacial lineation pattern frequently occurs in cross cutting relationships with other lineations. If these lineations are invariably parallel to flow, as we believe them to be, they are evidence of different directions of flow at different times and must imply changes in the location of ice domes and a change in their relative strengths. Mapping the area covered by the Laurentide Ice Sheet from LANDSAT images coupled with careful examination of the cross cutting relationships reveals evidence of discrete flows and their relative ages. It has been discovered from this data that the ice dome to the west of Hudson Bay (Keewatin Dome) underwent considerable migration (>500 Km) and that there was movement of the dome to the east of Hudson Bay (Labrador Dome) during the course of the last glacial period (Boulton and Clark, in press).

We attribute this contrast in interpretation (stationary domes versus migrating domes), to the way in which different scales of remotely sensed data sample the frequency of natural variation of glaciogenic landforms in specific ranges.

3. THE FREQUENCY OF NATURAL VARIATION IN GLACIOGENIC LANDFORMS

Streamlined landforms produced by glaciers show a wide range of spatial frequencies. Flutes have typical spacings of 1-10m, with length of 10^2-10^3m; drumlins and megaflutes have spacings from 10-10^3m, and lengths between 10^2-10^3m (drumlins) and 10^2-10^4m (megaflutes); whilst the mega-scale glacial grain shows spacings of 10^3-10^4m and lengths of 10^4-10^5m. We do not believe that these groups represent arbitrary subdivisions of a "white" spectrum of landforms across the whole frequency range, but rather natural frequencies at which spectral power is concentrated. This implies that these landform assemblages are natural generic groupings which reflect specific modes of operation of subglacial processes. As in many other natural systems the scales of operation of a process are determined by natural scales within the environment. Most flutes, for example, are formed by

sediment flow into incipient grooves in the lee of lodged boulders, and thus flute frequency will be determined by boulder frequency. Similarly most theories of drumlin formation suggest a limiting size for drumlins (Menzies and Rose, 1987). Thus we expect a number of interacting scale-dependent processes or single processes controlled by scale-dependent boundary conditions to form a multi-modal spatial frequency distribution. Figure 1. is a schematic diagram showing the principal spatial frequencies which we believe to be characteristic of glacially lineated sediment masses. The fact that names exist to describe different scales of landform (flutes, megaflutes, drumlins etc.) intuitively leads to the same conclusion. Quantitative research from the realm of desert geomorphology provides an analogy (see Figure 2). Wilson (1972) measured wavelengths of aeolian bedforms from the world's major deserts and found that they formed a multi-modal distribution curve with each peak corresponding to a particular bedform type and associated process.

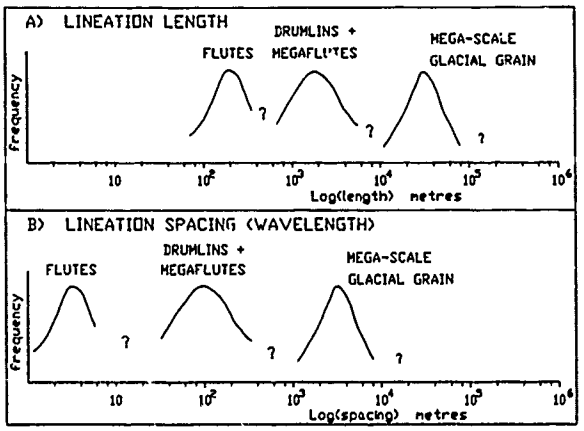


Figure 1.

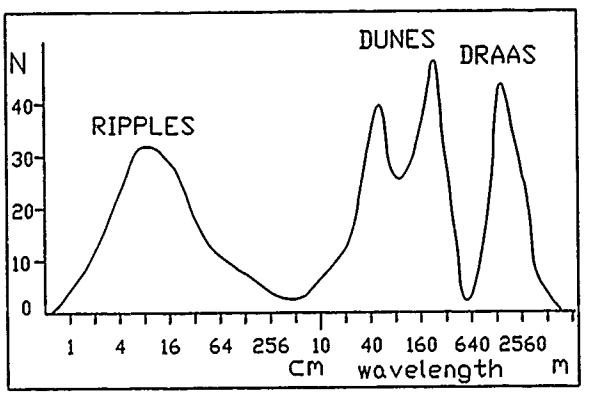


Figure 2.

The derivation of precise spatial frequency distribution curves for ice flow landforms is a non-trivial task involving analysis of all scales of pattern for all orientations over extensive areas. Current attempts to do this at Edinburgh University involve computation of Fourier transforms by 2 dimensional spectral analysis of digital images, and also by optical image processing using binary amplitude filtering (Renshaw and Ford 1983, Renshaw and Mugglestone 1987, Boulton *et al* in preparation). The potency of these techniques is the speed of operation, which is instantaneous in the case of the optical method. Furthermore, polar representation allows the directional component and scale of pattern to be analysed separately and with no *a priori* input.

4. REMOTE SENSING SCALES

Each type of remotely sensed image filters the natural spatial frequencies into discrete frequency ranges which have high frequency boundaries determined by resolution and low frequency boundaries controlled by the instantaneous working area (IWA). An air photograph, for example, at a scale of 1:50,000 has a potential resolution of 1.5m and an IWA of 13 by 13Km, while a LANDSAT MSS image has a resolution of 80m and IWA of 185 by 185Km. Detectable frequencies clearly need to be of wavelengths which are significantly smaller than the IWA. Comparison of Figures 3. and 4. illustrates that each remote sensing scale samples only one of the three natural frequency ranges typical of glacially-streamlined landforms. If the air photo (Figure 3.) is used to map ice flow direction the dominant drumlin pattern will be discerned giving a flow orientation of 255 degrees but the low frequency orientation will be missed. The LANDSAT scene (Figure 4.) shows a very clear orientation at 225 degrees, but the higher spatial frequency orientation at 255 degrees is virtually filtered out. The limited IWA of the air photograph causes lower frequencies to be missed, and the resolution of the LANDSAT image filters out the higher frequencies. Comparison of both scales allows orientations at both spatial frequencies to be discerned. Once low frequency forms are detected on LANDSAT for example, they can often be traced across air photographs where they previously went undetected.

Thus mapping at one scale enables only part of the natural data set, which reflects former ice flow, to be captured. Because of this our analysis of drift lineations in Canada utilised six different scales: air photographs, air photo mosaics, LANDSAT TM and MSS, MSS mosaics and AVHRR imagery. In Figure 5. we relate the overlapping spatial frequency ranges of these data sets to the ice flow data they attempt to capture by means of an adapted Venn diagram. The box represents a theoretically complete data set of ice flow landforms and the circles correspond to the remote sensing scales that were used, with the area of each circle representing a subjective estimate of the proportion of data that can potentially be captured at that scale. The Venn diagram illustrates the interplay between resolution and IWA in influencing the detection of natural frequencies as the scale of sampling increases. Air photographs capture only a fraction of the available information of palaeo-ice flow, and so this fraction is represented by a circle of small area in the Venn diagram. For air photo mosaics, the IWA is considerably increased thus allowing lower frequency landforms to be detected (this increase in data capture is represented by the area shaded 1), but high frequency flute patterns will be lost because of the coarsening of resolution from 1.5 to 15m (area 2 shows this loss of data). The region of overlap between the circles is the proportion of data that is discernible at both scales. The LANDSAT MSS and TM circles are positioned concentrically as they have the same IWA, but the resolving power of MSS is considerably less. Mosaics of LANDSAT MSS cover a 500 by 500Km area which allows very large features to be easily identified. AVHRR

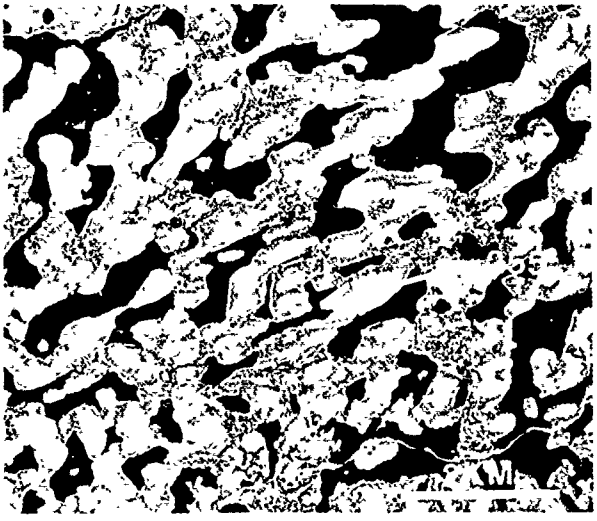


Figure 3.

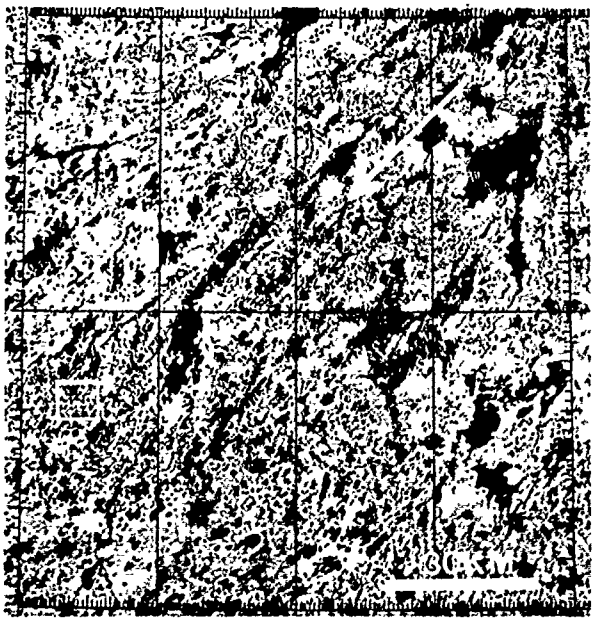


Figure 4.

imagery, of larger IWA and decreased resolution is able to pick out only the lowest frequencies and is particularly valuable in large scale regional interpretations and for detecting very low frequency patterns (eg. Cracknell *et al* 1987). However a great deal of AVHRR derived information is duplicated by LANDSAT mosaics, illustrated in the Venn diagram by a large degree of overlap and small area representing acquisition of new data. The remainder of the area outwith the circles and inside the box indicates the existance of still uncaptured information. The AVHRR scale of 1000's of Kms is at the same order of magnitude as the Ice Sheet and thus excludes the possibility of missing any extremely low frequency ice flow patterns. The missing information is at the field working scale of cm's and metres. Ice flow indicators at this scale tend to be isolated and unlike lineations seen at LANDSAT scale, do not spread contiguously over large distances. Small scale lineations are thus difficult to group into discrete flows and hence infer the gross ice sheet flow patterns.

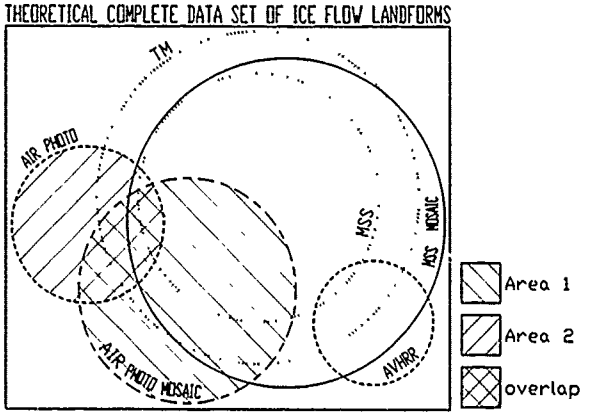


Figure 5.

A further complication to the sampling strategy arises from the use of the human eye for interpreting spatial frequencies. The eye introduces a bias by preferentially enhancing some frequencies at the expense of others (Drury 1987). Looking at the same terrain using different scales this bias is considerably lessened, further emphasising the importance of utilising as many scales as possible.

5. CONCLUSION

The direct matching of remote sensing scales to the schematic natural frequency curve is illustrated in Figure 6. We believe that by using these six scales of remote sensing we have covered the entire range of natural variation of glacially streamlined landforms with

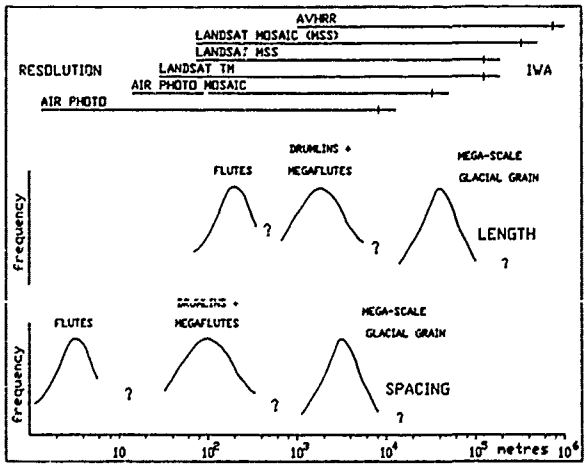


Figure 6.

the exception of very small ice flow indicators (cm's to m's in size), which are too small to provide reliable ice sheet-wide flow patterns. Thus instead of capturing information within only 2 peaks of a tri-modal natural frequency distribution, the whole range of reliable palaeo-ice flow indicators has been sampled (ie. a larger subset of the Venn diagram) and as a result it has been discovered that the major ice domes of the Laurentide Ice Sheet underwent considerable migration.

It has been shown that incomplete use of the frequencies of natural variation can lead to misleading interpretations (ie. a stable position of domes), and therefore as many scales as possible should be examined. This clearly has implications for the remote sensing of many other natural phenomena.

ACKNOWLEDGEMENTS:

We would like to acknowledge the Natural Environment Research council for providing the grant supporting this work and the cost of transporting CDC to the conference. Also thanks to Sarah Conrado for discussion, and help in preparation of this paper and to Dr E. Renshaw and Matthew Bennett for discussion.

REFERENCES:

1) Prest, V. K., Grant, D. R. & Rampton, V. N. - 1968: *Glacial map of Canada. Scale 1:5 million.* Geol. Surv. Canada Map 1253A.

2) Boulton, G. S. & Clark C. D. - 1989: The dynamics of mid-latitude ice sheets through glacial cycles. *Trans. Royal Soc. Edinburgh: Earth Sciences.* In press.

3) Menzies, J & Rose, J. (eds.) - 1987: *Drumlin Symposium.* A. A. Balkema, Rotterdam: pp. 360.

4) Wilson, I. G. - 1972: Universal discontinuities in bedforms produced by the wind. *J. Sed. Pet.* Vol.42(3): 667-669.

5) Renshaw, E. & Ford, E. D. - 1983: The interpretation of process from pattern using two-dimensional spectral analysis: methods and problems of interpretation. *J Royal Statistical Soc. Series C.* Vol. 32(1): 51-63.

6) Renshaw, E. & Mugglestone, M. - 1988: Detection of process from pattern using spectral techniques. *Proc. IGARSS '88 symposium, Edinburgh, Scotland.* Vol. 1: 610.

7) Boulton, G. S., Humberstone, I., Mugglestone, M.; Vass, D. G., & Renshaw, E. - 1989: in preparation.

8) Cracknell, A. P., Johnston, A. C., Boulton, G. S. & Clark, C. D. - 1987: Identification of ancient glacier marks using AVHRR imagery. *Proc. of AVHRR user meeting, Dec 1987,* Oxford.

9) Drury, S. A. - 1987: *Image interpretation in geology.* Allen & Unwin, London: p 19-20.

METHODE ASSISTEE PAR ORDINATEUR D'IDENTIFICATION ET D'ANALYSE DES LINÉAMENTS GÉOLOGIQUES À L'AIDE D'UNE IMAGE ROS (RÉSULTATS PRÉLIMINAIRES)

Jean Daoust[1], Q.H.J. Gwyn[1] et Christian Roux[2]

[1]Centre d'applications et de recherches en télédétection (CARTEL), Université de Sherbrooke, Sherbrooke, Québec J1K 2R1; [2]Département de mathématiques et de systèmes de communication École nationale supérieure de télécommunication (Bretagne), B.P. 832, 29285 Brest, France

RÉSUMÉ

De nombreuses études ont démontré que les zones d'anomalies géométriques des linéaments, identifiées sur les images acquises par télédétection satellitaire, peuvent être reliées à des secteurs présentant un fort potentiel de minéralisation. La présente étude propose dans un premier temps, une méthode d'extraction automatique des linéaments au moyen de filtrages directionnels et de transformations morphologiques. Les traitements sont effectués sur l'image SEASAT des monts Stokes. situés au nord-est de Sherbrooke, Québec. Le résultat obtenu est une image binaire correspondant à la carte des linéaments de cette région. Dans un deuxième temps, on effectue l'analyse des caractéristiques des linéaments à partir d'une carte de linéaments numérisée. Cette étude vise essentiellement à fournir aux géologues une méthode d'identification et d'analyse des linéaments dont les résultats sont normalisés. Ceci afin de permettre, d'effectuer sur une même base, la comparaison des potentiels de minéralisations de différentes zones.

ABSTRACT

Many studies have demonstrated that areas of anomalous lineament geometry appearing on satellite images, can be linked to sectors of potential mineralization. The first part of this study presents an automatic process to extract linear features from a SEASAT SAR image of the Stokes mountains, north-east of Sherbrooke, Québec. The process use directional filtering techniques and morphological transformations. The result of this method is a binary image corresponding to the lineament map of the area. The second part of the study consist in an automatic analysis of the caracteristics of the linear features appearing on a numerised lineament map. The goal of our work is to provide a quick method that gives normalised results, to extract geological linear features and to analize their caracteristics. These results would be helpful to compare different areas on the same base.

1.0 INTRODUCTION

Plusieurs études ont démontré que les zones d'anomalies géométriques des linéaments, identifiées sur les images acquises par télédétection satellitaire, peuvent être reliées à des secteurs possédant un fort potentiel de minéralisation. Les méthodes classiques d'identification des linéaments sur images satellites et d'analyse de leurs caractéristiques (longueur en fonction de l'orientation, densité et entrecroisement) se résument généralement à une interprétation visuelle des données. Les résultats étant dépendant de l'analyste, il est difficile de comparer les caractéristiques des linéaments d'un secteur à celles d'un autre. Afin de normaliser ces résultats, notre étude propose une méthode d'extraction des linéaments assistée par ordinateur et une méthode d'analyse automatique de leurs caractéristiques géométriques.

2.0 EXTRACTION DES LINEAMENTS D'UNE IMAGE SATELLITE

L'objectif de cette première étape est d'élaborer une méthode d'extraction des linéaments assistée par ordinateur. Cette méthode doit être en mesure de traiter les images de façon uniforme afin de produire des résultats normalisés permettant de comparer différentes régions sur la même base. Les résultats escomptés consistent en une production rapide d'une image binaire (deux niveaux de gris) correspondant à la carte des linéaments de la région étudiée.

2.1 SÉLECTION ET DESCRIPTION DU SITE

Le secteur d'étude (20 km x 20 km) est compris dans la bande volcano-sédimentaire des monts Stokes, Québec. Ce site fut choisi parce que sa géologie est bien connue, qu'il suscite un vif intérêt de la part des compagnies minières, que les données satellitaires le concernant sont abondantes et qu'il a déjà fait l'objet de quelques études concernant l'identification des linéaments sur images satellites (Pouliot, 1987; Deslandes, 1988). Ce qui offre la possibilité de comparer les résultats.

2.2 DONNÉES

L'image sur laquelle furent effectués les traitements est une image ROS acquise par le satellite SEASAT-1 le 17 septembre 1978. Le système imageur du satellite opérait en bande L (λ = 23.5 cm), polarisation HH, quatre "look", pixels de 25 m par 25m. Le capteur était orienté vers l'est. L'angle d'incidence variait de 20° à 26° déterminant ainsi au sol un couloir de balayage d'une largeur de 100 km. L'imagerie radar en bande-L a été retenue pour ses aptitudes à pénétrer le couvert végétal et à rehausser les éléments du relief.

2.3 PRÉTRAITEMENTS

Dans le but de comparer les résultats que l'on obtiendra aux cartes de linéaments provenant des interprétations visuelles des images SEASAT et SPOT du même secteur, nous avons effectué les traitements suivants (utilisés par Deslandes (1988)): correction de l'image par rapport à la carte topographique de Sherbrooke (1:50000), rééchantillonnage pour réduire la dimension des pixels à 20m x 20m, application à deux reprises d'un filtre médian de taille 3x3 afin de réduire la granularité. Nous avons de plus extrait une sous-image de 512 x 512 pixels que nous avons décimé à 256 x 256 pixels (un pixel sur deux) afin de réduire le temps de traitement.

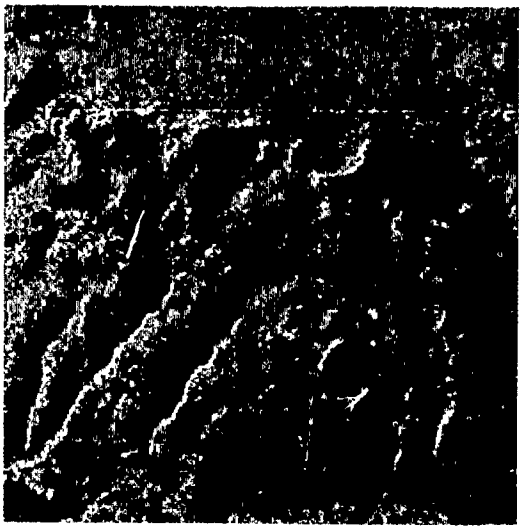


Fig. 2.1 Sous-image des monts Stokes (256 x 256 pixels) sur laquelle les traitements furent effectués.

2.4 MÉTHODOLOGIE

L'extraction des linéaments s'effectue en trois étapes. La première consiste à rehausser l'image pré-traitée de façon à obtenir le plus fort contraste entre les éléments linéaires et leur environnement, la seconde est le seuillage permettant d'obtenir une image binaire et la troisième consiste à affiner les éléments extraits par le seuillage (Fig. 2.2).

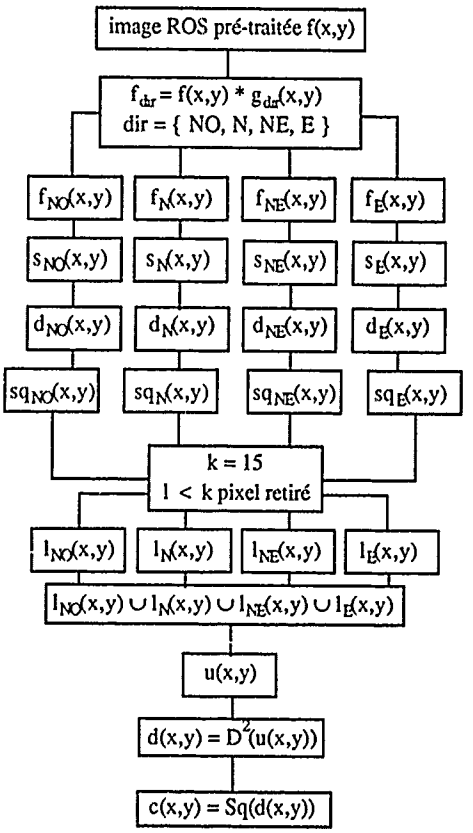


Fig. 2.2 Méthodologie d'extraction des linéaments d'une image numérique.

Le rehaussement de contraste s'effectue au moyen du filtrage directionnel de l'image pré-traitée f(x,y). On effectue le produit de convolution entre l'image et quatres filtres gradients directionnels (Nord-Ouest [NO], Nord [N], Nord-Est [NE], Est [E]) de taille cinq de façon à obtenir quatre images.

$$(f_{NO}(x,y), f_N(x,y), f_{NE}(x,y), f_E(x,y))$$

Chacun de ces filtres rehausse les éléments perpendiculaires à leur orientation et atténue les éléments parallèles. Ainsi un filtre Nord rehaussera les éléments qui sont orientés Est-Ouest et atténuera les éléments orientés Nord-Sud. En effectuant un filtrage directionnel de l'image selon les quatres directions, on se trouve à rehausser tous les éléments présentant une orientation préférentielle.

Les quatres images résultantes sont ensuite seuillées.

$$(s_{NO}(x,y), s_N(x,y), s_{NE}(x,y), s_E(x,y))$$

Les seuillages s'effectuent visuellement de façon à conserver le maximum d'éléments linéaires et le minimum de surfaces. Les images obtenues comprennent une multitude d'éléments plus ou moins linéaires de longueur variable. Afin de reconnecter les éléments distants de seulement un ou deux pixels et qui selon toute vraisemblance devraient appartenir aux mêmes linéaments,une dilatation de taille 1 est effectuée. On obtient alors les images :

$$(d_{NO}(x,y), d_N(x,y), d_{NE}(x,y), d_E(x,y))$$

On effectue ensuite une squelettisation de chacune des images dans le but d'obtenir des éléments linéaires (épaisseur égale à un pixel) et non des surfaces alongées.

$$(sq_{NO}(x,y), sq_N(x,y), sq_{NE}(x,y), sq_E(x,y))$$

Les images de lignes ainsi obtenues comportent encore plusieurs éléments linéaires de faible longueur qui seront retirés au moyen d'un programme effectuant une sélection des éléments selon leur longueur. Les éléments de longueur (l) inférieure à un seuil (k) sont retirés. Ce seuil a été établi empiriquement à 15 pixels. Les images obtenues sont :

$$(l_{NO}(x,y), l_N(x,y), l_{NE}(x,y), l_E(x,y))$$

L'étape finale consiste à effectuer l'union des quatres images de linéaments (Fig. 2.3) de façon à obtenir une seule image correspondant à la carte de linéaments de la région u(x,y).

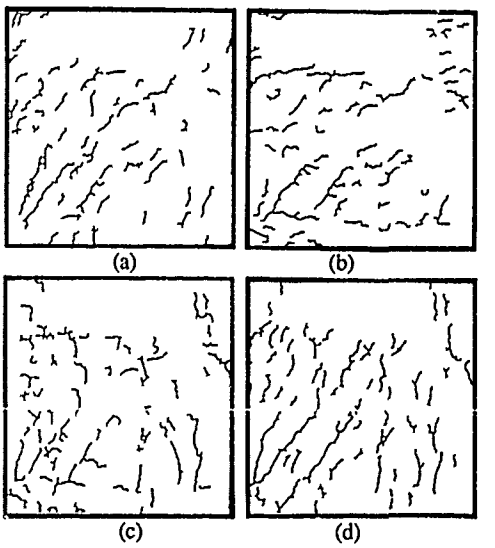


Fig. 2.3 images de linéaments (a) Nord-Ouest ; (b) Nord ; (c) Nord-Est ; d) Est

2.5 RESULTATS

La combinaison des images semble produire une redondance de l'information. Certains éléments paraissent dédoublés. Pouliot (1987) ayant déjà soulevé ce problème conclue qu'étant dans l'impossibilité de savoir si les linéaments qui paraissent redondant correspondent à une même information, il faut les considérer comme différents. En dépit de cette remarque, on a tenté de fusionner en un seul élément les linéaments présentant des comportements similaires et étant très faiblement espacés. La fusion fut effectuée au moyen d'une dilatation (D) de taille 2 suivie d'une squelettisation (Sq). Les éléments qui semblaient dédoublés sont presque tous devenus uniques. Cette étape a du même coup eu pour effet de translater les éléments de quelques pixels par rapport à leur position sur l'image originale. On observe ce phénomène en superposant l'image de linéaments résultante c(x,y) à l'image ROS originale (Fig. 2.4).

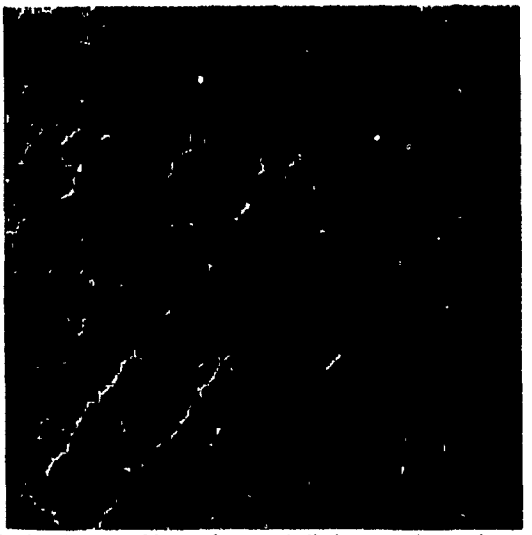


Fig. 2.4 Superposition de la carte de linéaments obtenue à l'image ROS.

L'image binaire obtenue du traitement de l'image ROS présente les principaux éléments linéaires identifiables visuellement. Une comparaison avec une carte des linéaments des monts Stokes, réalisée par photo-interprétation, montre toutefois que certains éléments, présentant un faible contraste avec leur environnement, ne sont pas détectés par la méthode.

3.0 ANALYSE DES CARACTÉRISTIQUES DES LINÉAMENTS

L'objectif de cette seconde étape est de mettre au point un programme d'analyse des caractéristiques des linéaments. Celles -ci sont définies comme étant la longueur en fonction de l'orientation, la densité de linéaments et l'entrecroisements de ceux-ci.

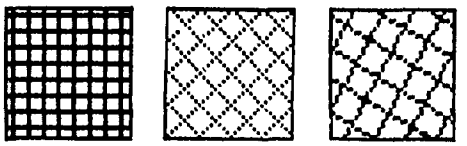
3.1 DONNÉES

Trois images-tests (32 x 32 pixels) générées par ordinateur ont été utilisées pour la mise au point de la méthode d'analyse des caractéristiques des linéaments. Chaque image comportait deux séries de lignes parallèles d'orientations connues et d'espacement régulier. Les lignes avaient une épaisseur de un pixel (Fig. 3.1).

Une carte de linéaments numérisée provenant de l'interprétation visuelle, effectuée par Deslandes (1988), de l'image SEASAT des monts Stokes a également été utilisée pour mettre à l'épreuve la méthode d'analyse des caractéristiques des linéaments.

3.2 PRÉTRAITEMENTS

Une squelettisation de l'image de linéaments des monts Stokes a été effectuée afin de s'assurer que tous les éléments linéaires de l'image possèdent une épaisseur égale à un pixel (Fig. 3.2).



(a) (b) (c)

Fig. 3.1 Images-tests comportant des lignes d'orientation: (a) 0° et 90° [Im1] ; (b) 45° et 135° [Im2]; (c) 30° et 120° [Im3].

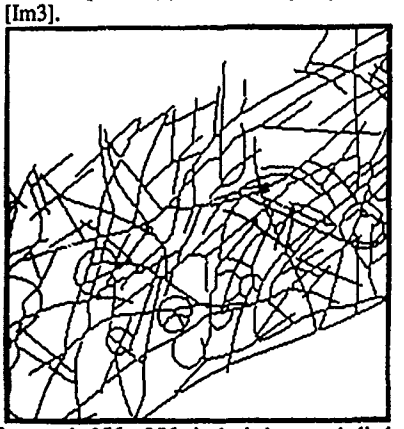


Fig. 3.2 Secteur de 256 x 256 pixels de la carte de linéaments numérisée et squelettisée des monts Stokes.

3.3 MÉTHODOLOGIE

L'analyse des caractéristiques est réalisée au moyen d'érosions directionnelles, de dilatations et de transformations de voisinage sur une image binaire squelettisée correspondant à une carte des linéaments (c(x,y)). On détermine la longueur des linéaments en fonction de l'orientation de la façon suivante. Pour chaque tranche de 15° (l (i)) comprise entre 0° et 165°, on effectue:

$$l(i) = Am^{1}[c(x,y) \cap D^{1}(D^{s(i)}(E^{s(i)}(c(x,y))))]$$

où i = 1, 2, ..., 12

1) une érosion directionelle (E) avec un élément structurant linéaire ou multiponctuel (s(i)) décrivant avec l'axe des ordonnées un angle correspondant à la direction recherchée.

2) une dilatation directionelle (D) avec le même élément structurant.

3) une dilatation isotrope (D) avec un élément structurant de taille unitaire.

4) l'intersection avec l'image originale (c(x,y)).

5) un ébarbulage (Am) afin d'éliminer les composantes parasites.

L'érosion directionnelle (1) a pour effet de retirer tous les éléments linéaires ne présentant pas la direction recherchée. La dilatation directionnelle (2) permet de reconstituer les sections de linéaments présentant l'orientation recherchée mais qui ont été partiellement affectés par l'érosion précédente. La dilatation

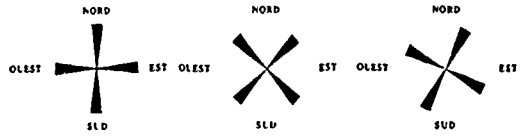
isotrope (3) avec un élément unitaire effectue un épaississement des éléments linéaires restants. L'intersection avec l'image originale (4) permet de reconstituer dans leur totalité les linéaments présentant l'orientation désirée et l'ébarbulage (5) permet d'éliminer les résidus laissés par l'intersection des deux images. Le résultat de ces opération est une image ne comportant que les linéaments présentant l'une des douze directions recherchées. Pour chaque image obtenue, on effectue une sommation des pixels non-nuls, ce qui correspond a calculer la longueur totale des éléments présentant une certaine orientation. À partir de ces images, on réalise une rosace exprimant pour chacune des tranches de 15°, la longueur totale des linéaments possédant cette orientation en fonction de la plus grande longueur rencontrée.

La densité étant définie comme la longueur des linéaments par unité de surface, elle est déterminée en dénombrant les pixels non-nuls présents dans une fenêtre correspondant, au sol, à une zone de 1km par 1km. Les linéaments étant d'épaisseur égale à un pixel, leur longueur est bien égale au nombre de pixels non-nuls. L'entrecroisement des linéaments est déterminé en extrayant les pixels qui possèdent plus de deux voisins non-consécutifs en trame octogonale. On utilise alors un voisinage déterminé par une fenêtre de 3 x 3 pixels.

3.4 RÉSULTATS

La méthode a été en mesure de déterminer la longueur en fonction de l'orientation des éléments de chacune des images-tests avec une très grande précision pour les directions correspondant à celles du maillage de la trame octogonale et une précision satisfaisante pour les autres directions (Fig. 3.3). La densité et l'entrecroisement ont également été déterminés avec une très grande précision comme en témoignent les résultats présentés aux tableaux 3.1 et 3.2.

Cette méthode fut ensuite appliquée à l'image binaire résultant de la numérisation de la carte de linéaments provenant de l'interprétation de l'image ROS des monts Stokes. Les résultats seront éventuellement confrontés avec ceux que l'on obtiendra de l'analyse visuelle de la carte de linéaments.



(a) (b) (c)

Fig. 3.3 Rosaces des linéaments, images comportant des lignes d'orientation : (a) 0° et 90° [Im1]; (b) 45° et 135° [Im2]; (c) 30° et 120° [Im3].

Tableau 3.1

Densité des linéaments (images 32 x 32 pixels)	
Image	corrélations analyse visuelle-analyse automatique
Im1	1.0
Im2	1.0
Im3	1.0

Tableau 3.2

Entrecroisement des linéaments (images 32 x 32 pixels)	
Image	corrélations analyse visuelle-analyse automatique
Im1	1.0
Im2	1.0
Im3	1.0

4.0 CONCLUSION

Les méthodes d'identification et d'analyse des linéaments, développées dans cette étude, permettrons d'obtenir rapidement des résultats normalisés. A partir de ceux-ci, il sera désormais possible de comparer les anomalies géométriques des linéaments de différentes régions afin de déterminer les secteurs à plus forte probabilité de minéralisation.

L'interprétation visuelle de la carte de linéaments obtenue à partir de la méthode d'extraction automatique appliquée à une image ROS des monts Stokes, montre qu'une grande quantité de linéaments ont été détectés. Toutefois, ne possédant pas de méthode rigoureuse d'évaluation des performances de la méthode élaborée, il n'est pas possible de conclure de façon définitive sur son efficacité. Une méthode d'évaluation quantitative des performances sera éventuellement mise au point. La méthode d'analyse des caractéristiques géométriques des linéaments à fait l'objet d'une évaluation plus rigoureuse. Une série d'images-test ont été utilisées à cette fin. Les résultats obtenus permettent d'affirmer que la méthode est en mesure de déterminer la longueur des linéaments en fonction de leur orientation avec un maximum de précision dans les quatre directions du maillage de la trame octogonale. Elle est également capable de déterminer la densité des linéaments et l'entrecroisement de ceux-ci.

La méthode d'extraction des linéaments a été appliquée à une seule image. On prévoit maintenant effectuer les mêmes traitements sur d'autres images présentant des caractéristiques géologiques différentes afin d'évaluer sa robustesse.

5.0 RÉFÉRENCES

Chermant, J. L. et M. Coster, " Précis d'Analyse d'Images ", Editions du Centre National de la Recherche Scientifique, Paris, France, 1985, 521 p.

Deslandes, S.," Méthodologie de détection des linéments géologiques: comparaison des images SEASAT et SPOT ", mémoire de maitrise (en préparation), Département de Géographie et Télédétection, Université de Sherbrooke, Sherbrooke, QC, Canada, 1988.

Pouliot, G., " Intégration de données satellites dans l'évaluation de cibles de prospection minière dans la bande volcano-sédimentaire Ascot-Weedon, Cantons de l'Est, Québec ", mémoire de maîtrise, Département de Géographie, Université de Sherbrooke, Sherbrooke, QC, Canada, 1987, 64 p.

Sekhon, R.," Application of Seasat-1 synthetic aperture radar (SAR) data to enhance and detect geological lineaments and to assist landsat landcover classification mapping ", par Computer Sciences Corporation pour Goddard Space Flight Center, Greenbelt, Maryland, U.S.A., 1981.

LINEAMENT IDENTIFICATION AND THE EDWARDS AQUIFER OF CENTRAL TEXAS

BY
T. KLEEMAN, CENTER FOR SPACE RESEARCH,WRW402
UNIVERSITY OF TEXAS, AUSTIN, TEXAS 78712-1085, USA
&
C. M. WOODRUFF, JR. CONSULTING GEOLOGIST
P.O. BOX 13252, CAPITOL STATION
AUSTIN, TEXAS 78711

ABSTRACT

This paper examines the analysis of a Landsat 4 TM scene for lineament detection in one area along the Balcones/Ouachita trend in Central Texas. Landsat imagery was enhanced to facilitate lineament identification. Comparisons of lineament orientation are made among lineaments from a variety of image sources. A short discussion of edge enhancement techniques employed is included.

INTRODUCTION

Lineaments, in the simplest terms, are straight lines seen from afar on the surface of the earth. Lineaments were described as early as 1863 by Dana [1]. The first systematic use of the concept devolved from works by Hobbs [2] in the early part of this century. Hobbs perceived linear features as being, "significant lines of landscape which reveal the hidden architecture of rock basement..." More recently lineaments have been described in regional tectonic analyses [3]. Lineament analysis has also been given tongue-in-cheek treatment by Wise [4]. Lines abound on the earth and misdiagnosis is always a real danger (e.g., the unwary investigator who maps an electric power grid).

Mindful of the pitfalls associated with such analysis, lineaments are examined in the Austin-San Antonio, Texas area for indications of the grain of the Balcones/Ouachita structural belt. Ongoing water related processes continue to accentuate the ancient structural grain. Water dissolves the near surface limestones and as channels are enlarged more water is introduced, thus causing more dissolution of the strata.

As described above, the Edwards Aquifer has been established as a high-permeability system with anisotropic properties which are suggested by the orientation of the lineaments. The underlying assumption is that preferred underground flow paths are indicated by these features. Locally, precise traces of these features are also sites of enhanced well production. Densities of lineaments and lineament intersection suggest loci of enhanced recharge owing to increased fracture porosity and related dissolution.

Woodruff's earlier work [5] mapped lineaments for a segment of the Edwards Aquifer area (see figure 1) after visual inspection of air photos and standard-product Landsat imagery. The study area (a 512 X 512 subset of the original satellite image) lies south-southwest of Austin, Texas. The Balcones Fault Zone cuts through it in a nearly diagonal orientation. This is a series of en echelon normal faults that have an aggregate movement in this area of several thousand feet. Maximum known displacement across a single fault in the study area is about 600 feet. The main aspect of motion is down toward the Gulf Coast. Most of the tectonic events responsible for this fault displacement probably occurred during the Miocene epoch. Tectonism is no longer active along this trend.

The Balcones Fault Zone is a near-surface manifestation of a deep-seated crustal discontinuity. The Ouachita orogen lies buried beneath the area and forms a hinge between the stable continental interior and the still subsiding Gulf Coast Basin [Flawn, et al, 1961]. Adjustments across this hinge were probably responsible for the down-to-the-coast dislocation of the Balcones fault system and for the emplacement of the igneous plugs that lie along the fault zone. The overall northeast-southwest trend of the Ouachita orogen was probably a major determining factor in the orientation of the main bounding faults of the system [Muehlberger & Kurie, 1956]

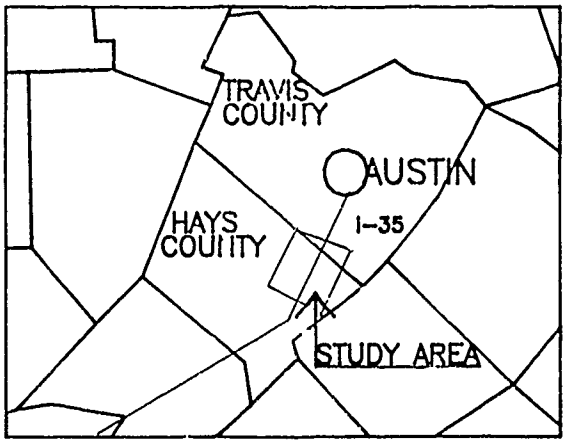


FIGURE 1 LANDSAT STUDY AREA

IMAGERY

Originally the analysis was to employ a digital tape for the Austin, Texas scene (Path - T27, Row -39), 25 January, 1984. Unfortunately, this tape proved to have data irregularities in the columns containing the study area (and in those columns only). Therefore, as an alternative, a later Landsat mosaic, 9 April 1984, was used. This mosaic was scanned (in black and white) with an Eikonix scanner. The digital product was then used in lieu of higher quality tape data. For this reason, the results discussed here should be considered preliminary. Subsequent investigations will employ Landsat digital tapes.

IMAGE ENHANCEMENT

Initially attempts at lineament enhancement employed the local adaptive techniques described by Driscoll and Walker [6]. The results, however, tended to be either too noisy or too dark and were never quite satisfactory.

Because of the earlier work by Woodruff [5], the orientation of the Balcones Zone related faults was anticipated -- the general northeast-southwest grain apparent in the imagery. These earlier findings influenced the selection of directional filters which were rotated in 5 degree increments through the areas 20 degrees to 40 degrees east and west of north. A standard trignonometric filter was employed [7]. For increments west of north the trig functions were positive. For increments east of north the trig functions were negative.

$$\cos A \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} + \sin A \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix} + \text{IMAGE}$$

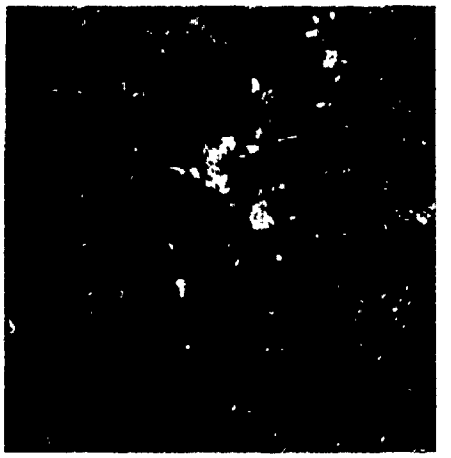


FIGURE 2
SCANNED IMAGE

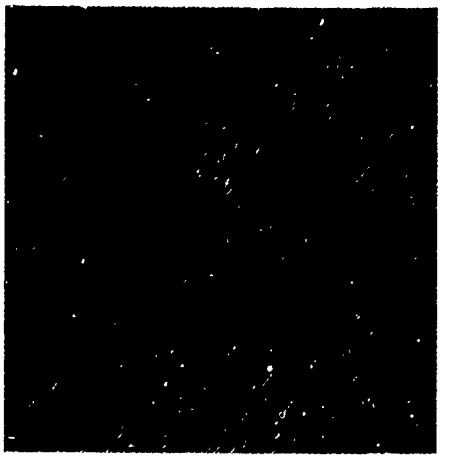


FIGURE 3
FILTERED IMAGE

The results varied little from one increment to the next with the optimal enhancements for visual interpretation occurring at 35 degrees east of north and 35 degrees west of north. Figure 2 shows the area scanned from the Landsat mosaic after stretching. Figure 3 is the same area after application of the directional filters. Figure 4 shows the results in figure 3 after histogram equalization. Equalization improved the visual qualities of the image.

Lineaments were mapped by hand off of hard copies from a Tektronix wax copier. The lineaments mapped on mylar were in turn scanned by the same Eikonix instrument. These were digitized so that length and azimuth could be automatically calculated.

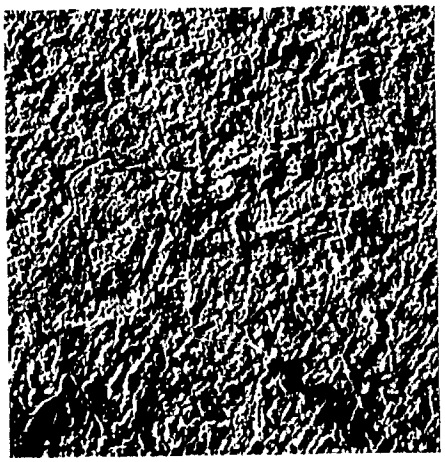


FIGURE 4
FILTERED IMAGE
AFTER EQUALIZATION

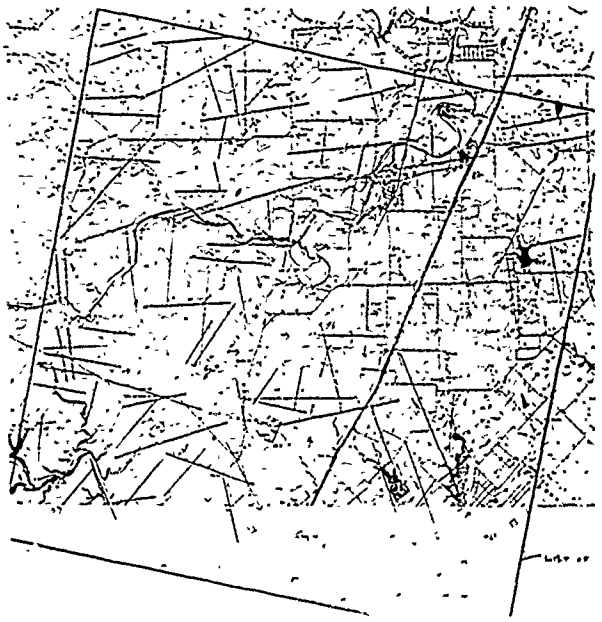


FIGURE 5
MAP OF LINEAMENTS

CONCLUSIONS

The use of scanned mosaics does not render results as satisfactory as those derived from tape. A significant swarm of lineaments in a vector oblique to the prevailing trend of known linear features may underscore the weakness with this scanning approach. Analysis of digital tapes will be necessary to determine whether these lineaments are the products of the scanning process or are indicators of previously unmapped physiographic features.

ACKNOWLEDGEMENTS

This research was supported in part, by the Texas Advanced Technology Research Project (TATRP) under Grants 149700 and 479410. The authors wish to express their appreciation to the Center for Space Research, University of Texas, for extensive and vital support with computation and graphics. The authors are also deeply indebted to the Texas Natural Resources Information Service (of the Texas Water Development Board) for making available the Landsat mosaic used in this study. ("Friends with data are friends indeed.")

REFERENCES

1. DANA, J.D., Manual of Geology: Treating of the Principles of the Science and SDpecial Reference to American Geological History, Theodore Bliss, Philadeophia, 1863.

2. Hobbs, W. H. "Repeating Patterns in the Relief and in the Structure of the Land" Geological Society Bulletin, v.22, p 123-176,1911.

3. Wise, D. U., "Regional and Sub-continental Sized Fracture Systems Detectable by Topographic Shadow Techniques," Proceedings, Conference on Research in Tectonics, Geological Survey of Canada Paper, p 68 - 72. 1968.

4. _____, "Linesmanship and the Practice of Linear Geoart," Geological Society of America Bulletin," v.93, no. 9, p886 -888, 1982.

5. Woodruff, C. M., Jr., "Lineaments and the Edwards Aquifer-Barton Springs Segment, Travis and Hays Counties, Texas," Report to the Edwards Aquifer Research and Data Center, Southwest Texas State University, San Marcos, Texas, 1988.

6. Driscoll, Edward C. J. and Walker, Chris, "Parametric and Non-Parametric Local Adaptive Enhancement," IGARRS, v.1, p. 6.1 - 6.5, 1983.

7. Sabins, Floyd F. Jr., Remote Sensing: Principles and Interpretation, [2nd ed.] W. H. Freeman & Co., 1987.

8. Flawn, P.T., Goldstein, A. Jr., King, R. B., & Weaver, C. E., "The Ouachita System," University of Texas, Bureau of Economic Geology Pub. 6120, 401 p. 1961.

9. Muehlberger, W. R. and Kurie, A. E., "Fracture Study of Central Travis County, Texas: A Preliminary Statement," Gulf Coast Association of Geological Societies Transactions, v. 6, p. 43 - 49, 1956.

REMOTE SENSING OF FRACTURE INDICATORS NEAR THE MOODUS, CONNECTICUT, SEISMIC AREA

Christopher A. Shuman
Department of Geosciences - The Pennsylvania State University - University Park, PA 16802

ABSTRACT

In this reconnaissance study, the Moodus, Connecticut, seismic area was examined primarily for the presence of tectonic fracture indicators. Fracture indicators, lineaments and or fracture traces, are thought to be the surface expression of subsurface fracture zones related to crustal stresses and therefore to provide clues to the cause of this area's anomalous seismic activity. Neotectonic and geochemical information was also evaluated for its possible relationship to the area's endemic seismic events. The purpose of this study was to assess the area's linear fracture indicators and to compare that information to independently obtained in situ and regional stress information. This study was motivated, in part, by the proximity of an operating nuclear power station to the Moodus area.

The area examined covers six U.S.G.S. 7.5 minute topographic quadrangles around the Moodus area. Geologically, it lies in deformed lower Paleozoic units of the Eastern Uplands adjacent to the Central Lowlands (Mesozoic basin) of Connecticut. The imagery used included 1:250,000 SLAR (mosaic), 1.100,000 SPOT (panchromatic); and 1:80,000 and 1:18,000 aerial photography (stereo, black and white). All quadrangles were examined at the three smaller scales while only the Moodus quadrangle was examined at the 1:18,000 scale. Fracture indicators were mapped and data on orientation, frequency, and length were compiled through standard techniques.

The results of the analysis indicate that unambiguous neotectonic geologic features related to the seismic activity were not in evidence. No supersaturated conditions were detected in 33 stream bottom samples and 8 water analyses. Analysis of orientation data from observed fracture indicators revealed prominent orientation peaks at 080-090 and 340-350 degrees. These orientations are essentially parallel and perpendicular, respectively, to the direction of the maximum compressive stress (σ_1) as determined by other researchers from analyses of in situ stress and earthquake focal mechanism information. This suggests a causal association with a thrust stress regime in this area. Lineament frequency was observed to increase in the Moodus quadrangle relative to the adjacent study quadrangles. This suggests that the seismic area may be related to an increase in fracture frequency. Analysis of the multi-scale lineament data revealed an exponential decay in lineament frequency per unit area and a linear increase in average lineament length, both as imagery scale number increases. These results suggest that fundamental scale-dependent relationships may exist for fracture indicator frequency and length data.

KEYWORDS - NEOTECTONICS, FRACTURES, LINEAMENTS, REMOTE SENSING, IN SITU STRESS, MOODUS SEISMIC AREA, CONNECTICUT

INTRODUCTION

The Moodus Seismic Area has been the site of numerous, generally low-intensity, seismic events in recent years and has a record of seismic activity dating from Connecticut's colonial era (Barosh et al., 1982). Scientific efforts to investigate the cause of this anomalous seismic activity have ranged from detailed geologic and surficial mapping of the area (London, 1989; LaFleur, 1980) to monitoring seismic activity and drilling and testing deep research boreholes (Zoback and Moos, 1988; American Geophysical Union, 1988). Some of these efforts have been motivated by the close proximity of the Haddam Neck (Connecticut Yankee) Nuclear Station to an area of possible seismic risk.

SETTING OF STUDY AREA

The area subjected to analysis in this investigation is a block of six U.S. Geological Survey (USGS) 7.5 minute quadrangles. The six quadrangles studied: 1) Middle Haddam; 2) Moodus; 3) Colchester; 4) Haddam; 5) Deep River; and 6) Hamburg; are centered on the town of Moodus and cover nearly 880 square kilometers (see Figure 1). Physiographically, the study area has a rolling, irregular topography characterized by steep hillsides, narrow valleys, and rare, broad uplands. The Connecticut River is the primary drainage and there are two major tributaries, the Salmon River and the Moodus River, which flow generally from the northeast to the southwest.

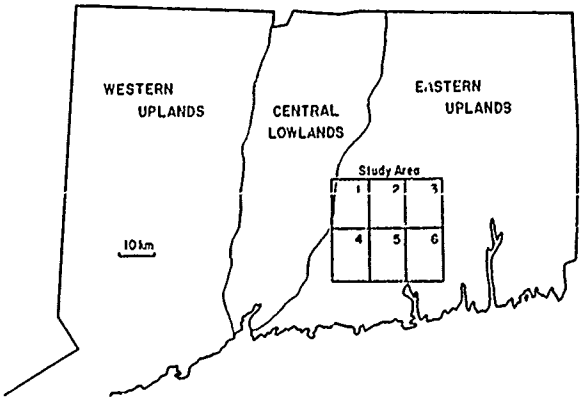


Figure 1, Map of study quadrangles and physiographic provinces

Geologically, the study block lies primarily within Connecticut's Eastern Uplands and extends slightly into the Central Valley physiographic province (Rodgers, 1985). The Eastern Uplands province is composed of severely deformed and metamorphosed, probable lower Paleozoic and possibly older units whose stratigraphic relationships are difficult to ascertain (London, 1989). While the details may be lacking, this sequence: Monson Gneiss - Middletown (Ammonoosuc) - Collins Hill - and Hebron formations; appears to reveal the emplacement of lower Paleozoic units onto an Avalonian basement complex (Lundgren, 1979; Weston Geophysical Corp., 1982).

GEOCHEMICAL STUDIES OF STREAM WATERS

In an attempt to identify areas of anomalous surface water chemistry that might be related to seismic activity or fault movements, streams in the Moodus area were examined and sampled during late summer, low-flow conditions. A review of available literature and outcrops had revealed no unambiguous evidence for or expression of neotectonic, surface disturbances. In some areas, the down-stream presence of travertine-marl deposits and anomalous concentrations of related chemical species has been used to identify fault and fracture zones. These deposits are apparently related to effluent, carbonate-rich, ground water from the dissolution of fault gouge precipitating in turbulent, aerated, stream riffles (Thornton, 1953). Because of the presence of carbonate-bearing gneisses of the Hebron Formation in the study area, it was hoped that fault and fracture features impacting the surface stream network might be identified.

In all, stream sediments were examined at 33 points by visual inspection and 8 water samples were taken from the area (see Figure 2). Despite low water levels, no evidence for precipitation of dissolved species was observed in the streams studied. Analysis of the water samples confirmed this, indicating that the stream waters were undersaturated (see Figure 3). With the exception of

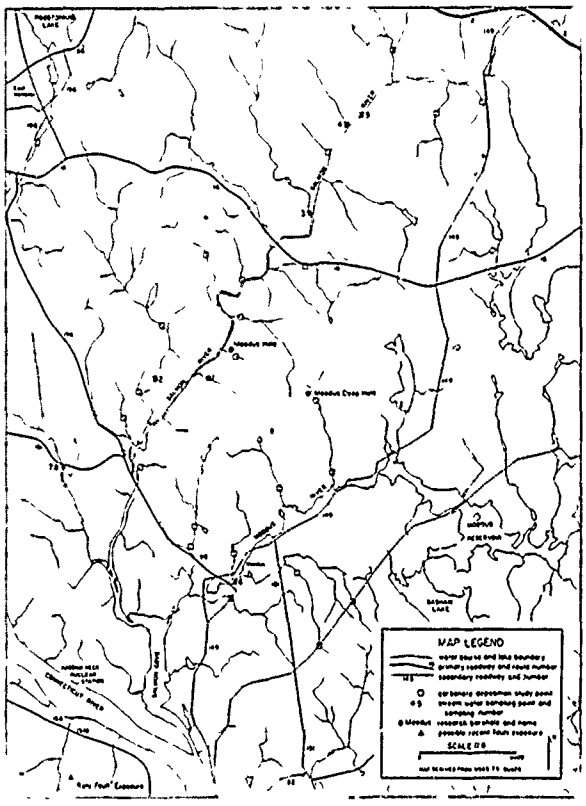


Figure 2, Map of geochemical sampling points and area features

salt contamination from a major highway observed in Sample 5, the stream waters appeared to be well buffered with surprisingly little mineral content. This may be the result of the abscence of dissolved species in effluent ground waters, the precipitation of constituents within the glacial overburden, or low-water levels preventing the discharge of key chemical species.

ANALYSIS OF FRACTURE INDICATORS

In order to characterize the orientation, frequency, and length of linear fracture indicators in and around the six quadrangle study block, lineaments and fracture trace features were mapped on available imagery types. As numerous studies in other areas have correlated these linear topographic, tonal, vegetative, and or combination features with discrete zones of fracture concentration related to tectonic stresses, assessing their distribution in the Moodus area was critical. This need was underscored by the labeling of the Salmon River valley as a lineament-fault by Barosh et al. (1982) and the proximity of both this feature and the Haddam Neck Nuclear Station to the seismic area.

Four different types of imagery were analyzed through the use of techniques adapted from Lattman (1958) and Melser and Earl (1982): a 1:250,000 scale, SLAR (Side-Looking Airborne Radar) mosaic; a 1:100,000 scale, panchromatic SPOT (Systeme Pour l'Observation de la Terre) image with partial coverage of the study block; a series of 1:80,000 scale, high-altitude, black and white, stereoscopic, aerial photographs covering the study block; and a series of 1:18,000 scale. low-altitude, stereoscopic, black and white aerial photographs that covered only the Moodus quadrangle. Through the use of this multi-scale approach, a more complete characterization of lineament features was obtained. However, because of the subjective nature of this technique, it must be noted that these results must be verified by site-specific techniques before they will yield conclusive tectonic or geologic information.

Side-Looking Airborne Radar Imagery

The SLAR imagery used in this study was acquired for the Connecticut Geologic and Natural History Survey (CGNHS) by the USGS in May of 1984. It has an east look direction and at 1:250,000 scale provides an excellent base for analysis of fracture indicators. In all, 48 lineaments were identified within or intersecting the six quadrangle study block. These features, primarily within the Eastern Uplands, ranged in length from 6 to 27 kilometers and may indicate first-order tectonic fractures. Both the Salmon and Moodus Rivers appear as lineaments that (coincidentally) appear to intersect at the Haddam Neck power plant. Also evident on the image is a section of the border fault that separates the two physicgraphic provinces in the study area. In addition, the results of this limited study compare well (about 70% correspondence) to the results of a SLAR lineament map at 1:125,000 scale prepared by Altamura (1985).

The SLAR lineament orientation data are presented in an azimuthal half-rose diagram (see Figure 4). This figure reveals that the features observed in the six quadrangle study block have a primary orientation maximum trending to the north-northwest between 340-350 degrees. Also, a secondary maximum trends northwest at between 300-320 degrees. These results compare well to lineament data in Weston Geophysical Corp. documents (1982) which note the presence of generally northwest trending features. It should be noted that a lack of east-west lineaments may be due to the east-look direction of the SLAR sensor.

SPOT Satellite Imagery

A SPOT 1:100,000 scale scene taken in March 1986 was selected for this study because of its 10 meter pixel resolution in the panchromatic band. As this was prior to the introduction of SPOT's scene-shifting process, only the three northern quadrangles of the study block were covered by the image. While

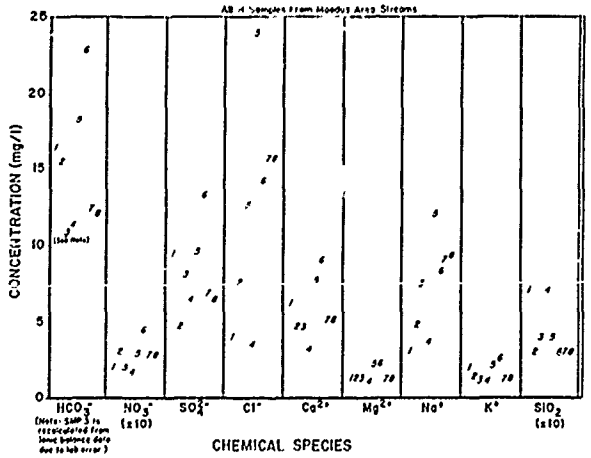


Figure 3, Results of geochemical stream water analyses

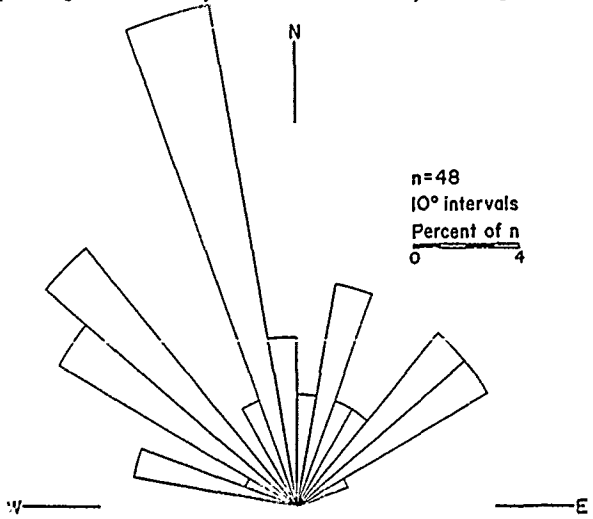


Figure 4, Orientation of lineaments observed on SLAR mosaic

this limited its utility, the scene provided a very detailed view suitable for identifying fracture indicators and other landscape elements. In all, 37 lineaments were noted over the visible portion of the study area. The features were found primarily in the Eastern Uplands and ranged in length from 4 to 10 kilometers.

The azimuthal half-rose diagram of SPOT lineament orientations reveals a different distribution of features than the radar imagery (see Figure 5). This distribution records two primary peaks; a broad one to the north-northeast between 020-040 degrees and a narrow one that coincides with the previously noted SLAR maxima between 340-350 degrees. The northeast orientation peak may be due to the subdivision of the Salmon River lineament-fault into a number of shorter lineaments. The overall change in observed fracture indicators may possibly be the result of the scale change, the relationship between first- and second-order tectonic structures (see Canich and Gold, 1977), or to the resolution of the SPOT imagery.

High-Altitude Aerial Photography

Fourteen 1:80,000 scale, black and white, aerial photographs from a 1980 flight were required to cover the six quadrangle study block. In all, 113 fracture indicators were observed on the stereo pairs of photographs used to analyze this region. They ranged in length from 2 to 13 kilometers and revealed both the Salmon and Moodus river lineaments to be a series of shorter, discontinuous features. Several features correlate closely to the border fault and possibly to a north and northeast trending dike (Rodgers, 1985).

The orientation data for the lineaments mapped on the high-altitude photography reveals two primary maxima: a narrow, northeast peak between 030-040 degrees and a broader, north-northwest peak between 330-350 degrees(see Figure 6). North oriented features between these two peaks are represented by a secondary maxima. The reasonably close correlation between the high-altitude and SPOT data suggests that types of imagery are capable of identifying the same sets of fracture indicators. Any differences between them may be due to the stereoscopic ability of the aerial photographs. This possibility could be tested by comparision with the SPOT satellite's own stereoscopic capability.

Low-Altitude Aerial Photography

Because of study limitations, only the Moodus Quadrangle was subjected to the 1.18,000 scale, lineament analysis. A total of 25, stereo, black and white, aerial photographs provided coverage of the quadrangle. A total of 159 fracture indicators, ranging in length from less than 1 kilometer to almost 3 kilometers, were mapped in this area. In some cases the mapped features appeared to correlate to previously mapped lineaments. These short features are thought to be related to joint and fault traces, lithologic bedding and strike lines, and possibly glacial scour and or foliation directions (Weston Geophysical Corp. 1982).

The orientation diagram for the low altitude data is distinct from the previously observed distributions (see Figure 7). This

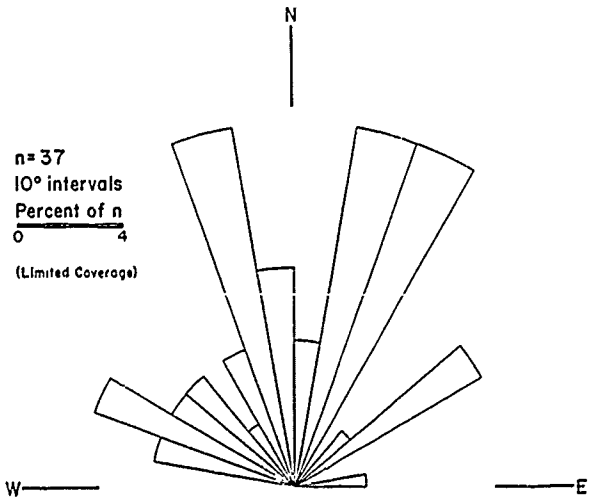


Figure 5, Orientation of lineaments observed on SPOT image

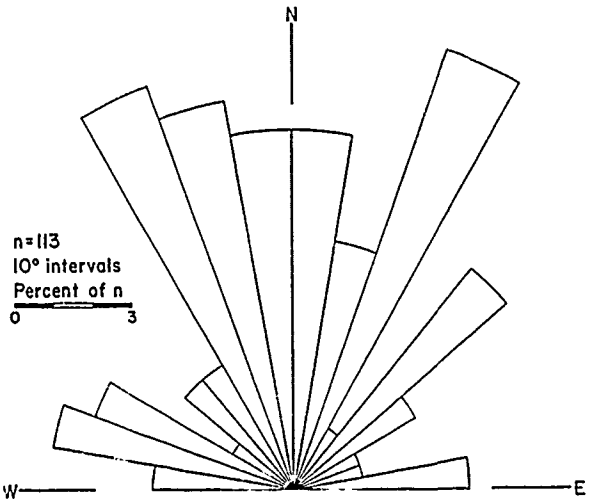


Figure 6, Orientation of lineaments on high-altitude photographs

data set shows a pair of orientation maxima; again one to the northwest between 320-330 degrees and the second in a new direction, to the east at 080-090 degrees. While the first peak is similar to other northwest trending orientation maxima, the second peak is nearly orthogonal to all previously observed lineament orientation peaks. This may be due to these features being related to small tectonic structures or they may indicate the nature of the area's current stress-state. In any case, this eastern orientation peak does not correlate to mapped geologic features.

Discussion of Results

Besides providing information on the overall distribution of fracture indicators across the six-quadrangle study area, the analysis results may provide insights about the low-intensity, seismic activity prevalent in the Moodus vicinity. Results of in situ testing at the Moodus Deep Hole provide detailed information on the state of crustal stress in the area. Hydraulic fracturing tests and borehole televiewer surveys indicate that: the direction of the maximum horizontal compression (σ_1) is essentially east-west at 086 degrees; the least principal stress (σ_3) is due to overburden pressure and is vertical; and the area appears to be under a reverse-thrust regime capable of causing low intensity events along faults oriented north-south (Zoback and Moos, 1988). This information is supported by unpublished focal mechanism solutions for a number of recent Moodus area earthquakes.

These in situ measurements are also supported, less precisely, by the results of the fracture indicator analysis. A plausible explanation for the singular lineament orientation peak (080-090 degrees) observed on the low-altitude aerial photographs is that these features represent fractures opened or forming under the current, in situ stress field. In addition, with north-northwest

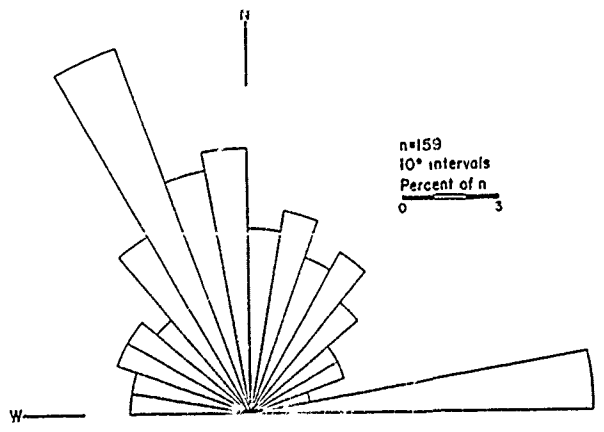


Figure 7, Orientation of lineaments on low-altitude photographs

(340-350 degrees), and to a lesser degree north to northeast, lineament orientations dominating the SLAR, SPOT, and high-altitude surveys, it is possible that these observed features reflect fractures affected by the reverse-thrust stress regime.

Determining the number of fracture indicators observed within or intersecting each quadrangle provides another indication as to the reason for endemic seismic activity in the Moodus area. Each imagery analysis that examined more than one study area quadrangle observed an increase in number of lineaments over the Moodus Quadrangle (see Figure 8). This measure of lineament "frequency" suggests that the Moodus area may be more fractured than adjacent areas. It is uncertain whether this contributes to or is the result of the increased seismic activity.

Results that are less specific to the Moodus area but of potentially greater significance are revealed when the results from all four scales of imagery are examined together. Figure 9 shows the observed relationship between the average number of lineaments intersecting a quadrangle and the scale of the image used in the analysis. The plot shows an exponential decrease in lineament frequency with increasing imagery scale number. This intuitively makes sense as it indicates that with an increase in size of a lineament feature, fewer features of that size will be observed in a unit area. It should be noted that as only one quadrangle was examined at the 1:18,000 scale, this data point is only an estimate. An additional relationship is illustrated in Figure 10. When the average lineament length observed is plotted against imagery scale number, a distinct, positively sloped, linear trend is revealed. This suggests that there may be an appropriate scale of imagery for the observation of a particular size of fracture feature. While both these observed relationships may be a function of imagery resolution and or human perception, they may also relate to the fundamental causes of lineaments in geologic environments.

CONCLUSIONS

The principal conclusions that can be drawn from this study are: 1) no distinctly positive evidence of neotectonic disturbances related to the seismic activity were observed; 2) the limited geochemical studies of local stream waters produced no indication of effluent ground water related to fractures or the dissolution of carbonate in the subsurface; 3) multi-scale lineament analysis revealed varying distributions of fracture indicator orientation, frequency, and length; 4) lineament orientation peaks that are concentrated in the east (080-090) and north-northwest (340-350) may correlate with in situ measurements recording an east-west trend of maximum compression and a north-south, reverse-thrust, crustal stress regime; 5) an increase in lineament frequency over the Moodus Quadrangle suggests that the seismic area may be more fractured and thus more subject to subsurface disturbance; 6) average lineament frequency per quadrangle block shows an exponential decrease with increasing imagery scale number; and 8) average lineament length increases linearly with increasing imagery scale number. Additional information specific to fracture indicators and earthquake activity in the Moodus area could be gathered by further study; however, confirmation of the indicated scale-dependent relationship for lineament frequency and length data may prove to more valuable in the long run.

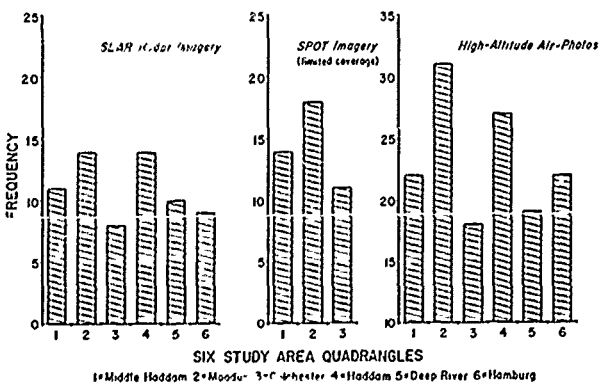


Figure 8, Areal lineament frequency by quadrangle

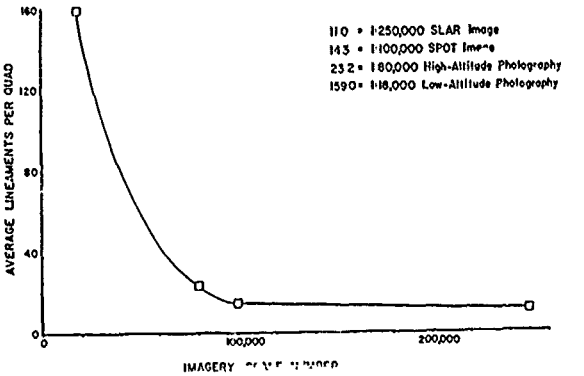


Figure 9, Observed relationship between average lineament frequency per quadrangle and imagery scale number

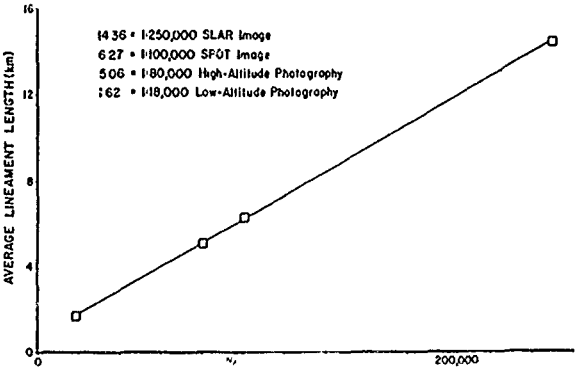


Figure 10, Observed relationship between average lineament length and imagery scale number

REFERENCES CITED

Altamura, R.J., 1985. Preliminary radar (SLAR) map of Conn., Open File Map, Conn. Geologic and Natural History Surv.

American Geophysical Union, 1988. Earthquakes, Geophysics, and Geology near Moodus, Conn., Spring Meeting Program, Sessions T51B and T52B, May 20, Baltimore, Maryland, pp. 74 and 79.

Barosh, P., D. London, and J. de Boer, 1982. The structural geology of the Moodus seismic area, south-central Conn., in Guidebook for Fieldtrips in Conn. and South Cen. Mass., eds., Joesten and Quarrier, #5, Conn. Geol. and Nat. Hist. Surv., pp. 419-451.

Canich, M.R. and D. Gold, 1977. A study of the Tyrone-Mt. Union lineament by remote sensing techniques and field methods, Off. of Rem. Sen. of Earth Res. Rep. 12-77, Penn State Univ. 59p.

Lafleur, R.G., 1980. Investigations of possible earthquake related deformation of glacial overburden deposits in the Moodus - Haddam area, Conn., Unpub. report, Renss. Polytech.Inst., 16p.

Lattman, L.H., 1958. Technique of mapping geologic fracture traces and lineaments on airphotos, Photo. Eng., vol. 19, pp. 568-576.

London, D., 1989. Bedrock geologic map of the Moodus Seismic Area, Conn., Rep. of Invest. 11, Conn. Geol. and Nat. Hist. Surv.

Lundgren, L., Jr., 1979. The bedrock geology of the Haddam Quad., Quad. Report 37, Conn. Geol. and Nat. Hist. Surv., 44p.

Meiser, E. and T. Earl, 1982. Uses of Fracture Traces in Water Well Location, Off. of Water Res. and Tech., Dept. of Int., 55p.

Rodgers, J., 1985. Bedrock geological map of Connecticut, Natural Resources Atlas, Conn. Geol. and Nat. Hist. Surv.

Shuman, C.A., 1988. Neotectonic, geochemical, and lineament studies of the Moodus Seismic Area, Unpub. report to the Nuclear Reg. Comm., Contract No. 04-85-111-01, 56p.

Thornton, C., 1953. The geology of the Mt. Jackson Quadrangle, Virginia, Unpublished Ph.D. diss., Yale University, 211p.

Weston Geophysical Corporation, 1982. Geological and seismological studies, Haddam Neck Nuclear Station, Parts I and II, Unpub. report to NE Utilities Service Co., 104p. and 66p.

Zoback, M.D. and D. Moos, 1988. In situ stress, natural fracture and sonic velocity measurements in the Moodus, Conn., scientific research well, Unpub. report to Woodward-Clyde Consul., 90p.

ORIGIN AND SIGNIFICANCE OF HEMAVATHI - TIRTHAHALLI MEGA-LINEAMENT - A CONCEPT

K GANESHA RAJ

National Natural Resources Management System
Indian Space Research Organisation
Anthariksha Bhavan, Nagashettyhalli
Bangalore, India

ABSTRACT

Landsat MSS False Colour Composite on 1:537,000 scale has been interpreted to study the lineaments in the study area (Frame 145-051 covering Hassan, Chikmagalur, Kodagu, Shivamoga, Mysore, Dakshina Kannada, Tumkur, Chitradurga Districts of Karnataka and Kasaragod District of Kerala). The lineaments were grouped into 6 classes based on their length. These six classes were divided into 3 groups and rose diagrams of the lineaments based on their direction were plotted. The most prominent lineament is about 190 km long and is named Hemavathi-Tirthahalli mega-lineament for easy reference. The direction of this mega-lineament is almost parallel to the west coast. This mega-lineament with considerable length and almost parallel orientation to the west coast indicates that it might have developed during the West Coast Fault period. Major lineaments like Yagachi, Tunga and a good number of other lineaments are also parallel to the west coast and might have developed during the same period. As most of the earthquakes (between 1967 and 77) in Southern Peninsula occurred in the vicinity of NNW-SSE to NW-SE trending mega lineaments, the Hemavathi-Tirthahalli mega-lineament should be studied to understand its seismic status.

Key Words : Mega-lineament, West Coast Fault, Rose Diagram, Landsat, Earthquake

INTRODUCTION

Lineament studies have become an important step in analysing the structural and tectonic aspects of an area. Lineament studies are useful in groundwater, mineral, oil explorations and in engineering geological applications as lineaments may represent fractures, faults, shear zones etc. Using Landsat MSS band 5 image on 1:1000,000 scale Srinivasan and Sreenivas (1977) had mapped the lineaments in the study area. In the present work Landsat 5 MSS False Colour Composite on 1:537,000 scale has been used to study the Origin and Significance of Hemavathi-Tirthahalli mega-lineament and surrounding lineaments.

STUDY AREA

Landsat 5 MSS False Colour Composite of a scene 145-051 covering Hassan and Chikmagalur Districts fully and parts of Dakshina Kannada, Kodagu, Mysore, Tumkur, Chitradurga and Shivamoga districts of Karnataka and Kasaragod district of Kerala, has been studied.

METHODOLOGY

Landsat MSS False Colour Composite has been interpreted visually to map the lineaments in the area. The lineaments were grouped based on their length into 6 classes i.e lineaments having length <10 km, 10-20 km, 20-30 km, 30-40 km, 40-50 km and >50 km. These classes were divided into 3 groups (1) Lineaments of length <10 km (2) Lineaments of length 10-20 km (3) Lineaments of length >20 km and rose diagrams based on their direction were plotted.

RESULTS & DISCUSSION

The lengthwise classification of the lineaments showed that the number of minor lineaments (281) having length of <10 kms is more than that of any other class. The lineaments with length 10-20 km is also considerable i.e 64, whereas lineaments with length >20 km are only 20. Among the last category those between 20 to 30 km are 12, 30 to 40 km are 4, 40 to 50 km are 2 and >50 km are only 2, thus making upto a total of 365 lineaments in all. The Hemavathi- Tirthahalli mega-lineament with a length of about 190 kms is prominently seen on the image. This is the longest of the lineaments in the area. The other major lineaments in the area are (1) Yagachi - 49 kms (2) Vedavathi - 50 km (3) Tunga - 33 km (4) part of Shimsa - 38 km (5) and part of Cauvery - 40 km (Fig.1).

The rose diagram of lineaments having length <10 km i.e minor lineaments showed the prominent direction between N85°E - S85°E. The next prominent direction is N45° - 55°E (Figure-2a) whereas the rose diagram of lineaments with length 10-20 km has two prominent directions (1) N85°E - S85°E (2) N15° - 25°W (Figure-2b). The rose diagram of major lineaments i.e having length >20 km clearly showed that the prominent direction is N25° - 35°W. The other prominent direction is N75° - 85°E (Figure-2c). The rose diagrams gave clear picture about the orientation of lineaments in the area. The major lineaments have prominent trend almost parallel to the west coast, whereas the minor have their prominent trend in the E-W direction. The Hemavathi-Tirthahalli mega-lineament, Yagachi, Tunga major lineaments and one set of lineaments with 10-20 km length are almost parallel to the west coast. The study of rose

diagrams of major (> 20 km) and intermediate (10-20 km) lineaments gives a clear idea that majority of these lineaments have a trend almost parallel to the west coast. Especially the Hemavathi-Tirthahalli mega-lineament with a length of about 190 km is running almost parallel to the west coast. It is inferred that all these might have developed because of the major West Coast Fault.

The direction of west coast in general is NNW-SSE. The presence of faulting along the west coast has been discussed by many geologists. They believe that the escarpment of the Western Ghats parallel with the Malabar coast has been formed by faulting (Wadia, 1975). The dislocation might have been in the nature of step-faults. Step-faulting will explain the three different levels, viz. the Deccan Table land, the coastal plain (often of rocky surface) and the submarine shelf, and the two linking steep slopes - the scarped face of the Ghats and the cliff line at the shore (Ahmad, 1972). Between Ratnagiri and north of Bombay there are a series of hot springs along a straight line parallel to the coast, which shows that they are located on a line of fracture. These facts on the land support the view that the coast is faulted (Krishnan, 1956).

The north and northeastward drift of Indian plate is engineered by the accretion of plate margin at the Mid-Indian Ocean Ridge where the current growth rate at various sites have been estimated to be 1 to 3 cm/yr (Le Pichon, 1968). During the Early Tertiary period, possibly, there was an imbalance in the accretion - shortening process, whereby the rate of plate growth exceeded the rate of crustal shortening. The "regional upwarp" (Radhakrishna, 1967) or "monoclinal flexure" (Subramanyan and Kirankumar, 1985) or doming (Subrahmanya, 1985) in the west coast, can be attributed to the imbalance in the accretion - shortening process during the northward translation of the Indian plate. Increasing accumulation of tectonic stress in the zone of upwarp must have led to the development of a major fault culminating in Western Ghats as upthrown block and west coast as the down thrown block (Ravindra and Krishna Rao, 1987). Studies carried out using geophysical data also showed the presence of faulting along the west coast (Chandrashekharam, 1985).

The Hemavathi-Thirthahalli mega-lineament might have developed during the west coast faulting. This lineament with considerable length and almost parallel orientation to the west coast indicates that it might have formed as a sympathetic fracture/fault (?) to the West Coast Fault. There are some other sympathetic major lineaments such as Yagachi and Tunga. A good number of other lineaments are also parallel to the coast. This parallel trend clearly indicates that these lineaments have developed due to a single tectonic activity. Majority of these lineaments may be fractures, but some may be faults. A number of streams/rivers follow these fractures. Hemavathi River, Sole Abbur Halla (tributary of Cauvery), Sanivar Sante Hole, Bilaha Halla (tributaries of Hemavathi) and tributaries of Tunga run along this mega-lineament. The mega-lineament extends from slightly above Tirthahalli to below Piriyapatna. Tirthahalli falls within this lineament and Hemavathi is the major river running along this lineament and so it has been called as Hemavathi-Tirthahalli mega-lineament. The mega-lineament is prominently seen all along on the image except in some parts.

Between 1967 & 1977 about 100 earthquakes (86 in Koyna area alone) of intensity '4' or above on Mercalli scale have been recorded in the Southern Peninsula. Most of these have been found to occur in the vicinity of NNW-SSE to NW-SE trending mega-lineaments. This indicates that a moderate to high seismic activity has been widely spread in the area. The proximity of this activity to mega-lineaments again indicates that sometime or the other they have been reactivated (NRSA, 1981). So the Hemavathi-Tirthahalli megalineament should also be examined to know its seismic status.

When correlated with Geological and Mineral Map of the area (1:500,000 scale, GSI, 1981) some mineral locations (eg. Magnesite) are seen to be along this lineament. The potential of this mega-lineament in mineral and ground water resource is to be explored. Whether the lineament is a fault or fracture has to be confirmed by field check. Other mega and major lineaments should also be studied in detail to understand the tectonics and structure of the area.

CONCLUSIONS

- The synoptic view provided by the satellite image helped in delineation of mega-lineament and other lineaments.
- The Hemavathi-Tirthahalli mega-lineament might have developed during the West Coast Fault period. The coast parallel trend and the length are the factors favouring the above inference.
- The occurrence of a number of Mega, Major and Minor lineaments indicate that the area is tectonically disturbed.
- As most of the earthquakes (between 1967 & 77) in Southern Peninsula occurred in the vicinity of NNW-SSE to NW-SE trending mega-lineaments, the Hemavathi-Tirthahalli mega-lineament should be studied to understand its seismic status.
- The potential of the mega-lineament for ground water, mineral and engineering geological applications can be fruitfully studied.

ACKNOWLEDGEMENTS

Author is grateful to Mr. G Behera, Manager, US&U, NNRMS, ISRO for his encouragement in the study and Mr D.N Setti, Head, RRSSC, Bangalore & Prof A.K Roy, Head, Geosciences Division, IIRS, Dehradun for their comments and suggestions. Thanks are due to Ms. Sumya for her assistance in the work.

References

1. Ahmad, E., "Coastal Geomorphology of India", Orient Longman, 1972.
2. Chandrashekaram, D., "Structure and evolution of the western continental margin of India deduced from gravity, seismic, geomagnetic and geochrono-

logical studies", Phys. Earth Planet Inter, Vol.41. pp. 186-198, 1985.

3. Geological & Mineral Map of Karnataka and Goa (Scale 1:500,000), Geological Survey of India, 1981.

4. Krishnan, M.S., Geology of India and Burma 6th edition, CBS Publishers and Distributors, 1956.

5. Le Pichon, X., "Sea floor spreading and continental drift", Jour. Geophy. Res.,Vol. 73. pp.3661-3697, 1968.

6. Radhakrishna, B.P., "The Western Ghats of Indian Peninsula" In Proc. Seminar on geomorphological studies in India, Univ. of Saugar, Sagar, 1967.

7. Ravindra, B.M., and Krishna Rao, B., "Relation of coastal faults and river morphology to sea erosion in Dakshina Kannada, Karnataka", Jour. Geol. Soc. India, Vol. 29, pp.424-432, 1987.

8. Regional Geological Map Part of West Karnataka, NRSA TR 0141-0150, 1981.

9. Srinivasan, R., and Sreenivas, B.L., "Some new geological features from the Landsat imagery of Karnataka", Jour. Geol. Soc. India, Vol. 18. pp.589-597, 1977.

10. Subrahmanya, K.R., "Morphology of the Mangalore environ", Abst. seminar on Coastal Geomorphology, Mangalore Univ. 1985.

11. Subrahmanya, K.R., "Evolution of the Western Ghats, India a simple model", Jour. Geol. Soc. India, Vol. 29, pp.446-449, 1987.

12. Subrahmanyan, V., and Kiran Kumar, B., "The evolution of Kerala coast", Abst. Seminar on Coastal Geomorphology, Mangalore Univ., Mangalore, 1985.

13. Wadia, D.N., Geology of India (Fourth Edition) Tata McGraw Hill Publishing Co., New Delhi, 1975.

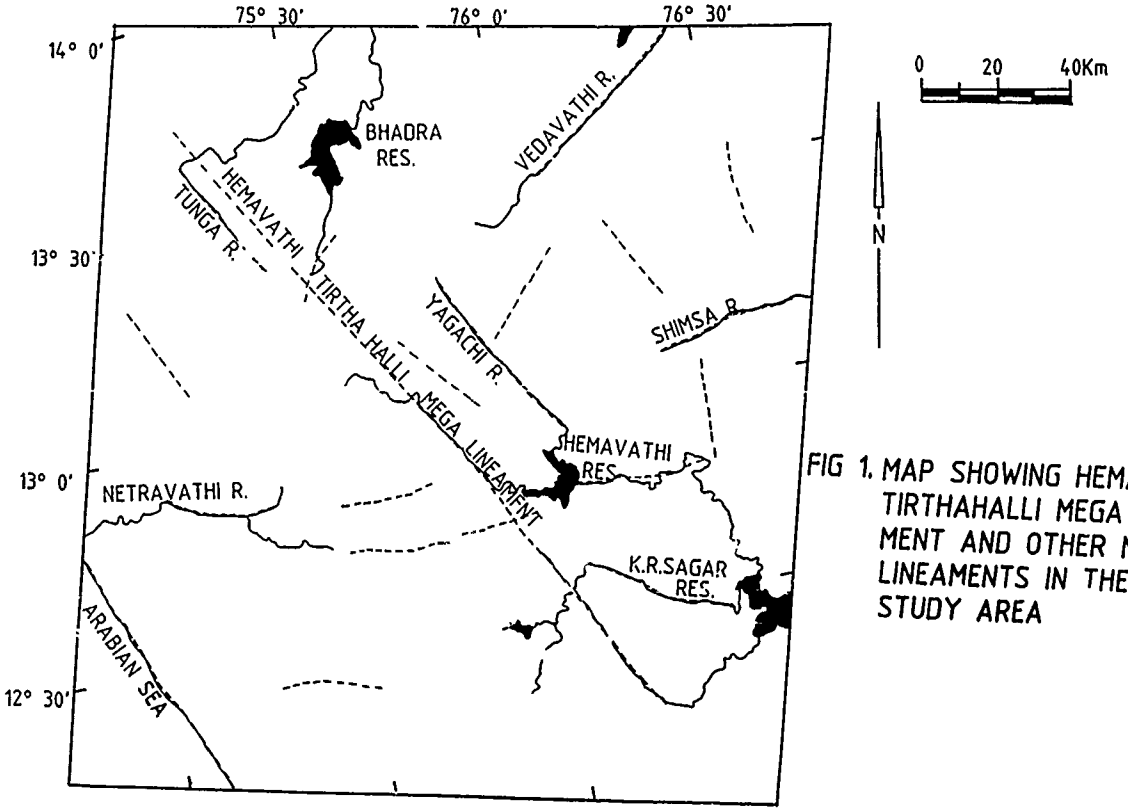


FIG 1. MAP SHOWING HEMAVATHI-TIRTHAHALLI MEGA LINEAMENT AND OTHER MAJOR LINEAMENTS IN THE STUDY AREA

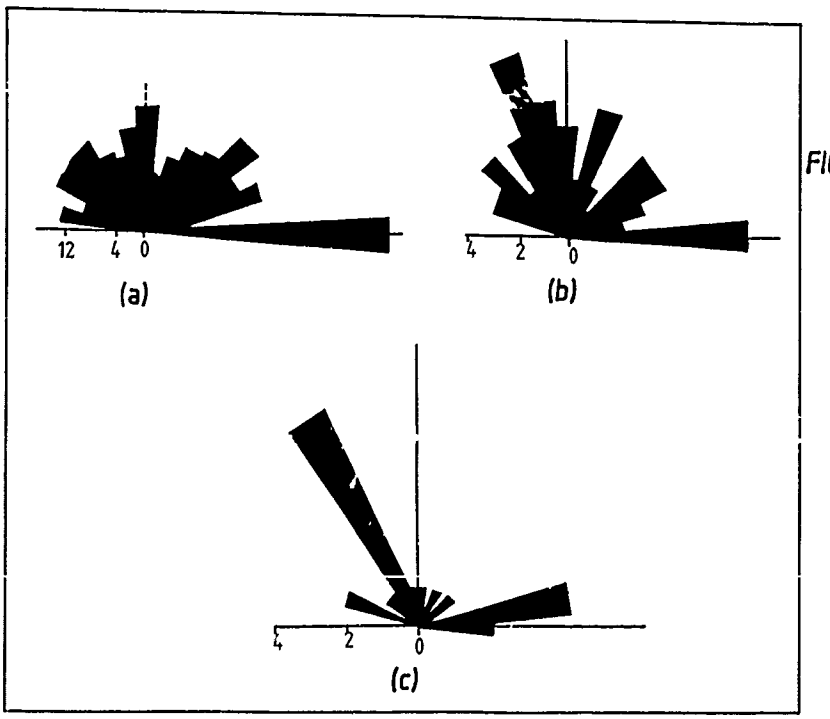


FIG 2. ROSE DIAGRAMS (SHOWING DIRECTION) OF LINEAMENTS OF

(a) LENGTH <10 Km
(b) LENGTH 10-20 Km
(c) LENGTH >20 Km

APPLICATION OF GIS AND REMOTE SENSING FOR LINEAMENT MAPPING OF STRUCTURAL FEATURES IN THE ROUYN-BEAUCHASTEL PROPERTY, QUEBEC

S. Gravelle and M. S. Akhavi

Remote Sensing Program
Nova Scotia College of Geographic Sciences

ABSTRACT

The Rouyn-Beauchastel property lies in the southern part of the Abitibi volcanic-sedimentary belt within the Superior Structural Province of northern Quebec.

Various geological and geochemical maps were digitized using an ARC/INFO GIS system and were transferred and registered to both Landsat Thematic Mapper (TM) and SPOT images. The resultant geological map was used as a reference theme file to detect the usefulness of enhancements performed on both raw Landsat TM and SPOT data for delineation of structural features.

Enhancements were performed to highlight image lineaments. The detected linears were transferred to the ARC/INFO environment and were integrated with the existing geologic theme files to produce an updated interpretive geological map.

More than thirty mjor lineaments were mapped. Most of these linears coincided with known faults or lithological contacts; however, six lineaments had not been detected previously. Specifically, a number of newly depicted northwest and northeast trending lineaments and their intersections, deserve further study because mineralization in this area often occurs in this type of structure.

KEY WORDS: Rouyn-Beauchastel property, lineament detection, GIS/RS database, TM and SPOT imagery

INTRODUCTION

The Rouyn-Beauchastel property is located in the Abitibi region of northern Quebec (figure 1) and is centered in the Rouyn-Noranda mining camp. The property lies in the southern part of the Abitibi volcanic-sedimentary belt within the Superior structural province. Numerous granitic and gabbrroic intrusions have been observed in this Archean assemblage which are occasionally unconformally overlain by the Huronian sedimentary sequence. The Archean basement has been cut by numerous faults and fractures developed during various episodes of deformation (Duplessis, 1987).

This study was conducted in support of geologic investigations by Groupe Conseil DOZ Inc. in the Rouyn-Beauchastel property (Gravelle, 1988). The main objective of this study was to detect and map lineaments and other structural features by integrating information residing in a geographic information system (GIS) with the geologic data derived from digital image processing of Landsat TM and SPOT satellite imagery.

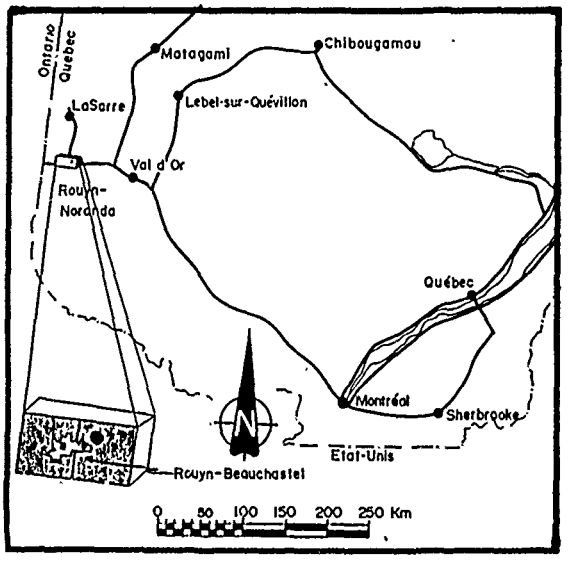


Figure 1 - Location of Study Area (after Duplessis, 1987)

PROCEDURE

Various geological and geochemical maps were digitized using an ARC/INFO GIS system resulting in a composite geological map. This composite map was then transferred and registered to both Landsat TM and SPOT geometrically corrected images. The geological map was then used as a reference theme file that was displayed over numerous enhanced images.

Landsat TM data provided the spectral sensitivity while SPOT data added detailed resolution. The Landsat TM data with 30m resolution was then registered to the SPOT data with 10m resolution. Figure 2 is an enhanced example of TM data covering the entire study area while figure 3 is a SPOT image of the southwest portion of the study area indicating resolution details of the two data types. An example of SPOT to TM registration is shown in

figure 4, which coincides with the location of figure 3. This image is slightly blurred as a result of the registration process, however it contained greater spatial and spectral details compared to each satellite image alone.

Enhancements such as look-up tables, principal components, Martin-Taylor and filtering were performed using a DIPIX/ARIES image analysis system to highlight image lineaments. Directional filtering proved to be one of the better enhancement techniques used to identify lineaments. Figure 5 shows an example of a NW directional filter of TM band 7.

The lineaments identified during the various enhancement techniques were outlined and classified as theme files in the DIPIX/ARIES environment and transferred to the ARC/INFO GIS environment. The linears were integrated with the existing geologic theme file and an updated interpretive geologic map was produced for the Rouyn-Beauchastel Property (Figure 6).

INTERPRETATION

Over 30 major lineaments or groups of lineaments were identified and mapped. Gravelle (1988) identified each lineament by a number on the resulting map (figure 6) and examined its relationship with known geological information. Due to the limitation of space, each lineament cannot be discussed in detail in this paper. However, an attempt is made to provide generalized information in respect to geologic characteristics and significances of these lineaments based on the interpretation of GIS/RS database aided with the existing geologic information which was reviewed as an integrated map layer.

The following is a brief description of these lineaments:

Nineteen lineaments detected could be associated with known geological structures (faults or shear zones) or related lithological contacts.

Seven other lineaments appeared to indicate the continuation of mapped faults or shear zones.

Two lineaments were found to correspond with geochemical and geophysical data examined in this investigation. However, these lineaments were not mapped on the existing geologic map (Duplessis, 1987).

Six lineaments detected in this investigation were not previously mapped. These are anticipated to be of structural origin or stratigraphic nature and deserve ground truthing and further geological investigation.

CONCLUSIONS

A geographic information database, supported by digital image processing of TM and SPOT imagery, provided an ability to transfer conventional map data to an image analysis system and extract pertinent geologic information.

This included the detection of structural, stratigraphical, and geochemical/geophysical relationships which were previously known. However, it was concluded that a number of structural units previously mapped in this area could be extended; as well, a number of newly identified lineaments could be detected. These included newly depicted northwest and northeast trending lineaments and their intersection with other lineaments, which are significant structures for emplacement of minerals in this property.

Gravelle's (1988) work was entrusted to Groupe Conseil DOZ Inc. to verify the exact geologic nature of newly identified lineaments and determine target localities suitable for conducting traditional geological investigations for mineral exploration.

ACKNOWLEDGEMENTS

The authors are thankful to Mr. Alaine Moreau of Groupe Conseil DOZ Inc. for providing TM and SPOT imagery as well as conventional data, and assisting in the interpretation of the geology of the study area. Mr. John Wightman, Principal of the Nova Scotia College of Geographic Sciences provided logistic support and encouragement. This paper benefitted from a review by our colleagues Mr. W. R. Trenholm and Mr. E. Wedler.

REFERENCES

1. Duplessis, D., "Rouyn-Beauchastel Evaluation Report" Groupe Conseil DOZ Inc., Rouyn-Noranda, Quebec, 1987
2. Gravelle, S., "Application of Remotely Sensed Techniques for Lineament Mapping of Structural Features in the Rouyn-Beauchastel Property", Co-op Project Report, Nova Scotia College of Geographic Sciences, 1988.



Figure 2 - Difference Between TM bands 7 and 1



Figure 3 - Panchromatic SPOT Image

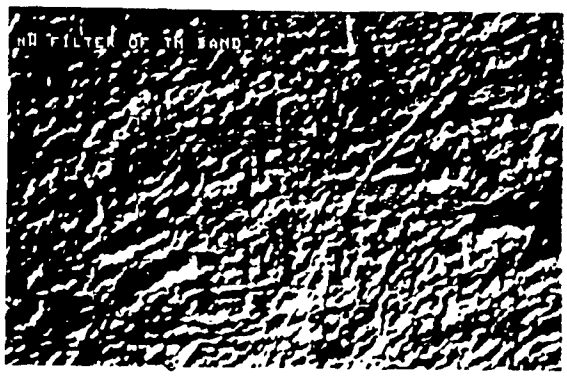


Figure 5 - Northwest Filter of TM Band 7

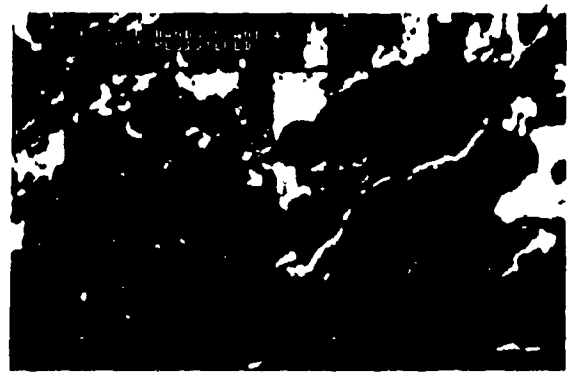


Figure 4 - Product of Bands 4 and 5
TM to SPOT Image Registration

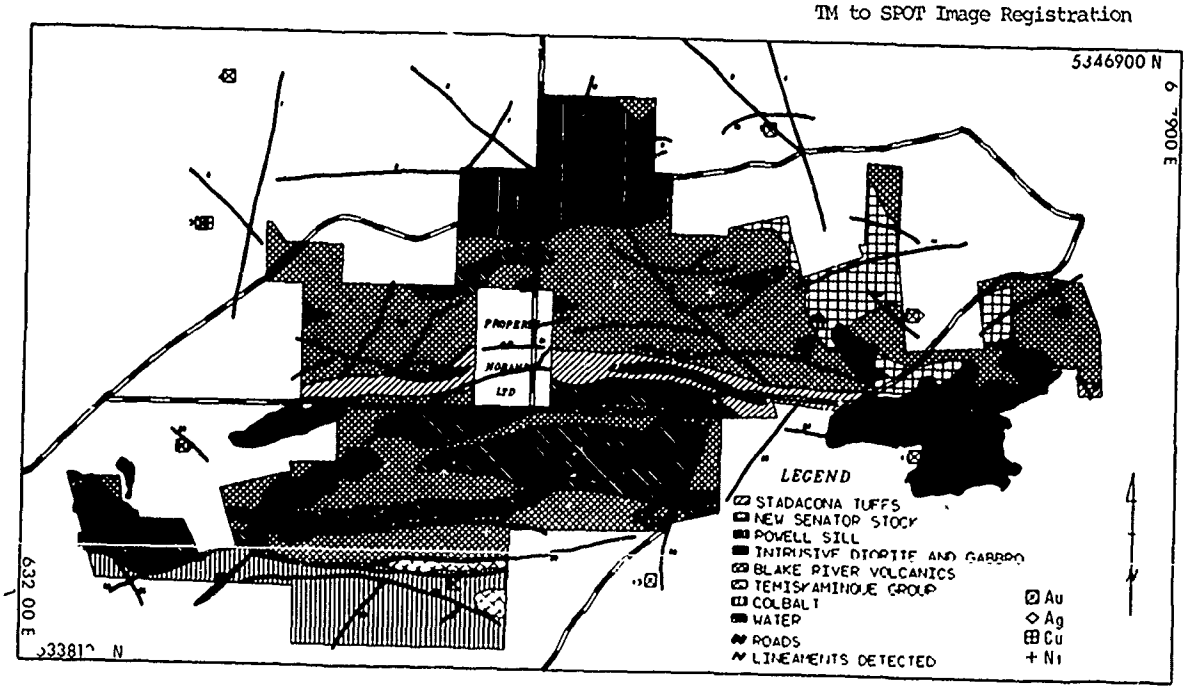


Figure 6 - Lineament Map of the Kouyn-Beauchastel Property

APPLICATION OF REMOTE SENSING IN THE EVALUATION OF SELECTED EARTHQUAKE & PRONE AREAS IN VIETNAM

DOÃN NGỌC LƯU
INSTITUTE OF GEOPHYSICS
CNRS OF VIETNAM
NGHIA DO - HA NOI V.N

Tel 52360 HANOI
Telex 4525 HANOI ACAVINAVT

ABSTRACT

The satellite imagery direct observation can display lineaments which are parallel or sometimes juxtaposed with faults infered from géophysical data (aéromagnétism , gravimétry , séismic) and détailed géologic map survey

This paper présent a study of the évaluation of selected earthquake - prone areas in Vietnam

The data from the images of the satellites : LANDSAT , SOYUZ , SPOT ... and from th aerial photo was analysed to find out the éléments of tectonic such as : linéament system , faulted system , folded system , the géological bounderies ...

Using the models mathematique and computer , to define the major élément-of-Tectonic : the major systems , and their directions

Using the results of géophysical data processing , the major charactéristics of géophysical fields of the tectonic units of Vietnam (the magnétic field , the gravity anomaly fields , the distribution of the séismic focus and épicenters , the deep géological structures)

Basing on Remote sensing data analysis , géophysical data processing and data synthèsis , to find out : the e évaluation of selected Earthquake - prone Aréas in Vietnam

DIGITAL MOSAICING TECHNOLOGY FOR SYNTHETIC APERTURE RADAR IMAGE PRODUCTS

Fern Schultz*, J. Bryan Mercer, Peter Button

INTERA Technologies, Ltd
Calgary, CANADA
**currently at The University of Calgary*
Department of Surveying Engineering

ABSTRACT

Since 1984, INTERA's STAR synthetic aperture radar (SAR) systems have provided high resolution radar imagery in digital form. INTERA has been motivated to develop the capability to perform all steps inecessary in the production of image mosaics in the digital domain. These include radiometric corrections, geo-referencing of the imagery by registering to a map projection, and the seaming of image strips together. These operations are collectively referred to as digital mosaicing. In a recent development, data are logged from a Global Positioning System (GPS) receiver on-board the STAR aircraft and referenced to ground-based GPS data, making it possible to compute "virtual" ground control points which substitute for those normally obtained from existing maps.

This paper describes the elements of INTERA's digital mosaicing technology, including the corrections applied, seam selection and geo-referencing techniques, and data management.

Keywords: SAR, Digital Mosaicing, GPS, Geo-referencing.

1. INTRODUCTION

Airborne radar images have been used for thematic mapping purposes for two decades. Typically, the imagery has been presented in the form of film products or "analogue mosaics" in which the film strips are fitted together, cut and warped, in a manner similar to air photo mosaic production. High resolution digital synthetic aperture radar (SAR) image data acquired by the INTERA STAR system were originally used in SAR mosiac production by initially transforming the data into film products and then treating these in analogue fashion. However, there are significant improvements that can be made by doing most of the processing in digital format before returning to a final hard copy output. INTERA's capability to produce digital SAR mosaics includes steps to perform radiometric corrections, geo-referencing and the seaming together of image strips all on a digital image processing system.

The rationale for digital mosaicing includes:

1) Improved preservation of radar image resolution and dynamic range, compared to the multi-stage analogue process, resulting in a sharper looking final image product which retains more of the original information.

2) Controlled radiometric corrections can be applied, providing a smooth visual transition from one image strip to the next in the mosaic by balancing across a strip and by matching from strip to strip.

3) Seams that can be made virtually invisible.

4) Geometric warping can be mathematically defined and the accuracy determined by residual error analysis.

5) This technology can be easily adapted for the use of imagery which has been fully rectified (i.e. with all terrain displacement effects having been corrected).

6) The technology has excellent compatibility with development leading toward automated processing.

2. STAR DIGITAL SAR DATA

STAR SAR data are acquired on-board the STAR aircraft in digital format and are recorded on a high density digital tape (HDDT) at 4096 8-bit bytes per scan line. The scan line is either a 46 km swath (with approximately 4.2 m by 11.6 m pixels) or a 23 km swath (with approximately 4.2 m by 5.8 m pixels). The data are preprocessed in flight to translate slant range to ground range. Flight line spacing typically allows for 60% overlap of the flight lines. The STAR system is described in detail in Nichols et al. (1986).

Post-flight data processing of the 46 km swath width data averages three scan lines to 12.6 m by 11.6 m pixels and provides transcibed data from the HDDT to CCT's in preparation for input to the digital image processing system.

3. DIGITAL MOSAIC PRODUCTION

3.1 Area of Coverage and Data Volume

The area to be mosaiced typically covers an area the equivalent of a 1:250 000 scale map quadrangle which in the mid latitudes covers about 110 km east-west by 80 km north south (1° latitude by 2° longitude). If flight lines are flown in an east-west orientation (with either all north-looking or all south looking), approximately six lines will cover the required area. For a north south orientation approximately eleven flight lines will cover the area. With the approximately 12 m^2 pixels each flight line will be almost 9800 records in length, resulting in approximately 440 Mbytes data input and 160 Mbytes data output (with 60% overlap of the flight lines). Therefore, a major problem encountered in the implementation of digital mosaicing is the management of this tremendous data volume. Since these quads must be produced at a commercially competitive rate, significant demands are placed on hardware and software processing speed, system storage and data transfer capacity and the operational interface.

3.2 Corrections Applied

Initial processing on the digital mosaicing system (Dipix ARIES III) consists of replacing any dropped lines by averaging or duplicating adjacent scan lines of data. This cosmetic correction retains the geometric fidelity of the data. Next, radiometric corrections are applied to each flight line to account for the across-track intensity fall-off which is inherent in the radar image. At the same time the mean is shifted to allow for a better distribution of the data across the 8-bit grey scale, optimizing the dynamic range of the digital data. A control image strip is selected which best represents the terrain in the area being mosaiced. Statistics are gathered for a sample rectangle in the control strip and the new mean and standard deviation for the distribution are selected based on the type of terrain which has been imaged and SAR data processing standards recommended by the Canada Centre for Remote Sensing (CCRS) (Wessels, 1987). For STAR data acquired with a gain setting of one, the appropriate mean ranges from 55 to 65 on the 0-255 grey scale, with a standard deviation of approximately 24, as derived from the following relation:

$$\text{standard deviation} = \frac{\text{mean}}{\sqrt{\#\text{ of looks}}}$$

where the STAR system has 7 looks. The correction based on the sample rectangle is applied to the entire image and statistics are output for matching rectangles in the area of overlap for adjacent strips. Radiometric corrections are applied and statistics output sequentially for all strips in a mosaic except in cases where a specified mean and standard deviation are applied to all strips. This can be justified if the data set is relatively consistent across the strips (ie. there is little variation in the type of terrain and cover imaged) as it reduces the computation time requirements considerably.

3.3 Ground Control and Resampling

In the situation where good map coverage exists (for example, in most of North America), each strip can be registered to the ground by selecting ground control points identifiable on both the image and the corresponding map. These are often man-made features such as bridges or road intersections which return a strong signal on the radar image and are easy to pin-point on the maps. The concept of spatial warping (Pratt, 1976) is applied through a second order polynomial transform. This is calculated and applied to the images in a cubic convolution resampling algorithm. The result is that the image strips are all in a common, absolute coordinate system, and are ready for seam selection and mosaicing which is described in detail below.

In areas without good existing map coverage or where registration of individual strips to maps cannot provide the level of positional accuracy required, an alternative approach is taken. Individual image strips are registered one to the next by selecting a master image, establishing an origin which can accommodate all the data with positive coordinates, and identifying common, or homologue points in adjacent images. The distribution of points is best when the area of overlap is spanned from corner to corner, in a pattern resembling a braced quadrilateral. The adjacent images are then resampled to fit the master image (which is not resampled at this stage), using first-order polynomial transformations derived from the homologue points. A resampled image then becomes the master image as the next adjacent image is registered to fit it (see Figure 1). Each image is resampled into the common, relative coordinate system established by the first master image strip.

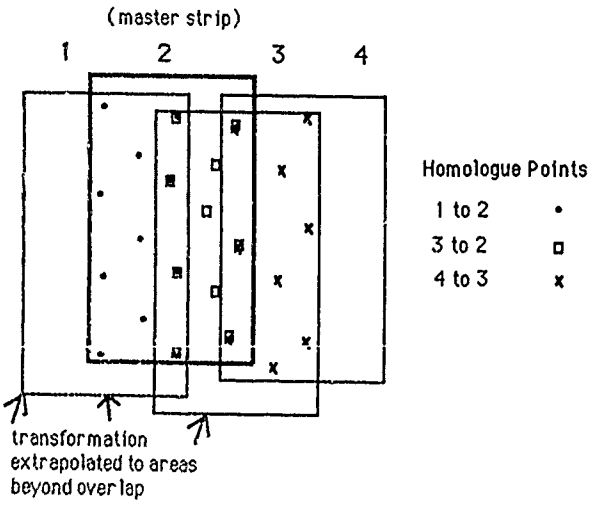


Figure 1. Image-to-image registration

The selection of a first-order polynomial transformation is based on experimentation with first and higher order transformations. The area of overlap between any two adjacent images is limited to 60 %, and the image strips are long and narrow, but of approximately even length. This results in good control over scaling and rotation in the along-track (azimuth) direction, but a narrow band of control across-track (in the range direction). The transformation must be extrapolated to the area beyond the narrow area of control in the overlap between the two image strips. Inconsistencies at the data acquisition stage complicate the image-to-image registration by introducing along-track variations in the strip geometry related to aircraft motion compensation, crabbing, drift and other factors. These variations in the data would be better handled by higher order polynomial transformations, but the higher order terms introduce a greater degree of warping in the area beyond the overlap between two images. The first order polynomial transformation minimizes this warping and therefore contributes to control across the entire mosaic area as the warped image is used as the master image for registering the adjacent strips in the sequential build-up of resampled image strips. The net result is a compromise in the registration of any two image strips, but a better fit of the final mosaic to ground control.

All the image strips are in a common coordinate system at this stage and can be overlaid to allow for placement of the seams in the areas of overlap. The seams are located primarily within the mid-range data, thereby minimizing shadow and near-range/far-range signal return differences due to the different angles of incidence. The path the seams follow ("meandering seams" - Scott, 1986) are intended to minimize detection of the seams in the final mosaic and are defined using an algorithm which allows for the greatest flexibility in dealing with registration offset. Seam files are generated which take data from image A to the left of the seam, and data from image B to the right of the seam line (see Figure 2).

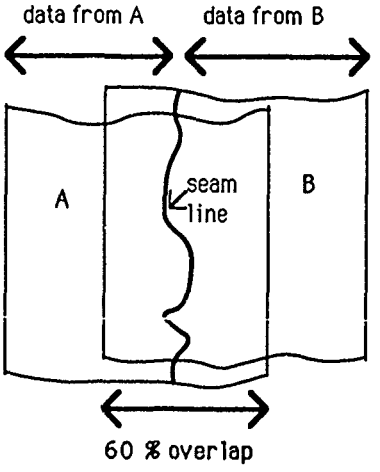


Figure 2. Seam Selection

The various components of a final mosaic are tiled together using a first-in-wins algorithm, with seam files laid down prior to the resampled image strip files. The complete mosaic is then ready for image-to-ground registration, using existing map information, and incorporating supplementary survey information (eg. from Doppler survey points with locations identifiable on the radar imagery). The final resampling is performed with pixels output to 12 metres square, using a cubic convolution routine with the second-order polynomial derived from the ground control points and according to a specified map projection. The digital radar image mosaic is ouput on CCT for lasar plotting and final print production at a selected scale, much like a semi-controlled photo mosaic. No correction has been applied to account for relief displacement effects which result in ground targets being displaced by an amount proportional to the target height.

4. VIRTUAL GROUND CONTROL WITH GPS

In a recent development a more refined mosaic product has been introduced which incorporates positioning information made available through the Global Positioning System (GPS). Conventional STAR data acquisition is done with Inertial Navigation System INS navigation on-board the aircraft. This system is inherently subject to drift problems that increase with mission duration. INS errors are typically described by a sinusoid of increasing amplitude and an 84 minute Schuler period. The ability to register adjacent image strips acquired in this fashion is therefore influenced by the performance of the INS.

With the introduction of GPS data, positioning information for the radar image strips is greatly improved. The data are logged from a GPS receiver located on-board the STAR aircraft and normally are referenced to correlated ground-based GPS data (differential mode). With precise aircraft navigation data and a model of the radar process it is possible to compute "virtual" ground control points which substitute for those normally obtained from existing maps. This is particularly relevant where existing maps are unavailable or unreliable.

The GPS data are acquired on board the aircraft simultaneously with the SAR image acquisition. Annotation data on the SAR image is correlated to the GPS data providing the geographic position for a specified line and pixels on the SAR image. Computations take care of all errors except the displacement of a feature due to its height above the terrain. Corrections that must be made to the position data are:

(a) Accurate computation of the cross track distance to the required pixel. The GPS gives an accurate height above sea level, whereas previously altitudes have been the pressure altitudes used in aircraft

navigation. In the tropics it is possible to have up 500 metres difference between the "barometric" and true height above mean sea level. The true height has a considerable impact on the true cross track distance to each pixel.

(b) Along track error due to the long term drift angle.Short term deviations from the planned flight are compensated for in the STAR motion compensation system. Deviations with time scales of several minutes cannot be removed by the internal motion compensation system. The net result is that imagery is displaced in the along track direction. It can be correctly positioned using "long term drift angle" data.

(c) Range dependent delay. This is a delay present in the SAR processing and is dependent on the range of the target. It causes a displacement of the image in the along track direction. Since the processing delays are known, the image may be restored to its correct location.

With this information, the positions corresponding to several pixels on a specific scan line are determined at regular intervals along the flight line, and input to a ground control file in the digital mosaicing system. These locations are referred to as "virtual ground control points" (GCPs). The image is resampled as before into an absolute coordinate system and specified map projection, using the low-order polynomial transform derived from the "virtual" GCPs. The positional accuracy of the warped image is greatly improved as a result of the number and excellent distribution of the "virtual" ground control points. Misregistration between two adjacent images resampled in this manner is minimized and seam selection is improved accordingly.

This method of using GPS as the prime navigation system on standard SAR surveys is not always economical at this time because of the restricted window of operation due to limited access to the GPS satellite constellation. As GPS coverage improves with the launch of more satellites, it is expected that it will be possible to use GPS as the prime navigation system on standard SAR surveys. The result will be better registration of adjacent image strips as warping of the flight lines will be reduced.

5. CONCLUSIONS

Digital mosiacing techniques for synthetic aperture radar imagery greatly improve the quality of the output products by preserving the radar image resolution and dynamic range, controlling radiometric imbalance by matching grey levels from strip to strip and by making seams that are virtually invisible. Geometric warping is mathematically defined and controlled through input of ground control which can be adapted to conditions and availability of relevant information. Output products can be produced at various scales, taking full advantage of the high resolution of the original digital SAR data.This digital mosaicing technology is used in a production environment and continues to be refined to improve the efficiency and accuracy attainable. The use of GPS to replace the need for GCP input further broadens utility and accuracy of the product. Future developments in automated image processing will serve to complement these techniques and will further improve the production of digital mosaics from SAR image data.

6. REFERENCES

1. Nichols, A.D., J.W. Wilhelm, T.W. Garfield, D.R. Inkster and S.K. Leung. "A SAR for Real-time Ice Reconnaissance". IEEE Trans. on Geoscience and Remote Sensing, Vol. GE-24(3). pp.383-389, 1986.

2. Wessels, G.J. Personal communication, 1987.

3. Pratt, William K. "Digital Image Processing". John Wiley and Sons, New York. pp. 430, 1978.

4. Scott, Andrew. Personal communication, 1986.

SATELLITE IMAGE PROCESSING WORKSTATIONS FOR THE EOS ERA

W. Markwitz, G. Schreier, H. Rister

DLR, WT-DA, German Aerospace Research Establishment
D-8031 Oberpfaffenhofen, FRG

1. Abstract

The German Aerospace Research Establishment DLR in Oberpfaffenhofen is currently establishing a User Support and Operation Center (USOC) for data processing support of the upcoming space application projects, mainly in the field of remote sensing. First clients of this center will be the ERS-1 and X-SAR-Projects, where DLR will act as a Processing and Archiving Facility (PAF).
In the framework of these projects, modern technology will be introduced to build up a new flexible data processing infrastructure. The basic elements are standardized UNIX workstations interconnected via Ethernet-TCP/IP, forming a powerful computing network.
These workstations are scaled to serve a large variety of different applications, such as control of data reception, data base applications, orbit determination, SAR processing and geocoding as well as interactive data analysis. Especially for image processing the configurations selected range from low-cost diskless machines up to superminis with number crunching capabilities, 32 bit depth real colour display and animation features.
To reach the needed performance, the workstations are supplemented with embedded special purpose processors which provide the hardware environment for a new image processing software. In preparation of the expected EOS-data streams and in order to meet the requirements for real time processing and for the generation of new products, new hardware add-on-modules are integrated and a modern powerful software package is under development.
Requirements, basic concept and current activities at DLR will be presented.
Keywords:
Operational Processing Systems, Image Processing Hardware, Image Processing Software, Standards, EOS-Systems, ERS-1 PAF.

2. Introduction

In the past much effort was spent to define payloads and missions of a future operational Earth Observation System (EOS) based on polar platforms. These platforms are scheduled by the american, japanese and european space agencies for the mid to end of the 90's.
Obviously, the sensor segment represents only one element of an operational end-to-end system. An other is the data end user, whose requirements as to amount, quality and flexibility of data are growing rapidly.
The interface between space segment and data user is covered by the payload ground segment, consisting of data receiving, data processing and archiving/ distribution components. This paper will especially address the authors views of the needs and manageable solutions for this essential element of the data chain. Basically, the physical / technical possibilities and restrictions must be made compatible to a wide range of different user requirements.
As the most predominant data set, regarding both amount and processing implications, remote sensing image data and their future treatment by image processing workstations is refered to in this article. Hardware and software requirements and forseen realization concepts will be shown.
Experience in the integration of the presented scenario is currently gained by DLRs Applied Data Systems Division, responsible for the German part of the Processing and Archiving Facilities (PAF) for ERS-1 and SIR-C/X-SAR. A quite elaborate system for data processing will be provided by this basic infrastructure, in order to handle SAR data for different sensors in terms of processing, validating, geocoding and image interpretation as well as orbit data computation and special sensor data generation (PRARE in case of ERS-1).
It will be shown, how the requirements of EOS data handling can be covered by the current PAF implementation.

3. Data Processing Requirements for EOS

To discuss the items in a more general context, the term EOS shall not be restricted to the polar platforms of NASA, ESA and NASDA, but is applied with respect to the general scenario of remote sensing satellites during the last decade of this century.
Although technological developments are moving to on-board real-time processing for SAR sensors, precision products need a more flexible approach in image generating, based in the ground segment. Therefore, raw SAR data are expected to be one of the major contributors to the data volume. While optical data only need a geometric and radiometric correction for image visualization, SAR holographic raw data must be treated in a very complex computational manner prior to the further evaluation for remote sensing.
In addition to the enormous amount of raw data, the various stages of data enhancement put a further load to the ground processing systems, both in computing power and archive/ access capabilities.
Furthermore, some EOS sensors and related applications will require real-time or near real-time processing to ensure on-site monitoring of environmental changes. In order to monitor environmental changes, an enormous quantity of data sets has to be archived. Finally, during the EOS era, there is an increased need for handling the data collected by different sensors not as individual entities. On the contrary, satellite remote sensing has to be understood as integral view of our earth's surface, under various physical conditions.
The issue therefore is the corregistration of different sensor data among each other or with geo-based raster or vector data. An important step towards corregistration is georeferencing and geocoding, aspired as standard data format for EOS for all kinds of sensor products. The huge amount of data sets from the sensors on one side and the challenging user data products on the other side will lead to requirements concerning hardware and software for an EOS image processing ground system:

a. Utilization of standard hardware, peripherals and software as far as possible to make use of the revolutionary technogical development in order to cope with the enormous throughput issue and to reduce the costs for update and maintenance.

b. Utilization of a standard operating system and standard programming languages to guarantee portability of code to advanced hardware solutions This also includes a careful documentation and development of SW according to common SW engineering standards

c. Incorporation of special processing hardware to meet the throughput requirements for (new) real-time processes or special data interfaces.

d. Long term storage media with short access times or on-line random access possibilities for archiving data and products.

e Communication between different processes and systems to ensure exchange of data and control information, including the possibilities of telescience

f. Flexible, versatile operator interfaces to control and maintain complex processes. For image processing, this means visualisation tools for many kinds of different digital image data.

For most applications of the EOS-era, the requirements stated above can be met by a modular, ot-the-shelf available system of distributed, dedicated workstations. Their specifications will be presented in the next chapters.

4. EOS Workstation Hardware Concepts

The term 'workstation' defines a wide range of computing equipment, from PCs to super-mini computers Their basic concept is that computing and/or interaction intensive jobs are dedicated to one appropriately scaled machine which therefore is optimized in hardware and peripherals

While simple text processing and I/O can be managed by a PC-like workstation, SAR processing for instance must be installed on a super-mini supported by special hardware. The analysis of this workstation-concept shows, that even number crunching applications can have a better price/performance ratio on workstations than on super-mainframe computers (figure 1)

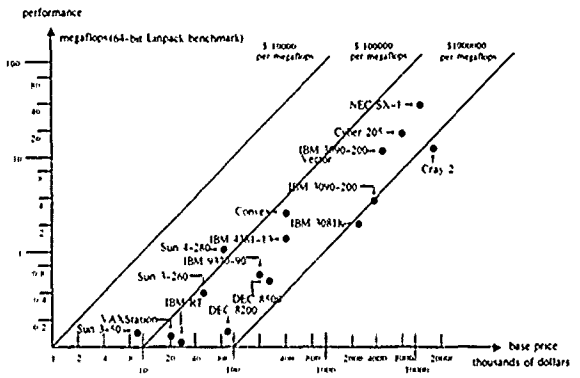


Fig 1.: Price/Performance ratio for computers (Modified from Dongarra, 1987)

The example mentioned above leads to the conclusion, that workstations could be the flexible tool to cope with EOS image processing. Workstations can cover a wide range of performance (measured in 'Millions of Instructions per second' or MIPS and 'Millions of Floating Point Operations per Second or MFLOPS), having the potential to upgrade main memory and disk capacity, to add special hardware via standard bus-systems or interfaces and finally interconnection to other workstations to build a distributed processing system with better performance than a large scale computer. Especially image processing workstations require the possibilities for raster and vector data visualization and a powerful man/machine interface.

To avoid that these requirements lead to incompatible, inhomogenous hardware solutions, some basic standards for the main components have to be considered and accordingly adapted to future concepts. These hardware standards are depicted in figure 2 and explained as follows:

a. Central Processing Unit

The avability of modern processing languages like 'C' have fortunately promoted the use of higher programming instructions rather than Assembler coding. Fortunatelly, this develc,ment is supported by the de facto standard UNIX operating system for nearly all used types of work stations Therefore, the application software should easily be portable to different type CPUs, assuming well documented source code and the appropriate compiler are available. Practically, distributed processing systems should concentrate just on a few different CPUs to ease maintainability As the development of CPUs is going towards 'Reduced Instruction Set Computers (RISC)' and other new technologies, the performance with now at least 10 MIPS and more than 1 MFLOP will increase accordingly.

b. Bus System and Special Processors

A standard bus system is one of the most stringent requirements for a workstation Although manufactures tend to integrate the whole set of basic computer functions on a main board (CPU, memory, terminal, LAN, etc), special peripheral high data rate and co-processing devices shall be connectable via a bus. One of the busses best supported presently is the VME-bus ('Virtual Memory Extended'), with a large variety of devices like array processors, image display peripherals, data formatters and I/O device controllers, that can be plugged into the system.

c. Network and Interfaces

Data and control information exchange between different workstations in a system is provided by a Local Area Network (LAN). Recent developments support Ethernet with TCP/IP protocol as the most widespread installation (10 Mbps). Ongoing developments will support optical networks with about 100 Mbps as a high data rate backbone (FDDI = Fiber Distributed Data Interface). Special software like NFS (Network File System) additionally guarantees the transparent share of disks and devices within the network.

Disks shall be connected via SCSI (Small Computers System Interface) or SMD (Storage Module Drive) or their derivatives. Recent developments are going towards higher data rates (>4 MBytes/sec) and increased control functions of SCSI.

d. Data Display Systems and Man/Machine Interface

A common trend in workstation terminal technology is the utilization of high resolution (approx. 1K * 1K pixel) bit mapped displays with color features. These screens offer multi window environment and graphical menue representations.

Especially for satellite image data visualization display systems have to provide at least 8-bit depth colors (256 out of 16 million). Preferably, 24 bit RGB true color should be used. Appropriate for true color display and for many applications is the integration of image processing boards, which provide additional processing power and number crunching capabilities. These boards shall be compatible with the workstations bus system (e.g. VME-bus).

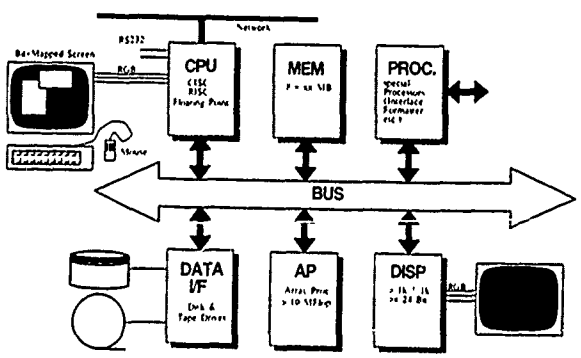


Fig 2 · Hardware layout of a basic image processing workstation

5. EOS Workstation Software Concept

As shown in the previous chapter, a network of distributed, dedicated workstations can be used to fulfil the computational tasks within an EOS ground data processing system. Obviously, as far as the price/performance ratio is increasing, the cost for software becomes significantly higher in comparison to the hardware. This will especially be true, if the design and integration of application software do not follow certain software engineering aspects and common standards.

A distributed system supporting dozens of different projects and used by different groups of engineers, easily becomes unmanageable, if each part of software is integrated within a different SW environment. Additionally, this would be contradictionary to the EOS requirement of system interoperability.

The interoperability and exchange of information can be realized, if the software system consists of a set of different layers or shells, beginning from the application menue and going down to the kernel functions of the processor A link between systems can be establisheded with two common parts of shells (interface driver) and software can be transformed from one machine to the next higher comparable shell of another machine.

As allready mentioned, an operating system which extensively supports this shell concept is UNIX with its derivatives. UNIX is supported by most workstation manufacturers Some major aspects of SW shells are stated below:

a. UNIX - Operating System Kernel

These basic shells cover hardware of the specific implementation and serve UNIX system routines UNIX implementations are available from PC to super-mainframe computer and are standard for the workstation domain. Despite slight differences in various UNIX implementations, application systems can be transformed easily, if they are based on UNIX.

b. Communication Drivers

The network communication SW concept itself consists of a set of different standardized layers. In the following, special emphasis will be put on the flexible application layer NFS (Network File System). NFS is widely implemented (even on non-unix systems) and guarantees transparent access to data and programms between linked machines. This allows a number of workstations to share a common data base, a special peripheral device or to exchange control status information in a complex processing system.

c. Programming Languages

Because UNIX itself is written in C, this programming language (Walker, 1984) is a standard for each UNIX implementation. Especially critical interface control structures can easily be transfered to other machines if written in C.

The need for structured programming, easy software maintenance and new object oriented programming made ADA a new standard in programming, especially for spaceflight applications (Pyle, 1981).

Besides, other programming languages like FORTRAN 77 and Pascal or LISP and Prolog are available on many UNIX implementations. They should be chosen where applicable.

d. Application Systems

Although a wide range of different sensor raster images will be produced by EOS, some common requirements in image processing are obvious for most data sets. Operational image processing can be subdivided in specific tasks, which can often be computed by pipelined modules. To meet the requirements for interoperability and flexibility in EOS image data processing it is therefore recommended to use modifyable and exchangeable image processing building block with a standard application shell, rather than to design a new end-to-end system for each sensor. Systems for image processing (Wharton, 1988), Geographic Information Systems (GIS World, 1988) and Data Base Management (ORACLE, 1986) are available, but must carefully be obeyed with respect to flexibility, maintainability and throughput power.

e. User Front Ends

User front ends for image processing applications are often integrated in image processing application systems as refered to above. Unfortunately, there is no commonly supported hardware or software standard in raster or image data display up to now. Portability of this software is very complicated and often implies completely new coding.

Standards in graphics such as GKS (Graphical Kernel System) and for window-systems in multiprocessing environments, such as 'X-Windows' and 'Look-and-Feel' are more and more anticipated and will hopefully be supported by all manufacturers of workstations (Kannegaard, 1988).

The standardization of remote sensing data formats is tackled by some international organisations Especially for SAR the CEOS (Committee for Earth Observing Satellites) agreed to support a common 'family' format for the forthcoming SAR sensors.

6. The Processing and Archiving Facility for SAR Data at DLR

The hardware and software requirements for EOS image processing systems, described in this paper have widely been adopted to the design of the German Processing and Archiving Facility (PAF) located ad DLR in Oberpfaffenhofen. The workstation concept is realized for all data processing functions.

Figure 3 gives an overview of the workstations envolved in the ERS-1 PAF. Approx. 50 workstations are already installed, most of them as target installations for the PAF. The distinct function of all major elements of the PAF is discussed elsewhere (Gredel, 1989). Here, we want to describe the geocoding system (GEOS) which is basically adaptable to other sensor requirements due to its general image processing capability.

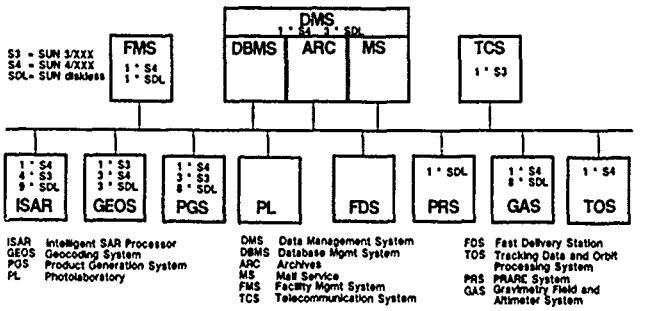


Fig 3.: Overall workstation installation at DLR ERS-1 PAF

As nearly all parts of the PAF, the geocoding system consists of SUN workstations. The software development and data base functions are hosted on SUN-3 WS, whereas the image processing functions are integrated on SUN-4 WS (figure 4).

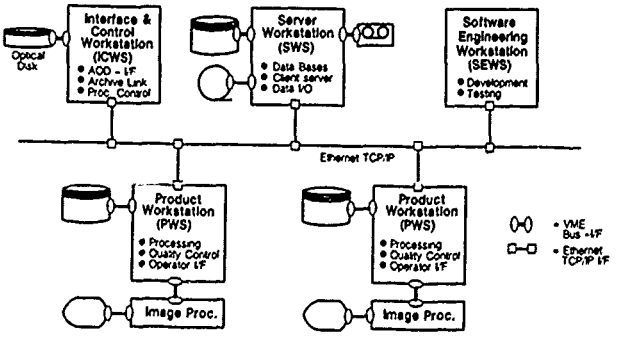


Fig 4 : Workstation HW of the Geocoding & Image Processing System

The Server Workstation (SWS) hosts the data bases needed for geocoding (Elevation Model, Ground Control Points) and provides the disk storage capacity for the diskless Software Engineering Workstations (SEWS) These SEWS serve as the engineers text, graphics, software development and test tool. Major parts of the software can be tested and compiled on these SEWS and easily be transfered to the target machines.

The Interface and Control Workstation (ICWS) is especially suited to perform the data I/O to the optical disks (AOD) which are used as archive media and to control and monitor the tasks performed on the Product Workstations (PWS).

The Product Workstations (PWS) are the workhorses for image processing. In principle all tasks of SWS, SEWS and ICWS can be handled by the PWS. Intentionally, these tasks are separated to ensure flexibility and to optimize operational production.

A typical PWS is integrated on the basis of a powerful SUN-4 workstation, equipped with enough main memory (at least 32 MBytes) and sufficient disk capacity for operational purposes (for ERS-1 geocoding: at least 1GByte).

Each PWS is equipped with a color monitor for interactive operation and image display. To speed up the image processing tasks, a 'Transcept Application Accellerator' (TAAC-1) is integrated, adding 8 MByte of data/image memory and approx. 12 MFlop to the system. Digitizers and image quick look facilities complete this image processing system, suited for SAR data geocoding.

The Software Concept is based upon the UNIX operating system (SUNOS 4.X) and the NFS features. The latter ensures, that the data read and written by ICWS are transparently accessable by PWS and that reference data hosted by SWS can be used by PWS as if the appropriate data disks are directly connected to the image processing workstation.

As throughout the whole PAF, ORACLE is used for data base implementations. The image processing software shell UPSTAIRS – a product of a DLRs working group (Klein+Stekl, 1988) – is used for integration of application tasks. The kernel of UPSTAIRS is mainly written in C, thus it takes only 1 day to transfer the whole system (approx. 100 MB of execode) from SUN-3 to SUN-4 systems.

Most image processing application modules of UPSTAIRS are written in FORTRAN, (intentionally from the first integration of UPSTAIRS on DEC-VAX computers) though C or other programming languages are possible. Upgraded features of UPSTAIRS allow to utilize the system both, as operational batch production tool and as highly interactive image processing system for scientific analysis.

7. Conclusions

We have presented our ideas concerning the requirements for flexible, operational image processing systems for the EOS era. It was pointed out that the need for standards in hardware and software is mandatory. Computing systems which are most likely adequate to cope with EOS data requirements are workstations. Nearly all manufactureres providing workstations apply common hardware standards and support portable software applications by using the UNIX operating system. It can be expected that the performance of processor hardware on one hand and of-the-shelf hardware boards on the other hand will increase accordingly to guarantee adequate EOS era image processing.

DLR is currently integrating such a processing system for SAR data, based on a workstation concept. Due to the experience derived therefrom and due to recent developments on the computer market we are well prepared to meet the requirements of the satellite data processing of the 90s.

8. References

1. Walker, A. N., "The UNIX Environment", John Wiley, 1984
2. Pyle, I. C., "The ADA programming language", Prentice Hall, 1981
3. Wharton, S W., et al, "The Land Application System (LAS) for Multispectral Image Processing", IEEE GRS, Vol. 25 No 5, 1988
4. GIS World, "The 1988 GIS Software Survey", GIS World Inc. Vol.1 No 1, PB 8090, Fort Collins, Co 80526, USA, 1988
5. ORACLE, "ORACLE Users Guide", Oracle Inc. 1988
6. Kannegaard, J., "Open Look: Outlook and overview", SunTechnology, Autumn, 1988
7. Gredel, J., Markwitz, W., Noack, W., Schreier, G., "Precision Processing of European ERS-1 SAR Data with high throughput", SPIE, Electronic Imaging, Los Angeles, 1989
8. Klein+Stekl, "UPSTAIRS Users Manual", Stuttgart, 1989
9. Dongarra, J., Martin, J.L., Worlton, J., "Computer benchmarking: paths and pitfalls", IEEE Spectrum July 1987

SATELLITE IMAGE PROCESSING SYSTEM (SIPS)

- DESIGN & DEVELOPMENT

T.C.Sarma, B.Lakshmi, D.S.Jain, B.Jaipal Reddy,
K.M.M.Rao, M.L.Mittal, K.Jairam Hebbar,
National Remote Sensing Agency,
HYDERABAD - 500037, INDIA.
TEL.NO:262572, TELEX:0425-6522

ABSTRACT

An Image Processing Work Station " Satellite Image Processing System " (SIPS) for the interactive analysis of remotely sensed, multi-spectral satellite aircraft data has been developed successfully.

The basic display processor features an image refresh memory in which the multi-band image data to be analysed is stored (upto 4 bands) in 4 planes of 1K x 1K pixels of eight bits resolution each. The selection of any band data from 4 memory planes to be input to the Red, Green, Blue colors is possible. There are three graphic planes of 1K x 1K pixels each for overlaying the text-annotation or any graphic information like boundaries etc on image data. Any area of 512 x 512 pixels from the storage can be displayed on the high resolution color monitor. A trackball controlled cursor is provided for annotation insertion, graphics & training set entry. The trackball is also used for panning and zooming of the images. A micro computer, Micro PDP 11/23 with 10MB Winchester disk, floppy drives and a terminal is used on the display controller.

Special care was exercised in system design, specifically the software, for easy interaction and user friendliness. The User interaction is completely menu driven and self explanatory with on-line HELP facility. This helps the uninitiated user to easily familiarize himself and use the system with minimal training.

The display is programmed to operate in Black & White, False Color or Color Sliced modes for raw data. It can also display classified data in assigned colors. Various image processing algorithms are implemented on this system which provides the capability for this work station to be used as a stand alone image processing work station. Based on the selection of the training sets, supervised classification can be carried out. Histogram generation and plotting,computation of statistical parameters etc are also provided for. Image enhancements using various techniques are realised by generating appropriate Look Up Tables (LUTS) and loading into the hardware for easy, quick on-line enhancements. Ratioing, principle component analysis etc can also be performed on this work station.

The system can be used either in stand alone mode or as an intelligent terminal to a 'Host' computer system. In the stand alone mode of utilization, the 'data' input can be through a 'floppy' disk or a magnetic tape(CCT). Regional Remote Sensing Centres in India will be using this SIPS work station as an image analysis terminal to the VAX 11/780 based system through an inter processor link to Micro PDP. When used thus, the data could be transferred through the 'Host' via the link and the image analysis carried out, 'Off-line' on SIPS. The result of the analysis could be transferred back to the 'Host' computer for further analysis or for storage.

This work station has been developed at the NRSA laboratory and the technology transferred to an Indian Small Scale Industry for productionization. The industry has completed the development of the prototype which has been successfully tested and evaluated by an expert team. The production units are being fabricated.

DIGITAL PHOTOGRAMMETRIC WORKSTATIONS

P. Lohmann

Institute for Photogrammetry and Engineering Surveys
University of Hannover

Abstract

Due to recent developments in the field of sensor and computer technology it seems now realistic to develop systems for the integration of photogrammetry, remote and GIS. The paper shows the design and development of such a system, which was based on a commercially available image processing system. The system includes a stereo display for the stereoscopic image data and three-dimensional graphics which might be generated by operator measurements or filed in/out to a geographic information system. Measurement and orientation programs as well as a package (hard- and software) for digital correlation and orthophoto production are part of the system. All operations are callable by soft menus on the stereo display using an optical mouse. Some first results obtained with this kind of photogrammetric workstation are presented. While the presented system design was very much depending on standard image processing hardware and components, new hardware developments in computer technology show, that processing power of super computers will be available on graphic workstation level. Available hardware components for stereo display, digital image storage, fast image access and computing power of some graphic workstations are presented. Transputer array technology and parallel processing of image data via parallel implementation of algorithms on these arrays permits even very time consuming tasks such as image correlation to be performed in near real time.

1. INTRODUCTION

The development of instruments in the field of photogrammetry in the last century was characterized by three phases. The first phase was the establishment of analog instruments being capable of measuring in analog photographs and producing line maps. This era of avoiding computations by opto-mechanical instruments was followed by the so-called analytical instruments, where still the images remained of analog type but the measurements and components of these type of instruments can be characterized as being computer assisted. Due to recent developments in the field of computer technology and because of the increasing importance of high resolution digital images the integration of digital image processing techniques, remote sensing and photogrammetry becomes possible. As an example some prototypes of digital photogrammetric workstations (DPW) have been developed in the last few years. The characteristics and advantages of such systems in comparison to analytical instruments can be expressed (Ref. 1 and 2) by the following items:

- no need for high precision opt-mechanical parts
- robust measurement system
- stable image geometry
- automatic, semi-automatic or manual operation
- high degree of interactivity
- data acquisition, storage and processing in one system
- on-line and real-time capabilities
- direct three-dimensional graphic superimposition possibilities
- possible use as data acquisition and processing instruments in GIS

Recognizing decreasing hardware costs, the possible advantages of such systems for the processing of digital images are the increase in accuracy, the high degree of automation, yielding in a faster availability of the results. A classification of the existing systems shows the existance of three major groups:

- Custom build systems, which are of great functionality and have a high performance, bute are generally very expensive.
- Systems based on conventional image processing systems, where image processing functions may be directly used, but which generally have a lack of openess with respect to hard- and software expansions.
- Modular systems, which are open with respect to the integration of hard- and software and generally can be considered very flexible, but which require a high effort in developing adequate user interfaces and data management software.

An overview of the existing systems is shown in Table 1

SYSTEM	NAME	REFERENCE
Custom build Systems	MATRA TRASTER T1ON DCCS	Euget, Vigneron 1988 (Ref. 7) Case 1982 (R. 18)
Mixture of Image Proc. & Custom build	KERN DSP 1 CONTEXT MAPPER	Cogan, Gugan 1988 (Ref. 9) Lohmann et al. 1988 (Ref. 3)
Image Proc.	GOULD DE ANZA DIPIX ARIES II KONTRON IPS 68K I²S Model 75	Alberts, Koenig 1984 (Ref. 10) El Hakim, Havetlock 1986 (R. 11) Gruen 1986 (R. 12) Gugan, Dowman 1986 (Ref. 13)
Modular Systems	DIPS II (SUN 3/110, MAXVISION) Sev. SUN based Systems	Gruen 1988 (Ref. 1) Dowman et al. 1987 (Ref. 14)
Mixture of Custom Build & Modular	TOPCON DCCS	Helava 1987 (Ref. 15)

Table 1: An overview of existing digital photogrammatric systems

2. THE DEVELOPMENT OF A DIGITAL PHOTOGRAMMETRIC WORKSTATION: THE CONTEXT MAPPER

2.1 System description

In designing a photogrammetric workstation one has to consider three major tasks, which a system should be able to work on (see Fig. 1):

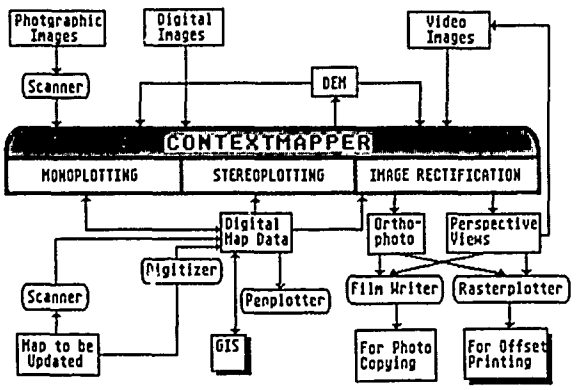


Figure 1: The CONTEXT MAPPER System

- MONOPLOTTING

  Monoplotting means a map revision system where digitized aerial photography or satellite imagery is transformed to the relevant map projection system and scale, taking elevation differences into account. An example for map revision is shown in Fig. 2 (Ref. 4).



Figure 2: Map revision using the CONTEXT MAPPER digital stereo-plotter

- STEREOPLOTTING

  Stereoplotting enables the three-dimensional evaluation of stereoscopic image pairs, automatical or operator assisted with the ability of stereo viewing of the model. In the current prototyp this is realized with two image processing displays being viewed by a stereoscope with free roaming images and fix floating marks. The operation of the workstation is controlled with a mouse and pop up menus (Ref. 3). Some of the main features of the stereo workstation are:

  - Stereo viewing
  - Handling of arbitrary image sizes
  - Real time zoom and roam
  - Sub-pixel measurement possibility
  - Vast set of photogrammetric orientation, adjustment and compilation software
  - Enhancement and feature extraction on-line
  - Automatic DEM generation (correlation)
  - Superimposition of 3-D graphics
  - On-line editing

- IMAGE RECTIFICATION

  Any images or maps in the system may be transferred to any projection system. If a digital elevation model (DEM) is introduced orthogonal projections in any coordinate system can be resampled. The mosaicing software enables the user to produce orthophoto maps from a series of adjacing photographs or satellite scenes.

Beside these three main tasks a variety of input and output routines for images, DEM, maps and other data are existing. An interface for the ESRI Arc/Info geographic information system has been developed to export map data from the GIS, superimpose the graphics to the images or in 3-D to the stereo-model, revise the data and import it back to the GIS.

The workstation is based on a commercially available image processing system GOP302 from TERAGON/CONTEXT AB. The host processor consists of a SUN 3/60 ™ supervisor processor with a 20 MHz MC 68020 DPU and 68881 floating point coprocessor running under UNIX 4.2 operating system. Special processors like the GOP (General Operator Processor) for fast complex filtering

and correlation and the GTP (Geometric Transfrom Processor) for fast geometric transformations and resampling can be accessed via the VME bus of the system. Two identical display control units (DCU) based on a MC68000 processor with a RAM memory of 1 Mega pixel (20bit), a display window of 512 x 512 pixel and a real time operating kernel(MODULA2) controlling the images as well as on-line image processing functions are part of the system. Due to the limited size of the display window, two DCU's have been selected, each controlling one color monitor. Both monitors are assembled in one cabinet and looked at with a mirror stereoscope. Because of the lack of DOVE prisma, the images have to be resampled to epipolar geometry, which is being done using the high speed GTP. The control of the image movement in x, y, z and the selection of processing functions is done via a three-button optical mouse. Digital elevation data (DEM) can be automatically generated by the use of digital correlation implemented on a specific filter processor (GOP). An example is shown in Figure 3.



Figure 3: Overlay of contour lines as correlated by the outomatic DTM-generator

3. DESIGN CONSIDERATIONS FOR FUTURE SYSTEMS

As referenced in the introduction the development of a DPW based on a standard image processing hardware is restricted in many ways, because in most cases the standard hardware architecture of image processing systems does not meet all operational requirements of a DPW. Three critical components will be looked at in more detail in the following sections

3.1 The display system

For precise measurement and operator comport the following parameters should be looked at during the selection phase of the monitor(s) (see Ref. 1):

- Good brightness
- High picture clarity and contrast
- Defocussing convergence
- Flicker
- Reflection glare
- Pixel non-linearity

A good monitor should be flickerfree, non-interlaced and have a refresh rate of more than 50 Hz. Examples can be found, like the SUN bitmapped display (1150 x 900 pixel), the AZURAY 2000/BW (2048 x 2048 pixel) or the HITACHI HM-6219 (2730 x 2048 pixel), to name just a few. High resolution is desireable to have a wide field of view (FOV). An image digitized with 12.5 micron on a 1024 x 1024 display gives only a FOV of 12.8 x 12.8 mm compared to standard 3 x 3 cm on photogrammetric instruments. For stereo display, only three methods are currently considered to be working operational, namely the anaglyphic method, the use of stereoscopes (KERN, TERAGON/CONTEXT) and the active and passive polarization techniques (TOPCON, MATRA TION).
In order to be comparable to analog/analytical photogrammetric instruments a sub-pixel measurement accuracy and method should be available, which can be achieved either by zooming or by performing an on-line resampling of the image data. The latter one implies a high processor performance which will be addressed in chapter 3.3.

3.2 Memory and Data Transfer Rates

In order to have as few disc I/O as possible, which will slow down operation speed, the image RAM memory should be as large as possible. Standard boards of 2 Mbyte (equivalent to 8 x 512 x 512 x 8 bit) are only a poor solution. Working with high resolution satellite images or digitized photographs will cause the system to be heavily engaged in reloading images to the RAM memory, while the operator roams over the entire image. This problem can be overcome by the use of bigger RAM boards or Real-Time-Video Discs. RAM boards in units of 16 Mbyte each are available and may be extended up to 256 Mbyte (GEMSYS 35 in KERN DSP 1). There are however even bigger RAM's available like the ISP1024 from DVS (see Ref. 5), starting with 64 Kbyte boards expandable up to 16 Gbyte with an internal data rate at the backpanel bus of the system of up to 480 Mbyte/sec. Real-Time-Video-Discs typically work at a speed of more than 10 Mbyte/sec. High speed movement of image plates in photogrammetric instruments is in the order of 2.5 cm/sec corresponding to a transfer rate of 4 Mbyte for digitized photograph at 12.5 micron. In normal measurement mode the speed is only 2 to 3 mm/sec corresponding to a transfer rate of 1.1 Mbyte for a 1024 x 1024 display.

3.3 Processor Speed

The total amount of floating operations in the CONTEXT MAPPER system during one loop including the resampling with bi-linear interpolation for sub-pixel measurement of a 16 x 16 window around the floating mark is in the order of 8000 Flops, resulting in a total of 240.000 Flops with 30 cycles/sec during normal measurement. In addition the on-line correlation and other processing functions, especially if the resampling window gets larger, demand a very high processing power. Standard processors like the MC 68020/68881 as they are used in the SUN 3 series computer offer a power of 3 Mips or 7 to 10 Mips (1.1 Mflops) on the SUN 4. Higher performances can only be achieved either with special VLSI processors (see Ref. 6) or by multiple processors computers, like the SUN TAAC-1 giving 25 Mips or 12.5 Mflops or the AT&T's Pixel Machine PXM 900, a multiple instruction multiple data MIMD computer with a performance of up to 820 Mflops.
Beside this dedicated processors the problem with multiple processors on bus coupled systems is, that

adding more and more processors on the same bus, does not linear increase the performance. From a number of 4 to 6 processors on the same bus on, the performance generally decreases. This can be avoided by using bus-less processors like transputers. The currently available INMOS T800 chips (see Fig. 4) give a performance of 1.5 Mflops/10 Mips each.

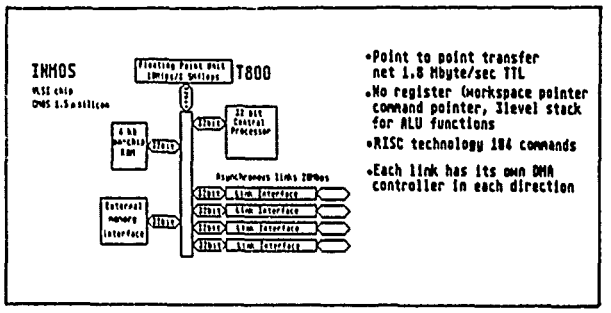


Figure 4: The INMOS T800 Transputer chip

They can be configured in any number and topology being appropriate for the application and they communicate via so-called links giving a net point to point transfer rate of 1.8 Mbyte/sec for each link. Arrays of transputers forming so-called computing clusters can be built in different topologies by using software controllable network configuration units. They may be used either as an add-on on standard graphic workstations or as stand alone systems of 64 up to 256 transputers. A 64 T800 super computing cluster has a performance of 640 Mips and 96 Mflops at a price of approximately 250.000 $. Parallel compilers of nearly all high level computer langues, like C, FORTRAN, PASCAL and PROLOG are available as well as a variety of software development tools.

4. CONCLUSION

The increasing performance of computer technologies has encouraged a number of companies and researchers to develop digital photogrammetric workstations. This is a development which can be seen in the light of the increasing importance of precise digital image processing of high resolution digital stereoscopic images for the purpose of map production and revision. Many of these developments have been based on standard available general purpose image processing systems. Due to the development of very powerful new processors and computer architectures, which have a processing power close to super-computers, and the existance of adequate display and storage capacities as well as algorithms and their knowledge based implementations it is expected that these systems will become operational very soon and shorten the time gap between data acquisition and the availability of results.

5. REFERENCES

1. Gruen, A.W.: Digital Photogrammetric Processing Systems - Current Status and Prospects, Proc. ISPRS 1988, Kyoto,Japan

2. Helava, U.V: On System Concepts for digital auto- Proc. ISPRS, 1988, Kyoto, Japan.

3. Lohmann, P., Picht, G., Weidenhammer, J., Jacobsen, K., Skog , L.: The design and implementation of a digital photogrammetric stereo-workstation. Proc. ISPRS, 1988, Kyoto, Japan.

4. Weidenhammer, J.: Aspects of digital photogrammetric workstations for GIS data acquisition and processing Proc. IGARSS 1989.

5. Digital Video Systems Hannover: The image sequence processor ISP 1024, DVS Hannover, Product description 1988.

6. Mueller, J.-P., Collius, K. A., Otto, G. P., Roleerts, J.B.G.: Stereo matching using transputer arrays. Photogrammetric Archives, ISPRS, 1988, Kyoto, Japan, P. 559 ff.

7. Euget, G., Vigneron, Chr.: MATRA Traster T1ON Digital Stereoplotter, Proc. ISPRS, 1988, Kyoto/ Japan.

8. Case, J.B., 1982: The Digital Stereo Comparator/ Compiler (DSCC). International Archives of Photogrammetry, Vol. 24 - II, PP. 23 - 29.

9. Cogan, L., Gugan, D., Hunter, D., Lutz, S., Perry, C.: Kern DSP1 Digital Stereo Photogrammetric System, Phot. Archives, ISPRS, 1988, Kyoto/Japan.

10. Alberts, J., Koenig, G.: A Digital Stereophotogrammetric System. International Archives of Photogrammetry and Remote Sensing, Vol. 25, A2, Commission II, 1984, Rio de Janeiro, pp. 1 - 7.

11. El-Hakim, S., Havelock, D., 1986: Digital image processing facilities at the N.R.C.C. photogrammetric laboratories. Int. Arch. of Photogrammetry and Remote Sensing, Vol. 26 - 2. pp. 155-156.

12. Gruen, A., 1986: The Digital Photogrammetric Station at the ETH Zurich. Int. Arch. of Photogrammetry and Remote Sensing. Vol. 26-2, pp. 76 - 84.

13. Gugan, D.J., Dowman, I.J., 1986: Design and Implementation of a Digital Photogrammetric System. International Archives of Photogrammetry and Remote Sensing. Vol. 26, Part 2, pp. 100-109.

14. Dowman, I.J., Gugan, D.J., Mueller, J.-P., O'Neill, M., Paramananda, V., 1987: Digital Processing of SPOT data. Proceedings 'Intercommission Conference on Fast Processing of Photogrammetric Data', Interlaken, June 2-4, pp. 318 - 330.

15. Helava, U.V., 1987: Digital Comparator Correlator System. Proceedings 'Intercommission Conference on Fast Processing of Photogrammetric Data', Interlaken, June 2-4, pp. 404 - 418.

Aspects of Digital Photogrammetric Workstations for GIS Data Aquisition and Processing

J. Weidenhammer

Institute for Photogrammetry and Engineering Surveys
University of Hannover

Abstract:

Enormous performance progress of computer systems in the recent years is the reason to use more digital map data. The nature of digital image aquisition systems demands digital photogrammetric intruments. Since SPOT offers high accurate satellite image data it does make sence to use digital image processing systems for measurement as well as analysing purposes.
The direct data flow from a digital photogrammetric aquisition instrument to digital maps has many advantages for monitoring and processing map and other information. Important features of digital systems are the possibilities to use fast prcoessing units for computation, interactive aquisition or update of GIS data. On the other hand the structure of data flow between components is getting more complex and needs therefore standardization and structuring within a wide context.
The following paper presents aspects of realizing a digitial photogrammetric workstation considering the environment of a conventional image processing system and GIS components. Practical implementation of such a system was a joined project with the University of Hannover and the image processing company TERAGON-CONTEXT. The system is based on a SUN- UNIX machine linked with special processors for fast rectification, correllation and interactive manipulation of images. Data collected by this system can be transfered via communication links to a geographical information system. There the opportunity is given to add the new information to existing data. Transfering data from a GIS to the image processing system offers the possibility to compare existing data with image information.
First experiences with the system and future development perspectives are presented.

Computer system overview:

The developed photogrammetric worstation is based on the GOP 302 image processing system. This system consists of a SUN-3 host running under UNIX 4.2 and serveral optional processors. Communication between the processors is realized by the 32 bit standard VME bus with a maximum transfer rate of 20 megabit/second. Modular designed hardware allows system configuration to particular needs. On the software side image processing tools and drivers are available; programming of processors is possible in higher languages instad of microcode. An overview about the system is given in figure 1.

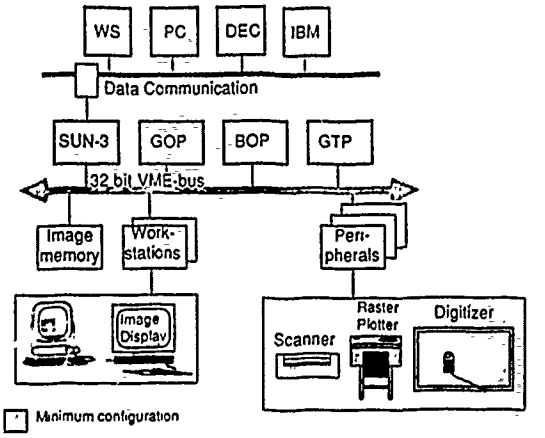


Figure 1 : Computer system overview

Conflict:

Data derived from photogrammetric instruments are one possible source for geographical information systems .
In a GIS many functions of displaying and manipulation of vector information are used and currently improved. On image processing systems we can find methods of extracting specific information from images.
GIS- hardware is normaly not considered to handle raster information, image processing systems are usually limited in graphics.
This conflict is solved by communication of both systems. Their advantages are used in the following way:

- Image processing system for data aquisition and update
- Adding GIS components to the image processing system
- Networking
- GIS for analysing and manipulation of vector information

Solution and required hardware:

Practically this is done at the photogrammetric workstation in the following way:

- The GOP filterprocessor serves mainly image filtering operations. Fast correlation is done by using the floating point part of this processor with about 0.27 seconds correlation time for 25 x 26 pattern and 50 x 50 search matrix size. The implemented method is product moment correlation (4).
- The geometric transform processor is used for resampling of images. After measuring the inner and relative orientation the stereo image pair is transformed to epipolar line geometry. Resampling an image of 2300 x 2300 pixels takes about 5 minutes.
- The display control processor is used for interactive image manipulation, manual measurements and visual control. The use of correlation supports measurements. The graphic capabilities of the display processor are used for GIS data overlay.
- The host system is used for supervised control and specific computations.
- A network enables communication to local or remote GIS systems.

Design, realization, test:

Measurement processes are classified in 3 categories:
- automatic
- semi - automatic
- manual

Therefore a design of interactions between those processes and communication protocols were defined. Offering programming utilities requires also a software interface without the need of specific system programming. A complex part of the stero workstation is the image and graphic display system with its interactions. The layout is shown in figure 2.

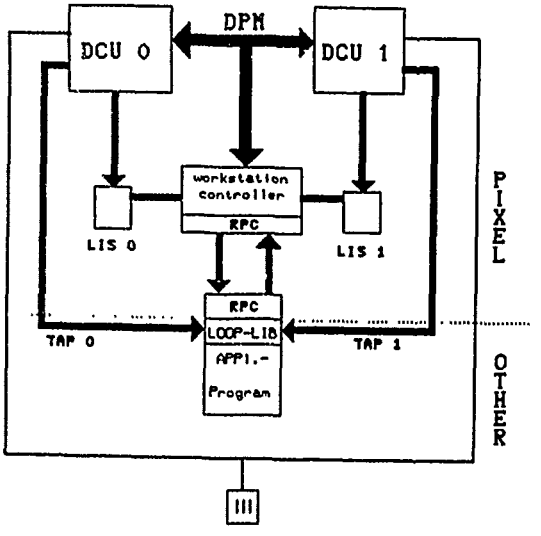


Figure 2: Control and data flow

Here we find hardware specific as well as hardware independent parts. The whole image and graphic display management is embedded into a controlling process running in the background of the system. Communication between processes is realized by message queues. Application programs communicate with the system over remote procedure calls so that no specific system knowledge and programming is required. Using different processors for displaying images and control has advantages in operation convenience but needs a careful synchronisation and data management. The 2-display stereo solution also offers a maximum viewing field for each of the stereo partners. For test purposes aerial photos were digitized at the OPTRONICS P 1700 scanner/plotter and results were compared with conventional derived data.

Operator interface:

Operator control at the stereo workstation is realized by screen menues in conjunction with the optical mouse. The 3 buttons of the mouse are loaded with special functions such as accepting/ignoring points or moving in x, y, z. A great help for manual measurement is the possibility of interactive image manipulation during the measurement. Most of the display functions such as zoom or table lookup operations are performed in real time and can be expanded to particular requirements. The maximum displayable image size is 512 x 512 pixel. Free realtime roaming is possible within an area of 974 x 960 pixel. Therefore image loading techniques from disk to memory either controlled or automatic were added to the display functions. For precise measurements a special roam mode is selectable: Mouse movements are interpreted as fractions of a pixel. The neighbourhood around the floating mark is then resampled according to a defined weight function and a better adjustment of points is possible. Howewer, this mode takes time because of the poor floating point capabilities of the display processor. Figure 3 gives an overview about the menues for image manipulation and photogrammetric functions.

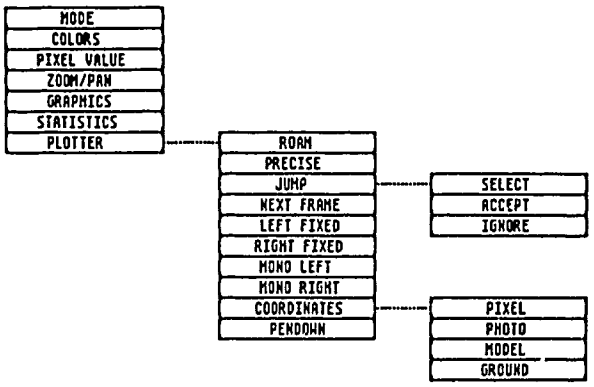


Figure 3: Menue system

Photogrammetric program packages:

With the modifications of the image processing system the implementation of a basic photogrammetric environment was possible. The range of programs covers orientations, rectifications, orthophoto, dtm generation, bundle block adjustment and graphical editing programs.

GIS features:

An important part of the system is GIS communication. The standard ethernet is not only considerable as a cable, it's a bus. Communication routines to the ARC/INFO system were developed so that data can be transfered to or from a remote system. Data stored in a local database on the image processing system can then be used as overlay information at the stereo or mono workstation. The local database serves as a tool for combining image and graphic information. Here a subset of GIS functions is available. An example of using existing GIS data as overlay information for map revision is shown in figure 4.



Figure 4: Map revision

Status:

The system structure is fisished, enhancements and applications within different fields are in hand. Presentation opportunities of the system were given at the Hannover Fair in April 1988, at the ISPRS Congress in Kyoto/Japan in July 1988 and at the Eurocarto 7 in September 1988. With the announcement from the manufacturer of the GOP 302 system to stop the hardware development of the GOP 302 system, the present advantages and limitations of the system are fixed. Image processing systems of the next generation offer higher resolution and more possibilities of combining images and graphics so that portation of the software to a new system is planned.

Requirements for future systems:

The future of photogrammetric workstations will depend on different conditions:

- digital image data: extensive distribution of digital image data and increased computer performance is the presupposition for the success of digital based measurement techniques.

- networking: communication is not only necessary between computer hardware; common data formats and structures have to be defined and used (!!) within a wide context.

- existing systems should be upgraded with a common data exchange possibility for other systems.

Literature:

(1) P. Lohman, G. Picht, J. Weidenhammer:

The design and implementation of a digital photogrammetric stereo-workstation, Photogrammetric Archives, Int. Congress for Photogr. and Remote Sensing, 1988, Kyoto/ Japan

(2) G. Konecny, P. Lohmann, L. Skog:

A Digital Mapping System. Photogrammetric Archives. Int. Congress for Photogr. and Remote Sensing, 1988, Kyoto/ Japan

(3) Context Vision:

Mapping, 1988, Kista/ Sweden

(4) V. Sasse, G. Altrogge:

Realisation of automatic correlation within a digital stereo plotter, Photogrammetric Archives, Int. Congress for Photogr. and Remote Sensing, 1988, Kyoto/ Japan

Image Processing Workstations and Data Bases for Quality Control of Geocoded Satellite Images

G. Schreier

DLR, WT-DA, German Aerospace Research Establishment
D-8031 Oberpfaffenhofen, FRG

1. Abstract

In the framework of processing facilities for future earth remote sensing sensors (ERS-1, SIR-C/X-SAR), the geocoding of Synthetic Aperture Radar images becomes an important aspect for supporting user data needs. DLR will offer such images, tackling with a wide range of processing and cartographic parameters.
As geocoding claims for comparability with topographic references, the geometric quality control of such data has to deal with image visualization, pattern correlation and topographic reference data bases, all together fast and flexible in access and manipulation.
The whole functionality of processing, quality control and data base management is integrated on high performance UNIX workstations, which allow transparent data access within large networks and visualization in monochrome or true color.
The displayed images are compared with stored geographic references such as Digital Elevation Models, Ground Control Point Chips and topographic maps. The hardware and software tools for such interactive, as well as automatic quality control are presented. The design of data bases for the reference information and the current work in integrating such data to form a basis for a Geographic Information System will be explained.

Keywords: SAR Geometric Quality, Digital Elevation Model Data Base, Ground Control Point Data Base, Maps, Map Data Base, Image Processing Workstations, Interactive Display Software.

2. Introduction

An important task for the operational Processing and Archiving Facilities (PAF) of the forthcoming remote sensing satellites is the quality ensurance for the variety of products they are going to generate. This is also valid for the German off-line ground segment for ERS-1 and SIR-C/X-SAR currently integrated at DLR (German Aerospace Research Establishment) in Oberpfaffenhofen. Especially for these radar sensors, extensive calibration and validation campaigns are planned (ESA, 1989) and rigorous product quality requirements must be met throughout the lifetime of the PAF.
Besides sophisticated monitoring of all engineering and processing parameters, validation and product quality assurance need interactive visualization of image data and reference to real earth's targets This is especially true for the SAR geocoding, where the image pixel geometry is resampled to be directly comparable to a cartographic reference system (Winter, 1989).
Besides, geocoding systems are often coupled to further image interpretation tasks such as mosaicking or ice-flow mapping (Fily, 1987)
Geocoding also is the entry to data synergism within Geographic Information Systems (GIS), thus establishing interfaces to data base systems being capable to handle geobased information or reference data. Part of it are visualization systems which shall enable the user to browse the data in different presentation formats.
The following chapters describe such an implementation at the german PAF for the purpose of quality control of geocoded SAR images. After having defined the characteristics of geographic fidelity in geocoded images, the interaction of data bases and visualization systems based on a workstation concept will be shown.

3. Geometric Quality of Geocoded Satellite SAR Images

The term geometric quality in this paper refers to the absolute measured position of one or more targets in the geocoded image and their deviation from the real geodetic position of the targets on earth In general, targets are in general not only artificial corner reflectors but all types of natural and man-made objects normally depicted in maps (streets, rivers, shore lines etc).
The basic geometry of SAR images is characterized by the active runtime measurement of the backscattered signal and the treatment of the signals frequency content in SAR processors (Elachi, 1988) This geometry is called 'slant-range'-image and is aligned to the orbit in one direction and to the signal travel time ('slant') in the other one Assuming an earth sphere or ellipsoid, the slant-range direction can be stretched by simple trigonometric interpolation, thus producing a more natural view but not enhancing geometrical quality. The geometric quality of SAR images can be enhanced by georeferencing and geocoding. The first procedure gives the geodetic location of a few image pixels or the image frame, whereas geocoding resamples each image pixel to be aligned to a given cartographic reference system (Guertin, 1981).
The basic image generation technique of SAR images enables georeferencing and geocoding systems to generate images with high geometric fidelity without any ground control pointing, provided the satellite orbit, the slant-range timing and the processor reference data are given very precisely (Curlander, 1984).
Unfortunately, the SAR image generation is severely affected by terrain undulations, introducing such effects as foreshortening, layover and shadow. These are based on a misregistration mainly in slant-range direction Δs, which can roughly be calculated as:

$$\Delta s = \Delta h \cdot \cot \theta ,$$

where Δh is the height difference and θ is the incidence angle of the SAR. (Example for ERS-1 : $\Delta s \approx \Delta h \cdot 2.35$)

Thus, even precise ephemeral and processing data could not give a pixel by pixel accuracy especially in hilly terrain. Terrain elevation effects could be tackled by taking a Digital Elevation Model (DEM) into account and calculating the SAR pointing function for each pixel rather than refering to an ellipsoid. These products could reach accuracies needed for topographic mapping, provided an accurate DEM is given and some geometric adjustments (using a few GCP) have been performed (Meier, 1985). These products are refered as GTC (Geocoded Terrain Corrected), whereas the products calculated with mere ellipsoid reference are called GEC (Geocoded Ellipsoid Corrected). Obviously, SAR Geocoding systems have to deliver three main further data sets: Digital Elevation Models (DEM), Ground Control Points (GCP) and Map References to ensure geometric quality.

3. Georeferenced Data Bases

3.1. The Digital Elevation Model Data Base

Digital Elevation Models (DEM) are used to enhance the geometric accuracy for the GTC-prodcuts.
DEM raster data are normally stored in a rectangular grid reference frame which is either cartographic or geographic The latter has the advantage, that any further cartographic projection can be calculated easily based on geographic reference. Additionally, latitude and longitude are well defined all over the world.

For operational satellite data processing systems, such as the SAR geocoding system of D-PAF, it is envisaged to store all DEM data online on magnetic disks. Today's high density storage technology makes on line storage feasible. For some land areas, DEM raster resolutions and an amount of 16 Bit encoded height data are given:

- FRG 248.601 sqkm 20m raster 1.243 MByte
- Europe 10 500.000 spkm 100m raster 2.100 MByte
- World 149 700 000 sqkm 250m raster 4 790 MByte

The requirements to be met when the DEM data base was designed are as follows:

- all DEM data is stored in geographical coordinates
- DEM data may have different resolutions for different areas
- spatial resolution is restricted to some (23) fixed values from 1" to 1 deg., to ease DB-organization
- height values are stored as 16 bit with definable offset data
- in flat areas (dh < 255 m), heights are stored as 8 bit numbers to reduce the amount of data
- for any location of the earth, only one valid height is stored
- to allow storage of multiple data sets, 99 test and 99 production data bases can be handled.

The data are organised in a binary tree structure, where the leaves of the tree, the nodes define different resolution sizes (see figure 1)

Nodes can be distinguished as index nodes, which point to sub-index nodes , or data nodes, which point to data buckets. The header of each node contains the geographic coordinates of the area boundaries, the resolution in north-south and east-west direction and several flag bits, indicating, whether the data set is complete. Additionally, some information on data update and auxiliary information such as data origin and quality is given.

Finally, DEM data itself is stored in data buckets of 4 Kbyte each. One node may point to up to 36 data buckets, giving 72 K-values of 16 Bit data. For level 6 nodes, this results in a finest possible resolution of approx. 0.25" DEM raster size to be stored (approx. 5 m at mid geographic latitudes) Each data bucket contains information about the offset of the height values, the data type and data organization. The offset of the height values is selected in such a way, that unknown or invalid values can be stored as 0.

Physically, up to 32K data buckets are stored in one extension' (= file). The actual design allows for up to 255 extensions, giving a maximum amount of 16 G-values of height data. These data base extension files can reside on different disks of one computer or even on different computers in a network (with NFS-software = Network File System of SUN Microsystems). Such a networked data base server may be an optical disk jukebox with several Gbytes of accessible data.

The access to the DEM data is realized by high level language procedure calls. Such calls can be performed by FORTRAN or C to read and write any particular area stored in the DEM-DB. All references to the area are given in geographic coordinates.

3.2. The Ground Control Point Data Base

The ground Control Points are used to refine the basic SAR projection function and to verify the geometric quality of the geocoding process.

The UPSTAIRS image processing system (Klein + Steki, 1988), installed at the DLR and used as a basis for the integration of geocoding software, actually supports the concept of GCP storage and management in different ways. So far, GCPs are understood as entities derived from distinct map (paper) sheets, establishing the reference between a map derived location on earth and the pixel coordinates of the digital image to be rectifed. Further information may be added to the GCP characteristics

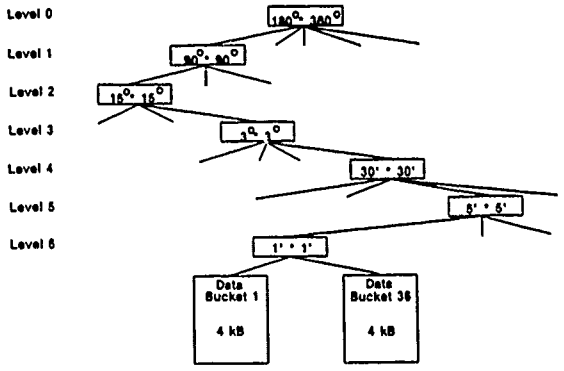


Figure 1: Organisation of the DEM-DB in geographical units

The SAR geocoding and quality control system will primarily be based on this implementation. Additionally, features are envisaged to handle GCP-chip information. A 'chip' is a small geocoded image showing a remarkable feature on earth. By image processing and correlation techniques this chip is matched with the target image, thus giving a distinct tiepoint from the known chip location.

The concept of GCP-chips will be used in conjunction with the DEM-DB. The GCP-DB points to small areas in the DEM-DB (steep slopes, valleys), where a 'SAR-simulation' is performed. This simulated image, with its well known location, serves as a Ground Control Point for the geocoding geometry adjustment process (Schreier, 1988)

The retrieval of GCPs will be performed by a programmable function call, specifying the region of interest (all over the world) in a geographic frame and returning a list of GCP entries in the GCP-DB. From this, detailed information of the GCPs can be retrieved The creation of new GCPs is also performed via function calls from higher level programming languages.

3.3. The Map Library System

The Map Library System (MLS) is used to store all relevant parameters of map paper charts, which are suitable for quality control by direct map-to-geocoded-image comparison.

As the geocoding of satellite data means resampling of raster image data to a format comparable to map geometry, topographic maps are the best (and actual available) choice for geometric quality control of geocoded images. Unfortunately, the organization and handling of map paper sheets in an appropriate scale for the areas of interest is a very difficult task. Maps from all over the world come along with different cartographic reference systems, and differences in sizes, quality, annotation and scale. In operational computer networks, the organization of such diverse data should be implemented in a relational data base system.

Therefore, the requirements for an operational Map Library System for SAR geocoding are as follows:

- Store parameter and annotation of more than 7000 maps for Europe
- Ensure flexibility for different parameters of maps
- Store internal-ID and national map name
- Store geographic coordinates of map frame corners
- Store country and reference map projection characteristics
- Store physical archive location
- Store archive update and utilization of maps

Figure 2 shows typical topographic map series frames for different countries Figure 3 gives a compilation of all relevant maps of Europe, suitable for geometric quality control.

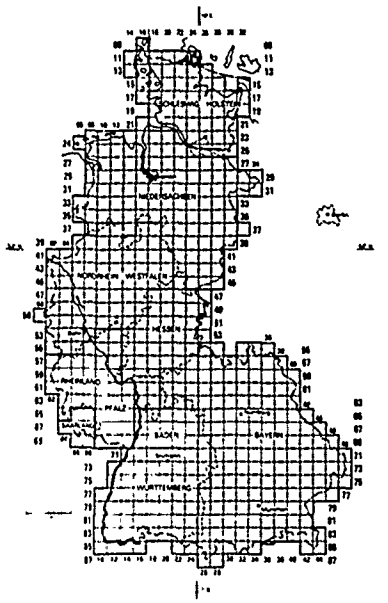


Figure 2. Topographical Map Series of FRG as sample input for the Map Library System

Topographical Maps for	Scale 1:50000 Number of Sheets		Scale 1:100000 Number of Sheets		Remarks
	all	of that published	all	of that published	
Austria	213	all	23	20	
Belgium	74	all	19	6	
Danmark	110	all	33	all	
Finland	615	all	349	all	
France	1094	about 950	293	all	
Germany-West	558	all	151	all	
Greece (only Geological M.)	about 420	about 210	---	---	
Geat Britain	204	all	62	all	without North.Ireland
Ireland, Northern	18	all	---	---	
Ireland, Republic	205	179 available	25	24 available	1 63360 inst of 1 50000 ; 1:126720inst of 1:100000
Italy	636	about 230	278	all	
Luxemb (only GeologicalM.)	---	---	1	all	
The Netherlands	110	all	---	---	
Norway	727	almost all	---	---	
Portugal	175	all	53	all	
Spain - topograph.	1106	about 275	296	about 150	
(- milar	1130	about 1015)			
Sweden	687 *	all	134	44	* part. Scale 1:100000
Switzerland	78	all	23	all	
TOTAL	about 7005	about 5388	1740	about 1487	without Spanish MilaryMaps

Figure 3: List of topographic maps of Europe

The implementation of MLS is performed in ORACLE on UNIX workstations (Wiesmann, 1988). Queries and entries to the database can either be performed by the SQL-query language or with a high level menue-interface called SQL*FORMS (ORACLE, 1986). Additionally, SQL calls can be embedded in FORTRAN programs, allowing direct access of application programs to the database. For the geocoding system, the major DB-calls are as follows:

- From a given national map ID, retrieve the map frame organized in geographic coordinates.
- From a given rectangular area, specified in geographic coordinates, retrieve all maps - or only maps with some specifications - which cover this area.
- From a given map-ID, retrieve the physical location of a map in the map archive.

The last two items are used for quality control of geocoded satellite images which will be performed as follows With the given corner coordinates of the geocoded scene, the Q/C program can query the data base for the appropriate maps The retrieved framing of the maps can be displayed as graphics on the workstation screen and the operator can interactively select a map. The physical archive location is used to search for the map at the right place.

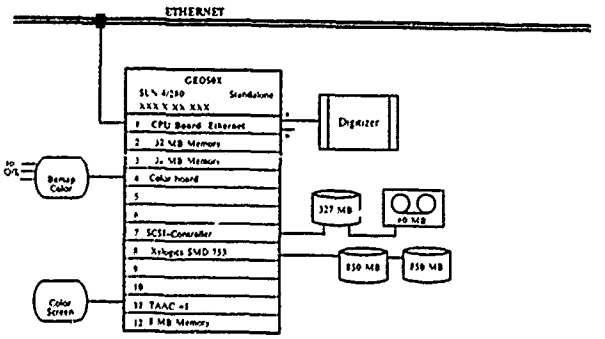


Figure 4: Image Processing Workstation Hardware

4. Image Quality Control Workstation Hardware

All data bases presented in this paper are installed on a network of UNIX workstation servers, used as target machines within the German PAF. These workstations are of SUN-3 and SUN-4 type.

Figure 4 gives the HW layout of a typical SUN-4 based geocoding and quality control workstation. The workstation comes along with a 10 MIPS RISC-architecture CPU and approx. 1.1 MFLOP arithmetic performance. For the purpose of ERS-1 geocoding, main memory is at least 32 MBytes and image data disk space about 1.7 GBytes. Operating system and programs reside on a smaller disk. Each workstation has a standard 1152*900 pixel color display of 8 bit depth (256 colors out of 16 millions). The availability of different colors is used to perform visualization and quality control. The reference points from map paper sheets are measured from an attached A0 digitizer Current plans foresee to upgrade the image processing and image visualization capabilities of this workstation by adding a special processor to the VME-bus of the SUN-4. This processor is the Trancept Application Accelerator (TAAC-1) (SUN, 1988), having true color display performance (32-bit depth) and 12.5 MFlop arithmetic throughput. First tests and SW integrations with the TAAC-1 have been performed.

5. The Geocoded Image Visualization Tool

On the basic of some prototypes and experiences with image visualization on color bit mapped workstation, an image visualization software was designed which is based on a display tool already used by University College London (Muller, 1988). This display tool divides the screen area in some major application windows (see Fig. 5).

- Main display window, where the original image can be displayed in several magnification or compression modes (Magnification factor from 16 to 1/16 including decimals).
- Overview image, showing a compressed (averaged) total image for orientation purposes. A box shows the actual part of the image displayed in the main window. This box can be used for scrolling the main image. An additional function allows a graphic overlay to be superimposed on the overview image. This actually displays the frames of the map paper sheets used for ground control pointing.
- Cursor tracing fields show the actual location of the cursor position within the main display in northing easting and longitude/ latitude This is possible for any magnification factor.
- A pull-down 'command & control window allows to display specific information about maps or control points
- Some buttons are associated with pull-down menues for magnification, cursor style, look-up-table selection etc.
- A tool, which allows to change LUT in monochrome and false color domain with various options.

The display-SW is optimized to handle large data sets (up to 12K*12K pixels), where the image window scrolling is performed with maximum

disk-I/O rate using memory paging mechanisms available under SUNOS 4.0

The real-time geographic coordinate tracing can be performed for any rectangular coordinate system of the geocoded image. The characteristics of this system and the image frame coordinates are supplied within a special data structure.

The implementation of this tool is performed in C using the 'SunView' graphical command syntax. However, it is planned to transfer this tool to the 'X-Windows' screen standard to have more flexibility in software and hardware platforms.

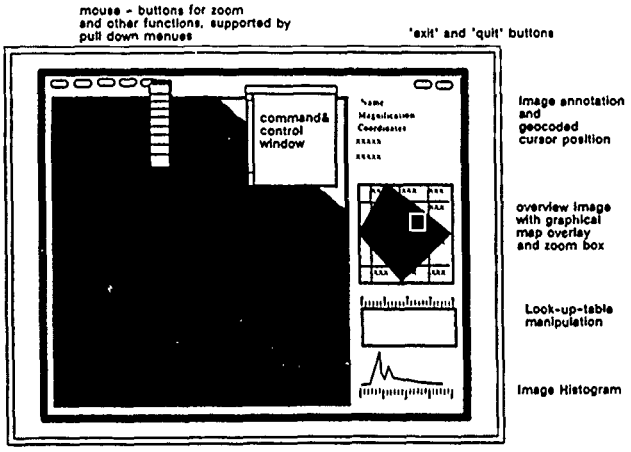


Figure 5: Screen Layout of the visualization tool

6. The Interaction of Visualization Tool and Data Bases for Geometric Quality Control

The data bases described in chapter 3 and the display tool described in the previous chapter will interact to form an operational tool for geometric quality control of geocoded images using map paper sheets. The procedure of the quality control can be described as follows:

1. Once the geocoded image is computed, its framing corner coordinates are known. By these coordinates, all map paper sheet data and the ancillary information are retrieved from the Map Library System, which might be suitable for control pointing.
2. The display tool appears on the operators screen, with subsampled overview image and superimposed map grid.
3. The operator can choose a map for tiepointing from the map grid information. MLS informs the operator about the physical location of the map in the archive.
4. The operator fixes the map on the digitizer attached to the workstation. MLS provides the information about corner coordinates and cartographic presentation of the map.
5. The display window shows that part of the geocoded image corresponding to the map area. The coordinates from digitized map points are transfered to pixel locations in the displayed image. This and the alphanumeric display of northing and easting is performed in a 'stream mode', thus moving the digitizer pen results in an automatic movement of the image cursor in a geocoded manner.
7. Once the operator finds a remarkable tiepoint in a map, he can fix it by a button stroke. Then he adjusts the screen cursor to the corresponding point and confirms it. The difference between 'mapped' and 'geocoded' point is stored in GCP-DB as 'residual'. The magnitude and direction of residuals in a geocoded image are an indication of for geometric fidelity.
   (If quality control is performed on ellipsoid corrected images, the terrain elevation for computing the height offset can be retrieved from DEM-DB.)
8. The operator can either select a new tiepoint on the same image (GOTO 7) or a new map for tiepointing (GOTO 3) or exit the display tool.

When exiting the quality control, a statistic about map-to-image tiepoints is computed. This statistic is delivered as an annotation data record to the end data user.

7. Conclusions

The current paper has tried to pinpoint the utilization of reference data bases and color graphics workstations for the geometric quality control of geocoded SAR images. It was outlined that current workstation hardware and software technology is well suited for the integration of such tools, especially in an operational environment. Nevertheless, further work must be done to include digitized vector data rather than map paper sheets and to also consider radiometric reference properties in the different data bases. Image display hardware and software must cope with these new possibilities by supporting multiple data visualization and overlay. The technological development in the computer market is going to support these features by providung adequate hardware and software.

8. References

1. ESA, "Calibration and Validation Task Definition", ER-RS-EPO-GP-1104, Feb, 1989
2. Winter, R., Markwitz, W.,,Schreier, G., Kosmann, D., "Satellite SAR Geocoding for Earth Mapping", SPIE, Los Angeles, Jan., 1989
3. Fily, M , Rothrock, D.,A.,"Sea Ice Tracking by Nested Correlations", IEE-GRS, Vol. GE-25, No. 5, 1987
4. Elachi, C ,"Spaceborne Radar Remote Sensing" IEEE Press, 1988
5. Guertin, F.,E., Shaw, E.,"Definition and Potential of Geocoded Satellite Imagery Products" 7th Canadian Symposium on Remote Sensing, 1981
6. Curlander, J. C.,"Utilization of Spaceborne SAR Data for Mapping", IEEE, GRS, Vol. GE-22, No. 2, 1984
7. Meier, E.,Nüesch, D R ,"Registration of Spaceborne SAR-Data to large scale topographic Maps, ERIM, Ann Arbor, 1985
8. Klein + Stekl, UPSTAIRS, Reference Manual, Stuttgart, 1988
9. Schreier, G , Kosmann, D., Roth, A.,"Design Aspects of a System for Geocoding Satellite SAR Images", ISPRS, Kyoto, Commission 1, 1988
10. Wiesmann, C.,"Map Library System, Version 0", Praktikumsbericht, DLR, 1988
11. ORACLE, "SQL*Forms Designers Guide", ORACLE Corp. 1986
12. SUN ,"TAAC-1 User Guide", Sun Microsystems, 1988
13. Müller, J. P.," Computing issues in digital Image processing in remote sensing", in "Digital Image Processing in Remote Sensing", London, 1988

Reduction of Ground Control Requirements for SPOT Imagery by Pass Processing

Bruce Sharpe and Kelly Wiebe

MacDonald Dettwiler & Associates Ltd.
13800 Commerce Parkway, Richmond, B.C.
Canada V6V 2J3

Abstract

Pass processing is a method to determine models for the geometric correction of satellite imagery. Most systems use a method we will call *single scene processing* in which the precision model is determined by marking ground control points (GCPs) within the area of the desired output scene. With pass processing, the correction model can be determined from GCPs located anywhere in within the same pass (orbit) as the desired output scene.

The goal of pass processing is to achieve an accuracy for a given number of GCPs that is comparable to that obtained with single scene processing. Since the GCPs are distributed over many scenes the average GCP requirement per scene is greatly reduced. Moreover, pass processing reduces the sensitivity to the location of the GCPs and allows for the correction of scenes which contain no GCPs at all.

This paper reports on the results of a study to extend the previous work on pass processing for Landsat imagery to SPOT and to measure the accuracy that can be achieved under various scenarios for the distribution and number of GCPs.

The results showed that the goals of pass processing have been achieved for SPOT: accuracies comparable to single scene processing were obtained using an average of 1/4 GCP per scene; large areas with no ground control were corrected.

Keywords: geometric correction, pass processing, SPOT, spatial triangulation

1 Introduction

SPOT imagery needs to be corrected to remove geometric distortions and in order to make the correspondence between feature locations in the imagery and on the earth's surface. The accuracy attainable using systematic corrections, that is, corrections based on *a priori* knowledge and orbit and attitude data, is only about 1 km. This error is mainly due to inaccuracies of the the orbit data and the fact that absolute attitude measurements are not available, only attitude rates.

Ground control points (GCPs) are features which are visible in the imagery and which have a known location on the earth's surface. Comparing their location in the imagery with their ground location allows for a refinement of the correction model so as to achieve subpixel accuracy.

The determination of the ground location of a GCP requires techniques such as surveying or photogrammetry, which can be expensive. In some areas, the determination of ground control locations may simply be infeasible. Thus in order to realize the potential of satellite imagery for cost-effective mapping applications, it is desirable to minimize the requirements for quantity and distribution of ground control.

One of the key techniques in MacDonald Dettwiler's approach to reducing ground control requirements is pass processing. Most satellite image correction systems use single scene processing in which correction models are determined using ground control only in the area of the scene to be corrected. With pass processing, ground control from anywhere in the same pass (orbit) can be used.

The goal of pass processing is to achieve the same accuracy as single scene processing while using about the same number of GCPs. Since there can be tens of scenes in a single pass, the average number of GCPs required per scene is reduced by an order of magnitude. Moreover, large areas with no ground control can be corrected with high accuracy.

Pass processing is similar to the concept of *spatial triangulation* in which blocks of stereo SPOT imagery are corrected. However, pass processing does not require stereo imagery and applies to one strip at a time. The techniques could be used as part of a spatial triangulation process.

Pass processing has been developed at MacDonald Dettwiler and implemented in our image processing systems for SPOT. This work extends the previous work for the early Landsat satellites [Friedmann *et al*, 1983] and the Landsat Thematic Mapper [Sharpe and Wiebe, 1988]. To assess the accuracy of the technique, a pass of SPOT imagery was corrected using various numbers and distributions of control points. The accuracy measured at independent check points.

2 Data

SPOT panchromatic imagery and ground truth were obtained for 15 scenes located in the province of Alberta (see Table 1).

The imagery was acquired at near-nadir viewing (incidence angle of −2.3°).

GRS number		Date
K	J	
537	231 – 246	September 24, 1987 (Day 267)

Table 1: Scenes comprising the study pass.

Ground truth was derived from 1:60 000 scale aerial photography for about 80 regions uniformly distributed throughout the imagery. About 6 GCPs per region were selected. There were 472 GCPs altogether whose ground truth was determined in this way. Their estimated accuracy was estimated to be 1–3 m in each of x, y and z.

3 Experiments

Once the image and geocentric coordinates of the GCPs were determined, subsets of them were selected to determine the pass processing correction models. We will refer to these GCPs as *model points*. The remainder of the GCPs were used as check points to measure the accuracy, as follows.

The correction models determine the transformation from image coordinates to geocentric location. The transformations were applied to the image coordinates of the check points and the results were compared with their ground truth location. The differences were decomposed into their along and across track components and statistics for each component were evaluated as a measure of the error in the model. Note that the elevations of all points were used so that there are no contributions from terrain errors.

The model points were selected so as to test the accuracy of the correction using varying numbers and distributions of them. The results are described in the sections which follow.

3.1 Uniform Distribution

For $N = 1..10$, N control points were selected as model points. These points were uniformly distributed over the pass. To increase the statistical sample, three different sets of model points were chosen for each N. In all cases, the chosen model points were those which were deemed to be clearly visible in the imagery.

Table 2 shows the accuracies measured. The combined error is the RMS of the along and across track errors, in other words, it is the two-dimensional RMS error, calculated over the whole pass. These errors (along, across and combined) are the RMS of the errors over all three runs.

The "min" and "max" columns show the minimum and maximum combined RMS error that was measured over all the scenes and all the runs.

The errors usually decrease with an increasing number of model points, as expected. A combined error at the subpixel level is obtained for 3 or more GCPs (1/5 GCPs per scene on average) although a subpixel level for all scenes is not obtained until 8 or more GCPs (≈ 1/2 GCP per scene) are used. It is noteworthy that the overall RMS error when using 9–10 GCPs is hardly larger than the model residuals when all 472 GCPs are used (5.9 m vs. 5.4 m). The latter represent the best results that could be achieved using this data.

#GCPs	Along	Across	Combined	Min	Max
1	6.3	22.8	23.6	15.8	37.6
2	9.4	26.1	27.7	14.7	45.8
3	6.7	7.4	10.0	6.5	14.0
4	7.0	5.0	8.6	5.7	15.8
5	6.1	6.2	8.7	4.5	16.4
6	5.0	4.4	6.7	4.2	13.5
7	4.7	4.3	6.4	3.5	12.3
8	4.5	4.6	6.4	3.5	10.1
9	4.5	3.9	5.9	3.5	10.0
10	4.5	3.9	5.9	3.5	9.5

Table 2: Two-dimensional RMS errors for models determined from a uniform number of control points. The along-track, across-track and combined errors are calculated over the whole pass. "Combined" means the RMS of the along and across-track error. The minimum and maximum errors are for the combined error when calculated scene-by-scene.

There are some anomalies in the results: the error increases in going from 1–2 GCPs and from 4–5 GCPs. These are most likely due to statistical fluctuations arising from the attempt to determine an inadequate model using a minimal number of noisy data points.

3.2 Interpolation

Ten different configurations of GCPs were tested, each of them three times. Basically, a small number of GCPs was used in each of one or two end scenes to correct the whole pass. Thus up to 13 scenes with no GCPs were corrected. We call this mode of processing "interpolation" because the correction model is interpolated between data points supplied at the ends to provide correction of the imagery in the middle.

The measured errors for the interpolation are shown in Table 3. The errors shown in Table 3 are defined similarly to those in Table 2.

Several conclusions can be drawn from Table 3:

1. Not much improvement in accuracy is gained by going beyond 4 GCPs. The overall RMS error is about 1 pixel or less for 4 or more GCPs.
2. It is somewhat better to have GCPs in two scenes at each end rather than just one.
3. Concentrations of GCPs in the south give marginally better results than in the north. This is probably because the northern GCPs were harder to locate in the imagery than those in the south and so were less accurate.

Summary: we were able to correct 11–13 scenes (660–780 km)

containing no GCPs to an overall accuracy of 1 pixel (2D, RMS) and with a maximum error of less than 15 m.

3.3 Extrapolation

Twelve different configurations of GCPs were tested, each of them three times. In this series of experiments, GCPs were selected from one end of the pass. The correction model was used to correct those scenes and extrapolated to correct the remainder of the pass where there were no GCPs. The GCPs were taken from 1, 2, 5 or 7 scenes and there were 1, 2 or 10 GCPs per scene.

The measured errors for the extrapolation tests are shown in Table 4. The errors shown in Table 4 are defined similarly to those in Table 2.

Extrapolation is the most difficult of the modes of pass processing and the accuracies achieved are not as good as for the uniform or interpolation modes. Nevertheless, reasonable accuracy was maintained for several hundreds of kilometers beyond the end of the control.

Our results can be summarized as follows:

1. Subpixel accuracy can be maintained in the along track direction for the entire pass (15 scenes) extrapolating beyond control contained in one or more scenes at one end. This is true as long as there are enough GCPs in the model (2 is not enough, 10 is more than enough).
2. Across track errors increase at the rate of about 4–5 m per scene away from the region of control. The errors appeared to be due to an oscillation effect. If this is the case, then the estimate of 4–5 m per scene is too pessimistic for inferring what the error would have been if we had extended the model further.

# GCPs			Along		Across		Combined		
	Per scene/# scenes								
Total	South	North	Min	Max	Min	Max	Min	Max	Total
2	1/1	1/1	5.0	12.5	20.7	46.4	21.5	46.7	34.2
4	2/1	2/1	4.3	11.8	2.8	14.7	6.1	15.6	10.6
10	5/1	5/1	2.2	9.1	2.8	14.0	4.3	14.5	9.3
11	1/1	5/2	2.7	7.7	2.7	14.3	5.2	14.8	10.1
11	5/2	1/1	2.3	9.5	2.5	13.3	3.5	14.1	9.3
20	5/2	5/2	2.2	8.1	2.3	11.5	3.3	13.2	8.2
20	10/1	10/1	2.2	9.5	2.6	14.1	4.3	14.5	9.1
21	1/1	10/2	2.8	7.6	2.5	14.3	4.5	13.5	9.5
21	10/2	1/1	2.5	9.9	2.4	13.0	3.5	13.7	9.1
40	10/2	10/2	2.2	7.7	2.3	12.1	3.3	12.9	8.1

Table 3: Interpolation errors. The GCPs were located in one or two scenes at the ends (south and north). Three sets of GCPs for each configuration were used; the RMS error of the three is reported. The minimum and maximum values are for the RMS errors calculated on a scene-by-scene basis. The combined error is the RMS of the along and across track errors.

# GCPs		Scenes		Along		Across		Combined		
Total	Per scene	#	Location	Min	Max	Min	Max	Min	Max	Total
1	1	1	N	4.2	15.2	10.5	50.8	13.5	51.5	29.0
1	1	1	S	5.0	14.4	8.6	52.4	10.0	52.6	29.5
2	2	1	N	4.2	15.1	2.9	44.3	5.9	44.7	22.2
2	2	1	S	4.7	13.0	4.6	58.8	6.9	59.4	31.2
10	10	1	N	4.4	6.3	2.5	43.3	5.9	43.5	22.3
10	10	1	S	2.3	9.5	4.0	54.3	5.3	54.5	29.9
20	10	2	N	2.1	7.7	2.3	42.3	4.9	42.5	22.0
20	10	2	S	2.1	8.3	3.0	40.3	4.2	40.7	25.8
50	10	5	N	2.3	8.1	3.4	39.9	5.7	40.0	20.0
50	10	5	S	2.6	8.6	3.7	33.6	4.5	34.0	20.6
70	10	7	N	2.0	7.4	3.2	38.1	6.2	38.3	18.9
70	10	7	S	2.0	8.7	4.1	27.1	4.5	27.5	16.8

Table 4: Extrapolation errors. The GCPs were located in scenes at one end of the pass, north or south. Three sets of GCPs for each configuration were used; the RMS error of the three is reported. The minimum and maximum values are for the RMS errors calculated on a scene-by-scene basis. The combined error is the RMS of the along and across track errors.

4 Conclusions

When good quality GCPs with a uniform distribution are used, it is possible to correct a pass of 15 scenes (900 km) to subpixel accuracy using only 1/2 GCP per scene on average. Using interpolation, we can correct 13 scenes containing no GCPs to subpixel accuracy. We can extrapolate control located in one or two scenes for 13 scenes maintaining subpixel accuracy in the along -track direction, and losing accuracy at the rate of about 4–5 m per scene in the across track direction. The error was typically about 50 m at the end of the pass opposite the control.

In summary, pass processing applied to this sample of 15 SPOT PLA scenes gave very satisfactory results for the cases of uniform distribution and interpolation. Extrapolation was less accurate, reflecting the difficulty of this kind of modelling. Nevertheless, the accuracies obtained using extrapolation would probably be quite acceptable for many applications.

5 Acknowledgements

The authors gratefully acknowledge the assistance of the Canada Center for Remote Sensing in selecting and acquiring the SPOT imagery for this study. This work was carried out under the Industrial Research Assistance Program.

References

[Friedmann *et al*, 1983] Daniel E. Friedmann, James P. Friedel, Kjell L. Magnussen, Ron Kwok, and Stephen Richardson. Multiple scene precision rectification of spaceborne imagery with very few ground control points. *Photogrammetric Engineering and Remote Sensing*, 49(12):1657–1667, December 1983.

[Sharpe and Wiebe, 1988] Bruce Sharpe and Kelly Wiebe. Reduction of ground control requirements by pass processing. In *Proceedings of the XVI International Congress of the ISPRS*, 1988.

SIMULATING MAPPING, SENSING AND FEATURE EXTRACTION IN A PC

T. N. Truske

BDM International, Inc.
1801 Randolph S.E.
Albuquerque, New Mexico, 87106

Abstract

A simulation model for testing feature matching and target extraction operations of low altitude radar and laser targeting systems is described which runs on a personal computer (PC). The model creates topographic maps with selected roughness and translates the map data into angle and range channel data of a targeting system. Model elements also represent noise in the sensor data channels. Other modules apply feature and edge extraction algorithms to the simulated channel data. A second channel provides reference data for correlation processing or for computing estimates of the consecutive positions of the sensor platform. (Key words: radar, sensor, simulation, mapping, image, feature)

1. INTRODUCTION

This work was motivated by the need for a test environment to examine the behavior of feature matching algorithms (Rosenfeld, V1, 1982) in smart missiles. The work included the design of the simulation model components discussed here and a missile dynamics, Kalman filter. The model was constructed to provide a tool which could be used on a PC with a math coprocessor but no video board or similar circuit card augmentation. The model provides a text bed, in the PC environment, for examining the behavior of different algorithms and operations to support the target tracking processes of smart missiles.

2. MODEL COMPONENTS

2.1 Map generation

The model operations are performed on a simulated ground map which represents profiles through a topographic map. The map is built on a data field of uniform random noise which is then smoothed by triangular weighting windows in the down (y) and cross (x) range directions. This process creates a surface cross-section which has reasonable properties in terms of the spatial correlation of the ground data.

The properties of the map are set by specifying: the extent of the field, the width of the smoothing window in x and y and the variance of the map contours. A ground feature of specified width, depth and height can be added to the map data field (Rosenfeld, V2, 1982). The map generation module does not consider earth curvature since the missile system operations represented here are for ranges where earth curvature is insignificant.

2.2 Missile position

The model specifies two missile positions in terms of its height, cross range and down range location. The respective positions can be used in a control loop for the vehicle dynamics or to compute position to position error signals based on the results of the feature matching operations.

2.3 Sensor parameters

The model requires that antenna beamwidth in elevation and azimuth, pointing angle and the elevation and azimuth antenna scan regions for both missile positions be specified. The type and level of noise in the angle and range channels of the system can also be specified (Berkowitz, 1965).

2.4 Translating field data to sensor data

The map of x, y and height data are translated into range modulation functions that correspond to the line of sight signal returns from unobscured terrain and ground features. The model determines the ground range and interpolated ground height on the lines connecting lattice points which intersect each elevation plane emanating from the missile position. This operation uses the map data in the x and y directions interpolated at the points where the azimuth track intercepts the map grid. The interpolated height values from the map are tested at progressively smaller depression angles to determine whether the ground point on the given azimuth track is obscured. The model then determines the angle and slant range to the visible points. The slant range and angle data define the range modulation function.

The range modulation function corresponds to the range data from the missile sensor in the absence of noise. In the actual missile operations the modulation function is contaminated with range and angle channel noise. Noise is added to the range modulation function data in the model to simulate the actual missile sensor data environment. The function data are then used to determine the ground height as a function of missile position and slant range to the ground. Ideally the range and angle channel data would provide values which represent the true field data weighted by the respective range and angle apertures of the sensor system. The model does not yet include range and angle channel signal aperture weighting. Figure 1 shows the operations of the model.

2.5 Feature matching operations

The model contains algorithms to identify man made ground features. The program encouraged the use of algorithms which were simple and robust. The algorithms: emphasize the horizontal and vertical feature elements and slopes, compute data intersections for the various feature and slope emphasis operations, apply gradient and Laplacian filters of arbitrary orders, apply amplitude weighted filtering using a priori feature information, and perform correlation matching of the data sets obtained from the other feature matching operations. The correlation function operations can be used to determine the change in antenna pointing angle between successive data observations. The statistics of the changes in antenna pointing angle between respective missile positions provide an error signal to support Kalman filter simulations of the missile behavior.

3. EXAMPLE SIMULATION RESULTS

3.1 Overview

Figures are shown for data fields which have respectively: no noise, moderate noise and high noise levels and combinations of ground terrain and target features with varying ratios of target to terrain signal to noise energy. Figure 2 shows the simulated ground map with a square feature in the center of the map. Figure 3 shows the same ground map modified by adding higher energy wavenumber components which overlap those of the desired target feature. This ground map is an example of a feature in the presence of a high level of terrain noise.

Figure 4 shows the observed map field height as a function of azimuth and elevation angle for two missile positions. The two positions are displaced in elevation by 10 degrees (the depression angle axis labeling refers to the leftmost data set) and in azimuth by 30 degrees. Figure 5 shows the map heights inferred from the slant range and depression angle measurements of the range modulation functions as a function of angle for two missile positions.

3.2 Noise free data

Figures 6, 7 and 8 show the respective results of the vertical slope emphasis, feature intersections and noise weighted operations. Particular points to note are the relatively clear feature identification obtained using the simple algorithms based on edge emphasis and data weighting. These results are expected with a high signal level feature in a low terrain noise and noise free sensor environment. The data weighted operation is effective because it uses the target feature data values to discriminate against other terrain surfaces.

3.3 Moderate sensor noise

Figures 9, 10 and 11 show the height from slant range, vertical slope emphasis and noise weighted operations for the case of moderate sensor range and angle channel noise. The noise in the angle channel is uniformly distributed with a maximum extent of one half the beamwidth in azimuth and elevation. The range channel noise is represented by a uniformly distributed error with a 10% maximum range extent. The noise introduces additional spurious feature components to the vertical slope emphasis data and broadens the extent of the noise weighted data operations. The range channel noise is seen by the system as high wavenumber energy similar to that of the target feature.

3.4 High sensor noise

Figure 12 shows the effect of high sensor range channel noise on the data after application of the vertical slope emphasis algorithm. Here the noise in the angle channel is uniformly distributed with a maximum extent of a beamwidth in azimuth and elevation. The range channel noise is represented by a uniformly distributed error with a 20% maximum range extent. A significant number of spurious feature components are present due to the higher channel noise. The spurious components are primarily due to the lack of inherent noise filtering in the algorithms.

3.5 Terrain noise

Figures 13 and 14 show height from slant range and noise weighting where the terrain has significant energy at wavenumber components which overlap those of the desired target feature. The poor feature identification reflects feature masking due to the similarity of the terrain and the target feature. Of the operations used, the data weighted operation is moderately effective because by using the information about the target feature properties (dimensions in this example) it is operating in the manner of a coherent detection process. It should be noted that other properties of the ground return, such as signal polarization, might also be used to obtain additional levels of target discrimination.

3.6 Correlation operations

The algorithm operations applied above were oriented to each individual field of inferred height data. To determine position changes in the missile from one data frame to another, we calculate the two dimensional correlation function of data which has been operated on by some of the feature enhancement algorithms. Figure 15 shows the result of correlating the intersections of the horizontal and vertical feature emphasis operations for the two missile positions. The data are based on the moderate noise level of one half a beamwidth in angle and 10% in range. The correlation operation correctly identifies the elevation (10 degrees depression) and azimuth angle (30 degrees) displacements between the two missile positions. The data demonstrates the benefit of the additional computational costs of the correlation processing.

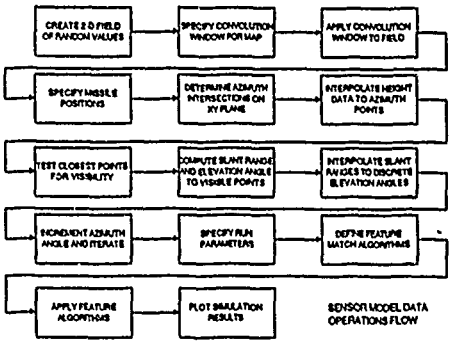


Figure 1.

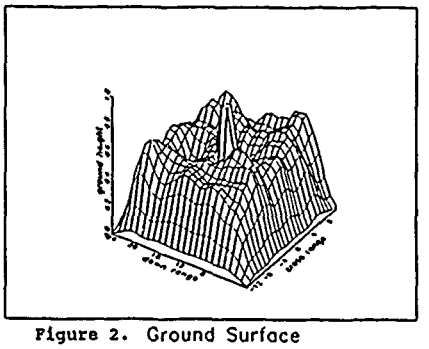


Figure 2. Ground Surface

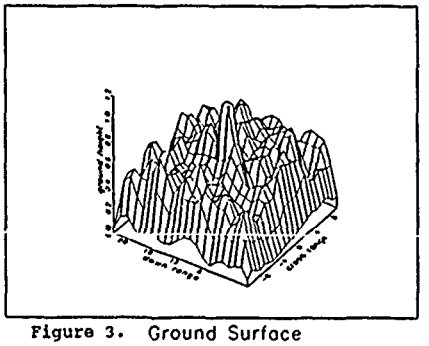


Figure 3. Ground Surface

4. CONCLUSIONS

A simulation model for testing feature matching and target extraction operations of low altitude radar and laser targeting systems has been developed which operates on a PC. The model includes modules for creating topographic maps with selected roughness and translating the map data into angle and range channel data of the sensor targeting system. Modules were developed which represent noise in the channels of the sensor system. Other modules apply feature and edge extraction algorithms to the simulated channel data. A second system channel represents a reference channel of external map data for correlation processing or consecutive estimates of the position of the sensor platform.

The model is programmed for the MS-DOS environment of an XT or AT class personal computer. Thus it is limited to relatively coarse representations of the ground feature maps and the number of data cells used to represent each sensor data field. In spite of these restrictions the model is useful in representing the qualitative and quantitative behaviors of various sensor systems and feature extraction algorithms.

5. ACKNOWLEDGEMENTS

The author wishes to express his appreciation for the support provided by colleagues and management of BDM.

6. REFERENCES

1. A. Rosenfeld and A. C. Kak, Digital Signal Processing Vol. 1, pp. 237-260, Academic Press Inc. Orlando, Florida (1982).

2. A. Rosenfeld and A. C. Kak, Digital Signal Processing Vol. 2, pp. 340-344, Academic Press Inc. Orlando, Florida (1982).

3. R. S. Berkowitz, Modern Radar, Analysis, Evaluation, and System Design, pp. 141-161, John Wiley and Sons Inc., New York, New York, (1965).

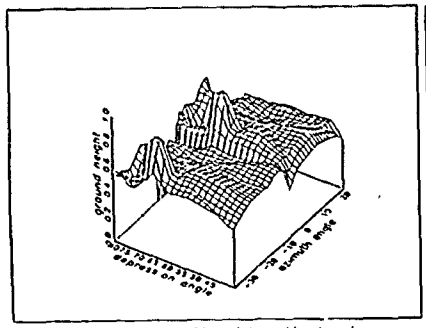


Figure 4. Observed Height with Angle

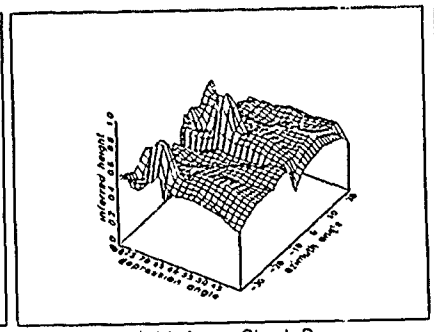


Figure 5. Height from Slant Range

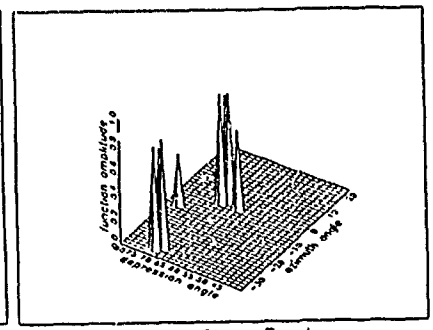


Figure 6. Vertical Slope Emphasis

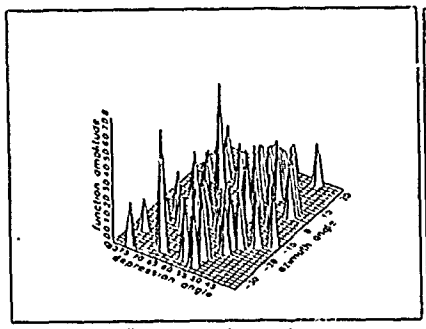


Figure 7. Feature Intersections

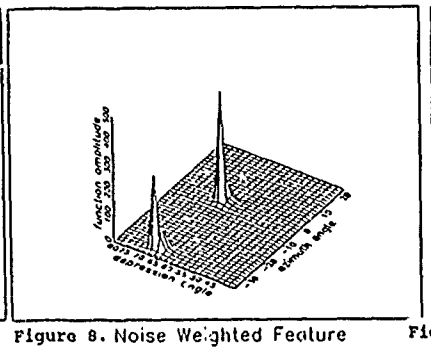


Figure 8. Noise Weighted Feature

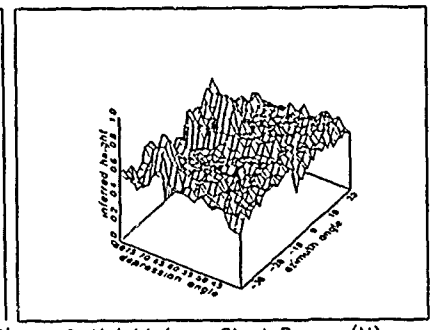


Figure 9. Height from Slant Range (N)

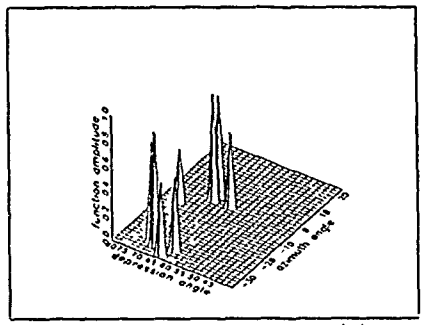


Figure 10. Vertical Slope Emphasis (N)

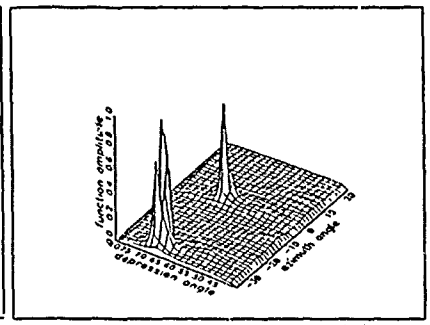


Figure 11. Noise Weighted Feature (N)

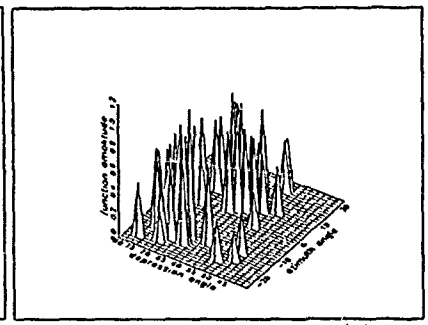


Figure 12. Vertical Slope Emphasis (N)

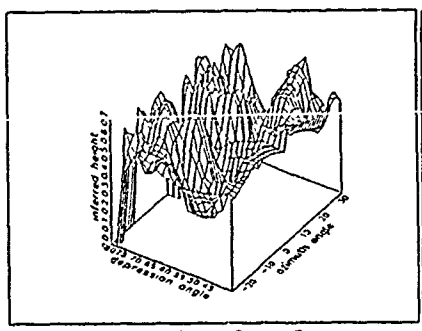


Figure 13. Height from Slant Range

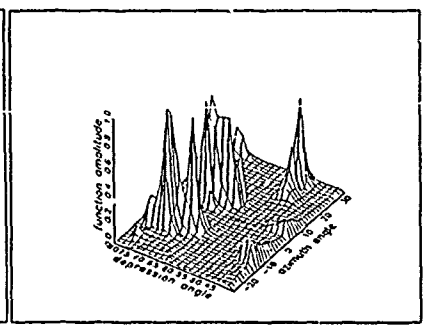


Figure 14. Noise Weighted Feature

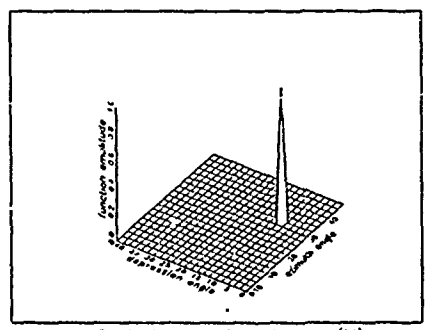


Figure 15. 2-D Correlation-hvfgn (N)

IMAGE ANALYSIS ON A MACINTOSH II.

David G. Barber, J. Douglas Dunlop, Joseph M. Piwowar and Ellsworth F. LeDrew.

Earth Observations Laboratory, Institute for Space and Terrestrial Science.
Department of Geography, University of Waterloo, Ont. N2L 3G1

ABSTRACT

Current research in remote sensing is directed towards understanding the relationship between image data and the physical features they represent. The Macintosh II provides a framework by which these relationships can be explored, implemented and tested quickly. Enhancements, filtering, image transformations and other image display functions allow for visual exploration of the image data. Numerical exploration is facilitated through statistical measures and graphs. Implementation of new and existing algorithms using numerical analysis programs allows the researcher to minimize the time between the idea and result. Rigorous quantitative analysis can be supplemented with visual presentation either as images or graphs. This combined visual and numerical approach and ease of movement between visual and numerical presentation, promotes the willingness to explore. In this presentation we illustrate image processing and analysis capabilities of the Macintosh II with a specific illustration of development and testing of texture algorithms for SAR sea ice classification We conclude with suggestions for stand-alone and networking configurations.

Keywords: Image analysis system, Macintosh II, SAR, Texture algorithms.

INTRODUCTION

In remote sensing applications research human interpretation is often necessary to assist computers identify features of interest. Feature identification usually consists of defining either deterministic or probablistic descriptors. Developing, testing and evaluating these descriptors is what keeps most of us employed.

Spatial and radiometric characteristics of modern digital imagery place limitations on the computing power required for image analysis. Until recently hardware requirements meant that image analysis was available only to those with enough resources to afford expensive image processing systems. With recent increases in processing and storage capacity of microcomputers, image processing is now available to a much more diverse group of scientists.

In this paper we present use of the new generation of Macintosh computers (Macintosh II, Macintosh IIx and Macintosh IIcx[1]) as typical examples of the evolution of the microcomputer into the arena of image processing. These computers (hereafter referred to collectively as Mac II) are particularly well suited for this work because of the intuitive interface which allows efficient manipulation of image data for both qualitative and quantitative analysis.

IMAGE PROCESSING ON THE MAC

The image analysis system we describe is in use at the Earth Observations Lab, Institute for Space and Terrestrial Science; we offer it as a model from which further development is encouraged. Changes from our configuration are premised on the current state of software development and individual requirements of the applications research. We currently use two Mac II's for image processing. Although multispectral

[1]Macintosh II, Macintosh IIx and Macintosh IIcx are trademarks of Apple Computer.

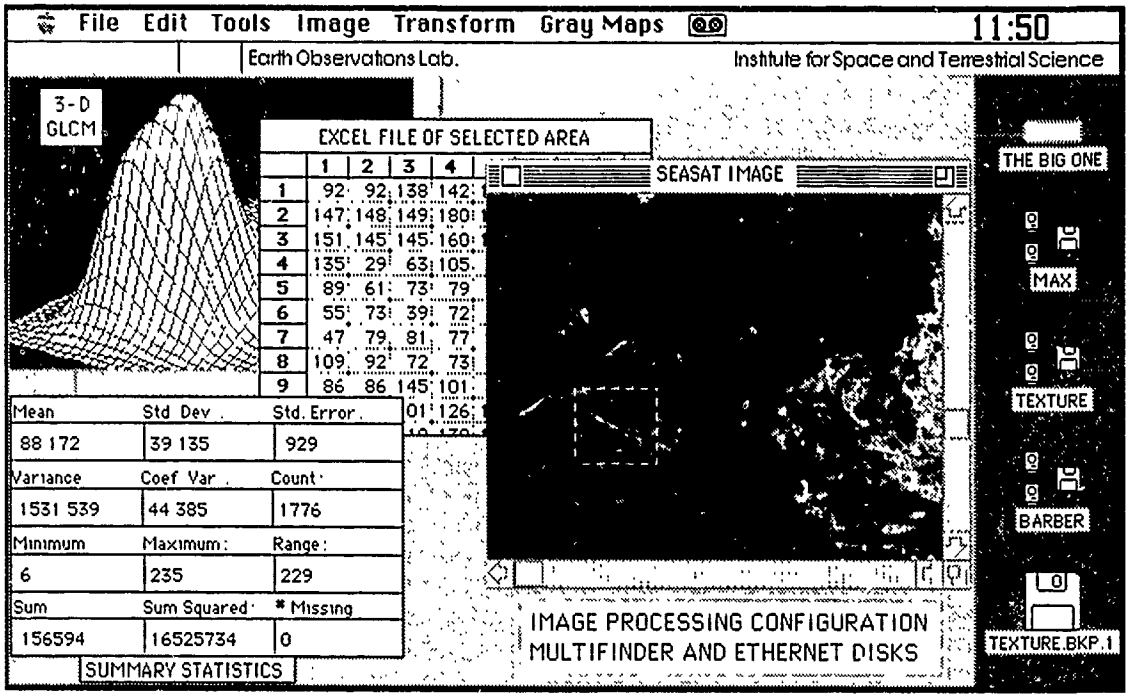


Figure 1. Illustration of Mac II image processing. Four programs are shown running under Multifinder.

processing is possible, our applications have been limited to single band images because of the limited capacity of certain software programs to handle large files.

Our image processing software consists of an image display/manipulation program (Digital Darkroom[1]), numerical analysis programs (Excel[2] and Mathmatica[3]), statistical analysis programs (Systat[4] and Statview 512+[5]); and a graphics/statistical analysis program (Data Desk[6]). In a typical analysis session the desired programs would be assigned a portion of the available RAM and are cooperatively multitasked using Multifinder[7]. A prerequisite to manipulation of data in these different software packages is format interchange software Most graphical, statistical and numerical programs accept ASCII format, so this was used as an interchange standard To illustrate use of the Mac II for image processing we provide the following example.

SAR Sea Ice Analysis

In this application we used the Mac II to create, test and evaluate Grey Level Co-occurrence Matrices (GLCM) for discrimination of sea ice types. These matrices are a means of obtaining textural information from different image features, and are typical of many spatial analysis approaches to feature discrimination. A GLCM is calculated at an interpixel spacing variable (δ) and orientation variable (α) over the entire homogeneous texture class. For each pixel in the class, at grey level i, the algorithm records the grey level for pixel j which is oriented at δ and α from i. The frequency of occurrence of these pixel pairs at grey levels (i,j) are normalized to reflect the probability distribution for the co-occurrence of all pixel pairs at grey levels i and j for each set of variables δ and α. A variety of statistics, derived from the GLCM, provide point estimates of image texture which are then compared between different features (Haralick, 1986). The utility of the Mac II comes in rapid turn-around between the idea, implementation, testing, and reporting. The speed of this integrated system and ease of use will be apparent to those familiar with the Macintosh interface or those who see this system in operation.

We wish to interpret a SEASAT SAR image of the Beaufort Sea (CRAGTAG[1] #RD2825), consisting of 978 lines

[1]Digital Darkroom is a registered trademark of Silicon Beach Software, Inc.
[2]Excel is a trademark of Microsoft Corp.
[3]Mathematical is a trademark of Wolfram Research Inc.
[4]Systat is a trademark of Systat Inc.
[5]Statview 512+ is a trademark of BrainPower, Inc.
[6]Data Desk is a trademark of Odesa Corp.
[7]Multifinder is a trademark of Apple Computer

[1]Canadian Radar AGc/Type Algorithm Group

by 964 pixels (Fig 1) for classification of ice type. Each ice class of interest is identified from the ground confirmation data. We intend to describe the differences between these classes using point estimates from various texture statistics.. The data from a homogeneous texture classes is selected in Digital Darkroom (Fig 1). The format interchange software allows us to move from the image to all other programs currently running under Multifinder. Summary statistics are calculated in Statview 512+, on the homogeneous texture class. In this same program we have access to many univariate and multivariate statistical analysis tools. The statistics program also has a useful graphics component which we use to conduct exploratory analysis, such as Principal Components Analysis or Multiple Discriminant Analysis (Barber *et al.* 1989).

In Excel a Grey Level Co-occurrence Matrix (GLCM) is calculated for the ice class. Development of the algorithms to produce the GLCM, and the statistics derived from each GLCM are programmed using a Macro language built into Excel. The Macro language is a high level programming environment which is interpreted rather than compiled. It is easy to implement analysis routines, but the processing speed is slower than compiled programs. We recommend using Macros for development and testing of analysis ideas, not for production oriented image processing.

Once the GLCM calculations are complete the results can be displayed graphically or numerically. We show an example of a three dimensional surface plot of the GLCM (Fig 1). This plot can be rotated and provides insights into the differences of each ice class in texture (or GLCM) space. Graphical displays are particularly useful when evaluating the outcome of a processing routine, or comparing the statistical characteristics of different image features. The other useful method of displaying results, is to convert the numerical data back into an image display format (part of the format conversion software) to view the results as an image.

Many ideas will not evolve if implementation is time consuming or tedious. The most significant contribution of the Mac II, in remote sensing applications research, is not the variety of operations that can be performed, but rather its ease of use. This efficiency promotes exploration, thereby challenging us to test our ideas in a structured analysis environment.

SUGGESTED CONFIGURATIONS

Stand alone

We recommend the following configuration if the MAC II is to be used as a stand alone system for remote sensing applications research:

- A minimum of four megabytes (MB) of RAM.
- A minimum of 40 MB of hard disk storage (80 MB is highly recommended for researchers working with large images).
- A backup system capable of fast backup and retrieve.
- A minimum 8 bit video resolution and a minimum 13 inch display.
- Format conversion programs are necessary to incorporate data from standard digital image formats (eg. LGSOWG[1]).

Networking

At the Earth Observations Lab we use a variety of computer systems for remote sensing and geographic information system research (Fig 2). To effectively utilize the capabilities of the Macintosh we have connected them to a campus-wide ethernet network. The speed and flexibility of the ethernet allows us to transfer an entire Landsat Thematic Mapper quadrant (7 bands) from one machine to another in approximately one minute. We use the tools available on the Macintosh to implement and evaluate analysis procedures. If results warrant, the analysis algorithms are ported to our SUN or MicroVAXes for implementation on a more production-oriented system.

The sharing of resources between systems is equally as important as the data transfer. For example, since the disks on the Mac II's are relatively small, we partition a section of the large disks on the MicroVAX-II/GPX for access by the Macintosh across the network. From the Mac II side, this partition looks like a locally mounted hard disk (Fig 1), from which the Macintosh can read files directly. The major obstacle which must be overcome for truly effective integration is the way in which the data are formatted on each machine. Even on the same machine, different programs require the data to be specified in unique formats. Until truly universal standards are developed and adhered to by software designers, there will always be a need for data conversion. The transfer of data between systems thus becomes a two-step operation: copy the data across the network, and translate it into an appropriate format. We are currently developing a 'wagon wheel' file server where the hub is a common file format and each spoke represents a translation capability to and from other file organizations.

[1]Landsat Ground Station Operators Working Group.

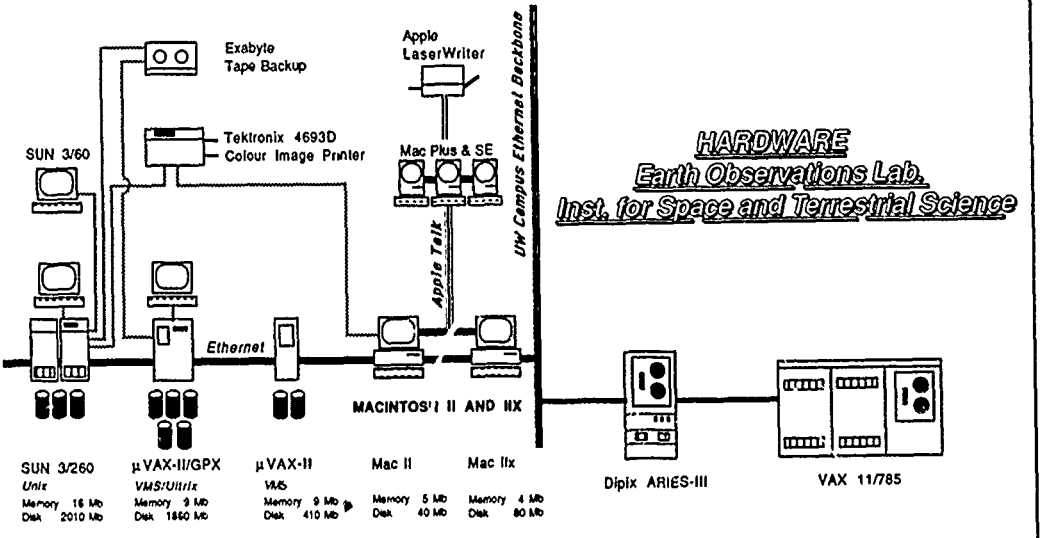


Figure 2. Networking at the Earth Observations Lab, Department of Geography, University of Waterloo

In either a stand alone or network configuration a variety of peripherals are available. Monochrome and color video digitizers allow importing of video images into the computer at video frame rates (1/30 sec.). The ability to digitize video imagery has benefits which range from quantitative analysis of airborne multispectral or thermal video (Barber *et al.* 1989), to digitizing textural or tonal images as standards on which to evaluate image analysis algorithms. Output to hardcopy devices is important for working copies of image analysis, presentations, and publication in technical and journal reports. Color, black and white, and slide presentation output devices provide hardcopy from draft to publication quality, in a surprising variety of formats and price ranges.

CONCLUSIONS

The Macintosh II, IIx and IIcx are the first members of the Macintosh family to allow image processing of remotely sensed images. The friendly interface and large base of commercial software make it possible to make simultaneous use of a variety of programs under Multifinder, to conduct image analysis. It should be noted that the Mac II in this configuration is intended for research purposes. We do not intend the Macintosh to be a production facility where computer intensive analysis is conducted. The strength of this system is in allowing the applications research scientist to formulate, initiate, test and evaluate their ideas with a minimum investment of time.

At an entry level hardware cost of $10,000, and a typical software applications costs of between $300 and $500, the Mac II represents a new research tool which brings image analysis to a larger and more diverse group of research scientists. It is the willingness to explore and accessibility to more scientists which makes the Mac II an important addition to the growing number of tools available to our profession.

ACKNOWLEDGEMENTS.

This research was supported by a Centre of Excellence grant from the Province of Ontario to the Institute for Space and Terrestrial Science, and an NSERC Operating Grant to E. LeDrew.

LITERATURE CITED

1. Barber, D.G. and E.F. LeDrew. Multivariate Analysis of Texture Statistics for SAR Sea Ice Discrimination. IGARS'89, July 10-14, Vancouver, BC. Canada.
2. Barber, D.G., P.R. Richard, and K.P. Hochheim. Thermal Remote Sensing for Walrus Population Assessment in the Canadian Arctic. IGARS'89, July 10-14, Vancouver, BC. Canada.
3. Haralick, R.M.. Statistical Image Texture Analysis. Handbook of Pattern Recognition and Image Processing. Ch. 11. Academic Press. 1986.

Modeling the Effects of Rain on Communications and Remote Sensing Systems

Robert K. Crane, Thayer School of Engineering
Dartmouth College, Hanover NH 03755, USA

Rain on a communication path may disrupt its operation. Rain may be the object to be sensed by a remote sensing system or may disrupt the operation of the system. In either case, the effects of the rain must be estimated statistically to provide information for system design. Two aspects of the rain problem are considered, the estimation of the statistics of the intensity of rain at a single geographical location (ie a point on a communication path) and the estimation of the joint statistics of rain at two geographical locations separated by up to 100 km (needed to estimate the spatial extent of the rain along a communications path or within the resolution volume of a remote sensing system).

Two approaches have been used for the estimation of point rain rate statistics, one employing rain climate regions and the other using local meteorological data to construct rain rate distribution estimates. Both approaches are empirical. For the former, empirical rain rate distributions are pooled from the limited number of observations made within a climate region and for the latter a number of years of observations at a single location are needed to provide a statistically stable estimate of the rain rate distribution. In practice, long time interval data sets are not available and either the data have been pooled from a climate region or large errors in the distribution estimates have been tolerated (often unknowingly). From the available rain rate distribution observations, it is evident that for mid-latitude locations, the use of rain climate regions (pooled data) is preferable to local data if the measurements at a location span a period of less than 2 years.

The estimation of the joint distribution for the simultaneous occurrence of rain at two spatially separated locations is difficult because the rain rate process is nonstationary for short time or spatial scales. Analyses of weather radar observations show that the spatial fluctuations in rain rate can be modeled by a two-dimensional turbulent process with an outer scale of the order of 250 km. Working with a spatial domain larger than the outer scale, a spatial correlation correlation function can be uniquely defined for the rain process. The correlation distance (1/e) for the rain process is 15 km. The spatial correlation function does not change from day-to-day or location-to-location. The temporal correlation function for the rain process is not defined for time scales corresponding to the time required for the rain process to move by an amount equal to the spatial correlation distance.

ELECTROMAGNETIC MODEL FOR PROPAGATION THROUGH CLOUDS

S.S.Şeker
The Department of Electrical and
Electronic Engineering
Boğaziçi University, İstanbul, TURKEY

ABSTRACT- Electromagnetic propagation through a sparse distribution of lossy dielectric particles in a cloud is investigated. Mathematical model is developed to aid in the interpretation of the interactions data obtained by electromagnetic remote probing of mixed ice crystals and waterdrop clouds. Such clouds can contain many possible crystals forms, most notably thin long cylinders, bullets and flat plate crystals.Bistatic reflectivity and attenuation are computed ,for waves of selected polarizations passing through clouds with specified size, shape, and distributions.

Keywords:Radiowave propagation, Stochastic cloud model, Multiple scattering.

1.INTRODUCTION

The increasing demand for communications is causing frequency-spectrum congestion. A problem in using short microwave or millimeter-wave bands for communications purposes exists in that strong interactions can occur between such radio waves and atmospheric hydrometeors (e.g. rain, snow and ice crystals), that significantly degrade the communications quality at frequencies above about 5 GHz.

Because of the small loss tangent of ice at the frequency range of 30-50 GHz, significant depolarization of satellite-ground communication signal in this frequency range could occur when cirrus cloud is present in its path.High-altitude cirrus cloud consists basicly of ice needles and plates(1). It is therefore necessary to establish a model capable of predicting the behaviour these systems. Because of the complexity and random behaviour of the different factors involved in the physical process or clouds, it is quite difficult to find a model which is simple and at the same time sufficently precise.

The widening interest in the scattering problem for applications to cross-polarization corrections to dual polarized links, as well as other applications, indicates a need for a more indepth understanding of the theory behind the conventional approach. This paper presents a systematic formulation of the problem of scattering from a sparse distribution of arbitrary shape in a random medium such as in clouds.The cloud model as we describe represents an initial step in modeling real mixed phase clouds. The cloud is taken to be a mixture of scatterers. Thin circular cylinders and disks represent either ice crystals or snowflakes respectively. Actual ice crystals and snowflakes have a huge variety of possible shapes. The circular disks and cylinders represent an average over these shapes, and averaged over all orientations.

2.PROBLEM FORMULATION

We consider the problem of scattering of time harmonic electromagnetic wave from N discrete lossy dielectric scatterers which have random position and orientation in a volume V. The particles have a volume V_p, a relative dielectric, constant ε_r and density ρ. The surronding medium is considered to be free space as in (2). In this way, one can obtain a mean field equation from Maxwell's equations, assuming the incident field on each scatterer is the mean field (Foldy approximation) and sparse distribution of particle which indicate fractional volume ($\delta=\rho V_p$)is very small. In the derivation of this result the assumption was made that the plane wave is scattered only once (single scattering) by the particles. This assumption is unrealistic, since the wave may be scattered several times by the particles before reaching receiver. When the scattering coefficients of the particles and the particle density are small, contributions to the scattered wave by second, third, and higher-order scattering may be ignored. There are cases, however, where the scattering coefficients of the scatterer are large and the multiple scattering contributions cannot be ignored.

The resulting approximate dyadic mean field and effective permittivity are given by (2).

$$\underline{E}(\underline{x})=e^{i\underline{k}\cdot\underline{x}}e^{\gamma(\bar{\underline{x}})}\bar{\underline{f}}(\underline{i},\underline{i}) \qquad (1)$$

with

$$\gamma(\bar{\underline{x}})=\frac{2\pi i}{k_o}\int_{\xi_{in}}^{\xi_{out}}\rho(\underline{x})d\xi \qquad \bar{\underline{x}}\in V \qquad (2)$$

and

$$\underline{\underline{\varepsilon}}(f)=\underline{\underline{I}}+\frac{4\pi}{k_o^2}\rho\,\bar{\underline{\underline{f}}}(\underline{i},\underline{i}) \qquad (3)$$

where $k_o=\omega\sqrt{\mu_o\varepsilon_o}$ is the free-space propagation constant, $\bar{\underline{f}}(\underline{i},\underline{i})$ is the average of the forward dyadic scattering amplitude of the hydrometeors with given shape and dielectric constant. The average is taken over appropriate distribution variables such as size and canting angle.

In order to write the components of $\underline{E}$, we first examine (2). The density is a slow varying function of

position and hence we take it as approximately constant and take it out of the integral in (2). Now $\int_{\xi in}^{\xi out} \nu(\underline{x})d\xi$ is simply the effective path length travelled by the wave inside the cloud. We call this pathlength L. Also $\underline{i}.\underline{x}$ is equal to the path length travaled, L. The components $E_{\alpha\beta}$ may be written as

$$E_{\alpha\beta}=e^{iK_{\alpha\beta}L}=e^{i(\frac{2\pi\rho}{k_o}\bar{f}_{\alpha\beta}+k_o)L} \quad \alpha,\beta\varepsilon(h,v) \qquad (4)$$

Here $K_{\alpha\beta}$ is the effective propagation coefficient and is given by

$$K_{\alpha\beta}=k_o+\frac{2\pi\rho}{k_o}\bar{f}_{\alpha\beta} \qquad \alpha,\beta\varepsilon(h,v) \qquad (5)$$

In a region containing a population of aligned ice crystals

$$K_{v,h}=k_o+\frac{2\pi\rho}{k_o}\int_o^\infty f_{v,h}\, n(a)da \qquad (6)$$

Here the forward scattering amplitudes f_v and f_h depend on the incident angle, and on the effective particle radius a; n(a) da is the number of particles per cubic meter with radii between a and a+da. The difference between K_v and K_h varies with incidence angle in the same manner as f_v-f_h.

The attenuation, and phase rotation are given as

$$A_{v,h}=8.686\ \mathrm{Im}\ (K_{v,h})\times 10^3 \ \mathrm{dB/km} \qquad (7)$$

$$\phi_{v,h}=\frac{180}{\pi}\ \mathrm{Re}\ (K_{v,h})\times 10^3 \ \mathrm{deg/km} \qquad (8)$$

In this model it is assumed that the drop axes align with the polarization directions, so no cross polarization in this coordinate system. Thus, the propagation coefficients are given in the preferred directions where no coupling exists.

In the physical problem, the drops will not in general align this way. Actually, it is the representative drop in its representative orientation we are dealing with, and in general, this drop will not aligh with the polarization directions. We may rotate to the preferred direction where cross terms are not zero.

The general form of the propagation coefficient in the preferred frame

$$K_{vv}=\frac{2\pi\rho}{k_o}\lambda^+ + k_o \qquad (9)$$

$$K_{hh}=\frac{2\pi\rho}{k_o}\bar{\lambda}+k_o \qquad (10)$$

where

$$\lambda^-=\left|\bar{f}_{hh}+\bar{f}_{vv}\pm\sqrt{(\bar{f}_{hh}-\bar{f}_{vv})^2+4\bar{f}_{hv}\bar{f}_{vh}}\right|/2 \qquad (11)$$

Using these results we can make the final analogy to the conventional approach, relating the transmitted fields to the incident fields.

$$\begin{bmatrix}\underline{E}^h\\ \underline{E}^v\end{bmatrix}=\begin{bmatrix}a_{hh} & a_{hv}\\ a_{vh} & a_{vv}\end{bmatrix}\begin{bmatrix}e^{i\underline{k}.\underline{x}}\,\underline{h}^o\\ e^{i\underline{k}.\underline{x}}\,\underline{v}^o\end{bmatrix} \qquad (12)$$

This is the same equation as in Brussard (3). It relates the transmitted fields to the incident fields given by the last term on the right of equation (12).

New microwave or millimeter communication systems are designed to use dual-polarization channels in order to increase communication capacity without increasing bandwidth. It is possible that the interfering crosstalk between the two chanels through depolarization caused by hydrometeors can severelly restrict the utility of these "frequency reuse" systems. The depolarization effect has thus attracted the concern of recent investigators, and the extensive theoretical studies have made. The degree of depolarization may be represented by the ratio of cross-polarized to copolarized signals at the receiver. We may define two kinds of cross-polarization factors. They are the cross-polarization discrimination (XPD) and the cross-polarization isolation (XPI). We can obtain these factors from out theory, for example XPD is, (3)

$$XPD_\alpha=\frac{E_{\alpha\beta}}{E_{\alpha\alpha}} \qquad (13)$$

Where α(h,v) and β is the other polarization directions Thus, we have for h polarization

$$XPD_h=\frac{1-e^{(\lambda^- -\lambda^+)\gamma}}{1-\frac{\lambda^+-\bar{f}_{hh}}{\lambda^- -\bar{f}_{hh}}e^{(\lambda^- -\lambda^+)\gamma}} \qquad (14)$$

It is required to determine the bistatic cross section of disk and cylinder Following the conventional defination, σ_{bpq} is defined in terms of the scattering amplitude with polarisation in the p direction due to an incident wave with polarization q as follows:

$$\sigma_{bpq}=4\pi\left|\underline{p}\cdot\underline{f}\,(\underline{Q},\underline{i},q)\underline{q}\right|^2 \qquad (15)$$

where $\underline{f}\,(\underline{Q},\underline{i},q)$ is given by refs. (4) and (5).

The total scattering by the assemblage is defined in terms of the reflectivity, Z_{pq}, given by

$$Z_{pq}=\int \sigma_{bpq}(a_0)\, N(a_0)\, da \qquad (16)$$

where $\sigma_{bpq}(a_0)$ is the bistatic cross section of a single hydrometeor of equivolume radius a_0, and $N(a_0)$ is the size distribution function of particle.

3.NUMERICAL CALCULATIONS

The scattering problems discussed have been analyzed numerically in this section. Before doing this, however, it will be necessary to have the dielectric constant, for the hydrometeors.

Ray (6) carefully examined the existing measurements made by many investigators at various frequency and temperature ranges and obtained an empirical model of the complex refractive index. According to Ray's model the real part of N ($N=\sqrt{\varepsilon}$) takes a nearly constant value of 1.78 independently of both frequency and temperature, and the imaginary part (the loss) takes an extremely small value as 0.036.

Snow particles are the complicated mixtures of ice with air, water, or both. The mixing rate and the shapes of the constituents may vary considerably depending on external meteorological conditions to which snow particles are exposed. The dielectric constants obtained by the above work have been used in the following calculation.

In the numerical work, two different media were considered The first medium consists solely of disk-shaped ice particles with radii ranging from 100 μm to 3.6mm, with a constant thickness/diameter ratio=0.1. The second medium consists solely of cylinder-shaped ice particles with radii ranging from 10 μm to 360 μm with a constant length/diameter ratio= 10. The particles are assumed to be aligned and lying in the horizontal plane (7).

Two different size distribution functions were assumed for each medium model (3,8). The first size distribution is a negative exponential size distribution, and the second size distribution is uniform. In both cases, the particle concentration is constant and 10^4 particles/m^3 . A total of four models of scattering media were studied They are;

Disk ensemble: 1) a negative exponential size distribution , 2) a uniform size distribution. Cylinder ensemble: 3) a negative exponential size distribution 4) a uniform size distribution.

The exponential size distribution is obtained by using 10 pairs of values The radius values form a geometric sequence with the first term equal to the smallest diameter and the 10th term aqual to the largest diameter in the distribution. The concentration for each of these radius values is 1000 m-3. The integrations in eqs. (6) and (16) are approximated by calculating the scattering amplitudes for these radius values, weghting them by corresponding concentrations and summing them over all ten radius values. For the cylinder ensemble, the smallest radius is taken to be 10 μm and the largest is 320 μm. These values correspond to the radius distribution

$$n_{ce}(a_0) = (1.21\times10^8)\ \exp(6.93\times10^4.a_0)\,m^{-3}.m^{-1} \qquad (17)$$

where the subscript ce stands for "cylinder, exponential".

For the disk ensemble, the following exponential radius distribution was used

$$n_{de}(a_0) = (1.21\times10^8)\ \exp(6.93\times10^3.a_0)\ m^{-3}.m^{-1} \qquad (18)$$

The uniform distribution used for the cylinder ensemble is

$$n_{cu}(a_0) = \begin{cases} 2.86\times10^6\ m^{-3}.\ m^{-1} & \text{for } 10^{-5}m < a_0 < 3.6\times10^{-4}m \\ 0 & \text{otherwise} \end{cases} \qquad (19)$$

and for the disk, it is

$$n_{du}(a_0) = \begin{cases} 2.86\times10^6\ m^{-3}.m^{-1} & \text{for } 10^{-4}m < a_0 < 3.6\times10^{-3}m \\ 0 & \text{otherwise} \end{cases} \qquad (20)$$

The bistatic reflectivities obtained for the disk ensemble are shown in Figure 1. The reflectivities Z_{hh} for horizontal polarization show a periodic variation with maxima in the forward and backward directions. The reflectivity for vertical polarization Z_{vv} is fairly constant for exponentially distributed radii. For uniformly distributed radii, a maximum in the forward direction and a minimum in the backward direction exist in Z_{vv} . The Z_{vv} values are larger than Z_{hh} values for the disk ensemble.

For a cylinder ensemble, the reflectivities for horizontal polarization show a peridodic variation similar to that of the disk ensemble. The vertical reflectivities are fairly constant, but they are smaller on the average than horizontal reflectivities Figure 2 shows the reflectivities for the cylinder ensemble.

The reflectivity calculations were made in the Rayleigh region. The results were qualitatively compared with published results. The following comments can be made:

1. The bistatic reflectivities of a medium containing aligned disk or cylinder scatterers show a similar angular variation pattern to those of a rain medium in the Rayleigh regime calcuiated by Charlton, Holt and Evans (9).

2.The reflectivity for vertical polarization remains fairly constant with scattering angle for both the cylinder and disk scatterers, while the reflectivity for horizontal polarization has peaks in the forward and backward scattering directions.

3.The reflectivity of a disk ensemble for vertical polarization is roughly six times larger than that for horizontal polarization wave. In the case of a cylinder ensemble the reflectivity for horizontal polarization has a maximum value about three times that for vertical polarization. Such a large difference in maximum of reflectivities for horizontal and vertical polarization are not observed in rain scatter.

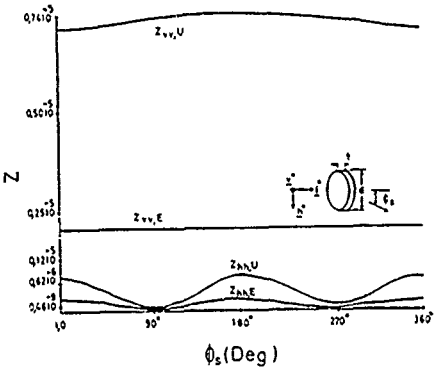


Figure 1: Bistatic Reflectivity Patterns for a Disk ensemble f=2.81 GHz, $\rho=10^4$ m^{-3} , 200μm<d<7.2 mm t/d=0.1 E= exponentially, U: Uniformly distributed radii

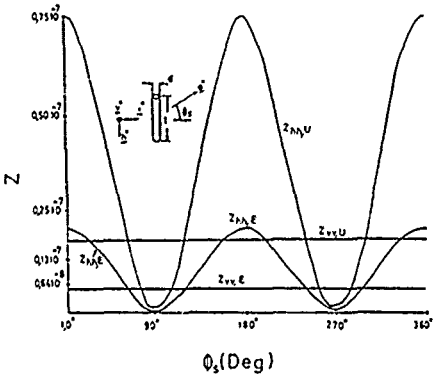


Figure 2: Bistatic Reflectivity Patterns for Cylinder ensemble f=2.81 GHz , $\rho=10^4$ m^{-3} , 20μ m<d< 720 μm, ℓ/d=10 E: exponentially, U: Uniformly distrubuted radii

The attenuation is also polarization and orientation dependent and for normal incidence, the attenuation of a disk ensemble with exponential size distribution is calculated to be 0.8 dB/km at 10GHz, which is approximately equivalent to the attenuation of a 30 mm/h rain.

In the second case, attenuation of the four model media were computed while the frequency was changed from 30 to 300 GHz. The same incidence and orientation angles were used as in the reflectivity calculations. The attenuation for the disk ensemble is shown in Figure 3. For the cylinder ensemble, the calculated attenuation curves are shown in Figure 4. The attenuation A_{hh} for horizontal polarization is larger than the attenuation A_{vv} for vectical polarization. This can be explained by the cylinders orientation being in the horizontal plane. The difference between the reflectivity values of the cylinder ensemble can be explained similarly.

Frequency characteristics of rain attenuation for various rain rates increases monotonously up to frequencies of roughly 100 GHz. Then it begins to decrease (1). A similar effect is occured in the attenuation by disk and cylinder ensembles. In the resonance region, the monotonocity of the frequency characteristics of attenuation may be disturbed. The oscillatory behavior evident at the higher frequencies in attenuation are shown in greater detail in figures 3 and 4 It is apperent that above 16 GHz for fig 3 , 60 GHz for fig 4 resonance play a major role in models and essentially precludes the utilization of simple models Similar oscillatory behaviour can be observed for bistatic reflectivity of disk and cylinder.

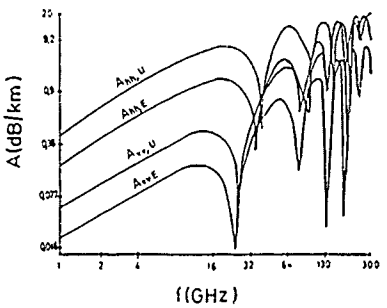


Figure 3 : Attenuation of the Disk ensemble $\rho=10^3$ m^{-3}
20 μm<d<20mm t/d=0.1 , E= exponentially,
U : Uniformly distributed radii

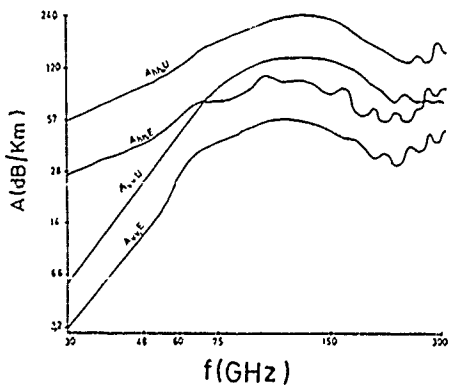


Figure 4: Attenuation of the Cylinder ensemble $\rho=10^3 m^{-3}$
20 μm <d<2 mm l/d =10.E=Exponentially,
U: Uniformly distributed radii

4.CONCLUSIONS

This work has attempted to give a new formulation of Electromagnetic propagation through a clouds. The areas of application of this model now corver a wide range from space communications to weather forecating.

The formulation as presented here offers several features for calculation of electromagnetic wave propagation through precipitation media. On the basis of our extensive calculations, we may summarize our observations as.

1. The formulation is matrix and stochastic in nature and easily accommodates arbitrary polarization states. The formulation is valid in all frequency regions when the criterions are met.
2. Complete characterization of medium depolarization effects from hydrometeors (e.g., attenuation, isolation, and phase shift).
3.Scattering particle distributions of particle size, shape, and orientation angle are directly included into the model and varying medium density along the propagation path can be accommodated.

It is hoped that the methods given in this work may provide a basis for improve interpretation of measured results from present cloud probing systems or to guide the design and deployment of future systems.

Acknowledgments: The author would like to thank Mustafa Kocatürk far carrying out the numerical calculations.
The author is also pleased to acknowledge the substantive contributions to the development of this model by Professor R.H. Lang of the George Washington Universtiy Wash.D.C.,and,Dr. D.M. LeVine, NASA,Maryland, U.S.A.

5.REFERENCES

1.Holt, A.R., Uzunoğlu, N.K., and Evans, B.C., "An integral equation solution to the scattering of electromagnetic radiation by dielectric spheroids and ellipsoids", IEEE Trans, AP-26, pp.706-712, 1978.

2.Lang, R.H. , Şeker, Ş.S. and LeVine, D.M., "Vector solution for the mean electromagnetic fields in a layer of random particles", Radio Science, Vol.21(5), pp.771-786, Sep.-Oct. 1986.

3.Brussard, G., "A Meteorological model for rain induced cross polarization", IEEE Trans A P, Vol.AP-24,(1), pp. 5-11, 1976.

4.LeVine, D.M., Meneghini, R.,Lang, H., and Şeker Ş.S., "Scattering from arbitrarily oriented dielectric disks in the physical optics regime", J.Opt. Soc Am., 73 (10), pp. 1255-1262, 1983.

5.Şeker, Ş.S. and Schneider, A., "Electromagnetic scattering from a dielectric cylinder of finite length", IEEE Trans.AP 36(2). pp.303-307 Feb. 1988.

6.Ray, P.S.,"Broadband Complex refractive indices of ice and water", Applied Optics, Vol. 11 (8), pp. 1836-44, August 1972.

7.Sekhon, R.S. and Srivastava, R.C., "Snow size spectra and radar reflectivity", J.Atmos. Sci. Vol.27. pp.299-307, March 1970.

8.Gunn, K.L.S and Marshall, J.S., "The distribution with size of aggregate snowflakes", J.Meteorol, Vol.15, pp. 452-461 , Oct 1958.

9.Charlton, D.G.,Holt,A.R and Evans, B.G., IEE Proc.Vol. 130 (4) Pt.F,pp.317-324, 1983.

Phase Matrices of Radiative Transfer Equation For Strongly Fluctuated Continuous Random Media

Ya-Qiu Jin

Department of Electronic Engineering, Fudan University
Shanghai, China

Abstract

Employing the strong fluctuation theory, the phase matrix and scattering coefficient of the radiative transfer equation for strongly fluctuated continuous random media are obtained. Comparisions with the conventional theory for weak-fluctuated media are discussed. By using the Gaussian quadrature and the eigenanalysis approaches, the vector thermal radiative transfer equation for a layer of random medium is solved, and is favorably matched with the experimental data.

1. Introduction

The radiative transfer theory (RT) has been extensively applied to the studies of multiple scattering and transmission of specific intensity in random media. Random media might fall into three kinds: random discrete scatterers, continuous random media, and randomlyrough surface. Discrete scatterer media may be treated as tenuously or densely distributed spherical Rayleigh, or Mie particles, nospherical particles (spheroid, disk, cylinder, etc.), or others. However, continuous random media are not as directly perceivable as random discrete scatterers. It is characterized by the covariance and correlation function of random fluctuation of the medium permittivity. The RT equation and its solution can be attributed to Refs.1 and 2. Reviewing the derivations of the RT equation, it can be seen that the validity criterion is restrictive to the weak fluctuation and small covariance. The approach of weak fluctuation will lead nonphysically negative effective propagation constant as the covariance takes a large value. Therefore, the conventional RT equation is only applicable to the weakly-fluctuated random media. Most of the geophysical media are often treated as strongly-fluctuated, noisotropic continuous random media. We have to reconsider the basis of the conventional RT equation, and enhance the applicability of the RT equation to strongly-fluctuated media. In the strong fluctuation theory (ref.3,4), the singularity of the mean dyadic Green's function and correlation of fluctuation have been properly taken into account. The effective permittivity of random medium is determined by the bilocally-approximated Dyson equation of the mean field. The criterion of small covariance can be eliminated. However, it's difficult so far to take account of all high-order multiple scattering in numerical calculation of scattered intensity, by employing the wave analytic theory.

In this paper, by using the strong fluctuation theory, the phase matrix and scattering coefficient in the RT equation are derived. A newly-introduced permittivity, as the zeroth order approximation of the effective permittivity, and the coefficient of the delta function for the singularity of the mean dyadic Green's function in nonisotropic random media are reformulated, and calculated. Then, we have a RT equation, which has new constituents, and is applicable to multiple scatteing, transmission, and thermal radiation in strongly-fluctuated random media. Comparisions with the conventional RT for weak fluctuation are discussed. By using Gaussian quadrature and the eigenanalysis approaches (ref.5), the vector thermal RT equation is solved. Numerical example is used to favorably match the experimental data of snowfield. The relationships between the brightness temperature and parameters (wavelength, polarization, observation angle, correlation length, layer depth, and composition of the medium, etc.) can be obtained.

2. The Phase Matrix of the RT equation

Consider a EM wave incident on a continuous random medium with the permittivity $\varepsilon(\bar{r})=\varepsilon_m+\varepsilon_f(\bar{r})$, where ε_m is the mean permittivity $\langle\varepsilon(\bar{r})\rangle$ and $\varepsilon_f(\bar{r})$ is the random fluctuation, and $\langle\varepsilon_f(\bar{r})\rangle=0$. The wave equation is written as

$$\nabla\times\nabla\times\bar{E}(\bar{r})-\frac{k_0^2}{\varepsilon_0}\varepsilon_m\bar{E}(\bar{r})=\frac{k_0^2}{\varepsilon_0}\varepsilon_f(\bar{r})\bar{E}(\bar{r}) \quad . \tag{1}$$

Taking a small volume V_1 with arbitrary shape in the random medium, the scattering from the scatterer V_1 to the point $\bar{r}$ outside V_1 is

$$\bar{E}_s(\bar{r})=\int_{V_1} d\bar{r}_1\bar{\bar{G}}_{01m}(\bar{r},\bar{r}_1)\cdot\frac{k_0^2}{\varepsilon_0}\varepsilon_f(\bar{r}_1)\bar{E}(\bar{r}_1), \tag{2}$$

where $\bar{\bar{G}}_{01m}$ is the mean dyadic Green's function. Assuming $\varepsilon_f\ll 1$, and the covariance $D_\varepsilon\ll 1$,

the Born approximation $\bar{E}(\bar{r}_1)$=the incident field $\bar{E}_0(\bar{r}_1)$. The small D_ε must be imposed, since large D_ε will lead nonphysical effective permittivity (ref.3). Then, the second moment of scattering field $\langle|\bar{E}_s(\bar{r})|^2\rangle$ is used to derive the phase matrix. However, for most of multi-phase random media composed by different constituents (such as snow, or some geophysical media), the fluctuation ε_f in any point and the covariance D_ε are often not much less than one. The scattered field from the source $k_0^2\,\varepsilon(\bar{r})\bar{E}(\bar{r})/\varepsilon_0$ in the homogeneous medium ε_m should be modified.

In the strong fluctuation theory (ref.3 4), a diagonal permittivity $\bar{\bar{\varepsilon}}_g$ is introduced, and is to be determined by the correlation function of fluctuation. Wave scattering and propagation in random media has an effective propagation constant. The effective permittivity takes account of the randomness and scattering effects in the medium, and is not equal to $\varepsilon_m\bar{\bar{I}}$. If high-order scattering is neglected, $\bar{\bar{\varepsilon}}_{eff}$ will approaches to $\bar{\bar{\varepsilon}}_g$. And $\bar{\bar{\varepsilon}}_g$ approaches to $\varepsilon_m\bar{\bar{I}}$ only if the fluctuation is weak. Thus, $\bar{\bar{\varepsilon}}_g$ is the zeroth-order approximation of the effective permittivity. From the strong fluctuation, the total field in the volume V_1 is

$$\bar{F}(\bar{r}_1)=\bar{E}_0(\bar{r}_1)+k_0^2\int_{V_1}d\bar{r}_1'PV.\bar{\bar{G}}_g(\bar{r}_1,\bar{r}_1')\cdot\bar{\bar{Q}}(\bar{r}_1')\cdot\bar{F}(r_1'),$$

where (3)

$$\bar{\bar{Q}}(\bar{r})=[\bar{\bar{I}}\varepsilon_n(\bar{r})-\bar{\bar{\varepsilon}}_{gn}][\bar{\bar{I}}+\bar{\bar{S}}\cdot(\bar{\bar{I}}\varepsilon_n(\bar{r})-\bar{\bar{\varepsilon}}_{gn})]^{-1}\quad,\quad(4)$$

$$\bar{F}(\bar{r})=[\bar{\bar{I}}+\bar{\bar{S}}\cdot(\bar{\bar{I}}\varepsilon_n(\bar{r})-\bar{\bar{\varepsilon}}_{gn})]\cdot\bar{E}(\bar{r})\quad,\quad(5)$$

$$\bar{\bar{G}}_g(\bar{r},\bar{r}')=PV.\bar{\bar{G}}_g(\bar{r},\bar{r}')-(\bar{\bar{S}}/k_0^2)\,\delta(\bar{r}-\bar{r}')\quad.\quad(6)$$

Here, the coefficient $\bar{\bar{S}}$ of the priciple volume is also diagonal, and is determined by the shape of V_1. This volume should be an equal-correlation surface. Averaging (3) under the bilocal approximation, the Dyson equation of $\langle\bar{F}\rangle$ can be obtained. Then, the effective permittivity can be obtained, and the restrictive criterion $D_\varepsilon\ll1$ can be eliminated. The scattering field at $\bar{r}$ outside the scatterer source V_1 is

$$\bar{E}_s(\bar{r})=k_0^2\int_{V_1}d\bar{r}_1\bar{\bar{G}}_{01m}(\bar{r},\bar{r}_1)\cdot\bar{\bar{Q}}(\bar{r}_1)\cdot\bar{F}(\bar{r}_1)\quad.\quad(7)$$

Making use of the distorted Born approximation, the second moment of scattering field

$$\langle|\bar{E}_s(\bar{r})|^2\rangle=\langle k_0^2\int_{V_1}d\bar{r}_1\int_{V_1}d\bar{r}_1'\,[\bar{\bar{G}}_{01m}(\bar{r},\bar{r}_1)\cdot\bar{\bar{Q}}(\bar{r}_1)\cdot\bar{F}_m(\bar{r}_1))\,[\bar{\bar{G}}_{01m}(\bar{r},\bar{r}_1')\cdot\bar{\bar{Q}}(r_1')\cdot\bar{F}_m(\bar{r}_1')]^*\rangle \quad(8)$$

where the mean dyadic Green's function in the far field approximation is written as

$$\bar{\bar{G}}_{01m}(\bar{r},\bar{r}_1)=\exp(ik_0r)[\hat{v}_s\hat{v}_s\exp(-i\bar{k}_{vs}\bar{r}_1)+\hat{h}_s\hat{h}_s\exp(-i\bar{k}_{hs}\bar{r}_1)]/(4\pi r),\quad(9)$$

and the mean field in the zeroth-order approximation is

$$\bar{F}_m(\bar{r}_1)=E_{vi}\hat{v}_i\exp(i\bar{k}_{vi}\bar{r}_1)+E_{hi}\hat{h}_i\exp(i\bar{k}_{hi}\bar{r}_1).\quad(10)$$

To obtain the RT equation, we define the Stockes parameters I and the scattering matrix $\bar{\bar{S}}_I$ as

$$\bar{I}_s(\bar{r})=\bar{\bar{S}}_I(\hat{k}_s,\hat{k}_i)\cdot\bar{I}_i(\bar{r})\quad.\quad(11)$$

Following the definition of the Stockes parameters, we calculate the elements of the scattering matrix by substituting (9,10) into (8),

$$S_{vv}=(k_0^4/16\pi^2r^2)\{\sum_t(\hat{v}_s\hat{t})^2(\hat{t}\cdot\hat{v}_i)^2\int_{V_1}d\bar{r}_1\int_{V_1}d\bar{r}_1'\langle Q_t(\bar{r}_1)Q_t^*(\bar{r}_1')\rangle\exp[i(\bar{k}_{vi}-\bar{k}_{vs})\cdot(\bar{r}_1-\bar{r}_1')]$$
$$+\sum_t\sum_{t'}(\hat{v}_s\hat{t})(\hat{t}\cdot\hat{v}_i)(\hat{v}_s\hat{t}')(\hat{t}'\hat{v}_i)\int_{V_1}d\bar{r}_1\int_{V_1}d\bar{r}_1'\,2\langle Re(Q_t(\bar{r}_1)Q_{t'}^*(\bar{r}_1'))\rangle\exp[i(\bar{k}_{vi}-\bar{k}_{vs})\cdot(\bar{r}_1-\bar{r}_1')]\quad,\quad(12)$$

where $t,t'=x,y,z$, and $t'\neq t$. Rewriting the scattering and incident directions in terms of $(\theta,\theta';\phi,\phi')$, we obtain

$$\begin{aligned}S_{vv}=(&D_{11}\cos^2\theta\cos^2\phi\cos^2\theta'\cos^2\phi'\\&+D_{22}\cos^2\theta\sin^2\phi\cos^2\theta'\sin^2\phi'\\&+2D_{12}\cos^2\theta\sin\phi\cos\phi\cos\theta'\sin\phi'\cos\phi'\\&+2D_{13}\sin\theta\cos\theta\cos\phi\sin\theta'\cos\theta'\cos\phi'\\&+2D_{23}\sin\theta\cos\theta\sin\phi\sin\theta'\cos\theta'\sin\phi'\\&+D_{33}\sin^2\theta\sin^2\theta')W_{vv}\end{aligned}\quad,\quad(13)$$

where

$$W_{ab}=\int_{V_1}d\bar{r}_1\int_{V_1}d\bar{r}_1'\,\exp[i(\bar{k}_b'-\bar{k}_a)\cdot(\bar{r}_1-\bar{r}_1')]C(|\bar{r}_1-\bar{r}_1'|),\quad(14)$$

where a,b can take different polarization v, or h, for all elements. Other elements are obtained in the same way. C(.) is the normalized correlation function, and the covariance $D_{11}=\langle Q_xQ_x^*\rangle$, $D_{12}=Re\langle Q_xQ_y^*\rangle$, $D_{13}=Re\langle Q_xQ_z^*\rangle$, etc. are defined. For more explicit formulations, we consider herewith the transverse symetric case, i.e. the correlation lengths $L_x=L_y\equiv L_1$, and

$$\langle Q_t(\bar{r}_1)Q_{t'}^*(\bar{r}_1')\rangle=D_{tt'}\exp[-|z_1-z_1'|/L_z-(|x_1-x_1'|^2+|y_1-y_1'|^2)/L_1^2]\quad.\quad(15)$$

Then, substituting (15) into (14), we obtain

$$\begin{aligned}P_{11}(\theta,\theta')&=\int_0^{2\pi}d\phi r^2S_{vv}(\theta,\theta';\phi)/V_1\\&=M(\theta,\theta')\exp(-A)\{[D_{33}\sin^2\theta\sin^2\theta'\\&\quad+D_{11}\cos^2\theta\cos^2\theta'/2]I_0(A)\\&\quad+2D_{13}\sin\theta\cos\theta\sin\theta'\cos\theta'I_1(A)\\&\quad+D_{11}\cos^2\theta\cos^2\theta'I_2(A)/2\}\end{aligned}\quad,\quad(16a)$$

$$P_{12}(\theta,\theta')=D_{11}M(\theta,\theta')\exp(-A)\cos^2\theta[I_0(A)-I_2(A)]/2\quad,\quad(16b)$$

$$P_{21}(\theta,\theta')=D_{11}M(\theta,\theta')\exp(-A)\cos^2\theta'[I_0(A)-I_2(A)]/2\quad,\quad(16c)$$

$$P_{22}(\theta,\theta')=D_{11}M(\theta,\theta')\exp(-A)[I_0(A)+I_2(A)]/2\quad,\quad(16d)$$

where

$$M(\theta,\theta')=(k_g^4L_1^2/4)\{L_z/[1+k_g^2L_z^2(\cos\theta-\cos\theta')^2]\}\times\exp[-(k_g^2L_1^2/4)(\sin\theta-\sin\theta')^2]\quad,\quad(17a)$$

$$A=(k_g^2L_1^2/2)\sin\theta\sin\theta'\quad,\quad(17b)$$

and I_0, I_1, and I_2 denote, respectively, modified Bessel function of zeroth, first, and second order. The scattering coefficients are obtained as

$$K_{sv,h}(\theta)=\int_0^\pi d\theta'\sin\theta'[P_{11(12)}+P_{21(22)}]\quad.\quad(18)$$

It can be seen from (16a-d) that there are contributions from D_{11}, D_{33} and D_{13} in P_{11}, and there is only one covariance D_{11} in P_{12}, P_{21}, and P_{22}. Since there are components of the vertical polarization vector in both xy and z directions, and there is no component of the horizontal polarization vector in z direction, the differences between L_z and L_1 and D_{11}, D_{33} and D_{13} effect the polarization transition from $\hat{v}'$ to $\hat{v}$. Moreover, if $L_x \neq L_y$, all elements in $\bar{\bar{P}}(\theta,\theta')$ shall be related with D_{xx}, D_{yy}, D_{xy}, and L_x, L_y. It can be seen from (13,14). The dependence of angles in $\bar{\bar{P}}(\theta,\theta';\phi,\phi')$ is similar to that of discrete Rayleigh scatterer, since we have used the distorted Born approximation in low frequency limit. The phase matrix $\bar{\bar{P}}(\theta,\theta')$ as shown in (16), and the scattering coefficients of (18) are used in our RT equation for a layer of strongly fluctuated, nonisotropic continuous random medium. When ε_f, D_ε are very small, then $\bar{\bar{\varepsilon}}_g$ approaches to $\varepsilon_m \bar{\bar{I}}$, and $D_{11} \doteq D_{33} \doteq D_{13} \doteq D_\varepsilon$. Then our results reduce to the conventional RT.

3. Coefficient $\bar{\bar{S}}$ of the delta function

The coefficient $\bar{\bar{S}}$ in (6) depends on the shape of the excluded principle volume, and is related with correlation function and correlation lengths. From (3), we have

$$\langle \bar{F}_s(\bar{r}_1)\rangle = k_0^4 \int_{V_1} d\bar{r}_1' \, PV.\bar{\bar{G}}_g(\bar{r}_1,\bar{r}_1')\langle \bar{Q}(\bar{r}_1') \int_{V_1} d\bar{r}_1'' \, PV.\bar{\bar{G}}_g(\bar{r}_1',\bar{r}_1'')\cdot \bar{Q}(\bar{r}_1'')\cdot \bar{F}_m(\bar{r}_1'')\rangle \quad (19)$$

If let $V_1 \to 0$, then $\langle \bar{F}_s(\bar{r}_1)\rangle$ should approach to zero. Then, we should have

$$\lim_{V_1\to 0} \langle \bar{Q}(\bar{r}_1') \int_{V_1} d\bar{r}_1'' PV.\bar{\bar{G}}_g(\bar{r}_1',\bar{r}_1'')\cdot \bar{Q}(\bar{r}_1'')\rangle = 0. \quad (20)$$

If V_1 has the shape of an equal-correlation surface around $\bar{r}''=\bar{r}'$, the correlation may be factored from inside the integral. Obviously, it implies (20), since

$$\lim_{V_1\to 0} \int_{V_1} d\bar{r}_1'' PV.\bar{\bar{G}}_g(\bar{r}_1',\bar{r}_1'') = 0 \quad . \quad (21)$$

Following Stogryn's approach (ref.6) to determine the coefficient $\bar{\bar{S}}$ in the singularity of the Green's function $\bar{\bar{G}}_g$, it can be finally obtained that

$$\bar{\bar{S}} = -\lim_{V_1\to 0} \frac{k_0^2 \int d\bar{r}\,\bar{\bar{G}}_g(\bar{r})}{\varepsilon_0^{3/2}} = \frac{1}{4\pi\sqrt{\varepsilon_{gx}\varepsilon_{gy}\varepsilon_{gz}}} \lim_{V_1\to 0} \int_{V_1} dr \, \nabla\nabla(1/r_g) \quad . \quad (22)$$

Taking the correlation function of (15), and integrating over the surface

$$\rho^2/L_1^2 + |z|/L_z = 1 \quad ,$$

the explicit formulations are as follows

$$S_1 = \varepsilon_0 b^{1/4}/\{\varepsilon_g(2b^{1/2}+1)\} \quad , \quad (23a)$$

$$S_2 = S_1 \quad , \quad (23b)$$

$$S_3 = \varepsilon_0/\{\varepsilon_{gz}(2b^{1/2}+1)\}, \quad (23c)$$

$$b = \varepsilon_g L_z^2/(\varepsilon_{gz} L_1^2) \quad . \quad (24)$$

Using $\langle \bar{Q}\rangle = 0$ to determine $\bar{\bar{\varepsilon}}_g$ (ref.3,4,6),

$$\left\langle \frac{\varepsilon(\bar{r}) - \bar{\bar{\varepsilon}}_g}{\varepsilon_0 + S_1(\varepsilon(\bar{r}) - \varepsilon_g)} \right\rangle = 0 \quad , \quad (25a)$$

$$\left\langle \frac{\varepsilon(r) - \varepsilon_{gz}}{\varepsilon_0 + S_3(\varepsilon(r) - \varepsilon_{gz})} \right\rangle = 0 \quad . \quad (25b)$$

Substituting S_1, S_3 of (23) into (25), and iterating numerically, the solutions of S_1, S_3, and $\varepsilon_g, \varepsilon_{gz}$ are obtained. We may compare (23) with some limiting cases (ref.7). If $L_z = L_1$, then $b=1$, $S_1=S_2=S_3=\varepsilon_0/(3\varepsilon_g)$. The volume is a sphere. If $L_z >> L_1$, then $b\to\infty$, $S_1=S_2=\varepsilon_0/(2\varepsilon_g)$, $S_3=0$. V_1 is a thin cylinder, or a pin. If $L_z << L_1$, then $b\to 0$, $S_1=S_2=0$, $S_3=\varepsilon_0/\varepsilon_g$. V_1 is a disk. The results are consistent with Ref.7 in limiting cases.

If the random medium is a two-phase medium composed by two constituents with ε_s and ε_b, and the fractional volume f_s and $1-f_s$, respectively. Then, (25) are written as

$$f_s \frac{\varepsilon_s - \varepsilon_g}{\varepsilon_0 + S_1(\varepsilon_s - \varepsilon_g)} + (1-f_s)\frac{\varepsilon_b - \varepsilon_g}{\varepsilon_0 + S_1(\varepsilon_b - \varepsilon_g)} = 0 \quad , \quad (26a)$$

$$f_s \frac{\varepsilon_s - \varepsilon_{gz}}{\varepsilon_0 + S_3(\varepsilon_0 - \varepsilon_{gz})} + (1-f_s)\frac{\varepsilon_b - \varepsilon_{gz}}{\varepsilon_0 + S_3(\varepsilon_b - \varepsilon_{gz})} = 0. \quad (26b)$$

4. Numerical Results of the Vector Thermal RT Equation

Using the results in the last two sections, we obtain the vector RT equation for a layer of strongly-fluctuated continuous random medium, which has new constituents of the phase function and scattering coefficient. We use the Gaussian quadrature (N=16) and the eigenanlysis approaches to solve the RT equation (ref.5), and obtain numerical results of the brightness temperature versus other geophysical parameters, such as wavelength, observation angle, polarization, correlation function and correlation length, layer depth, and medium composition, etc.. Comparision with the conventional RT is available.

The solid lines in Fig.1 give the functional relation between the brightness temperature and observation angle. The results

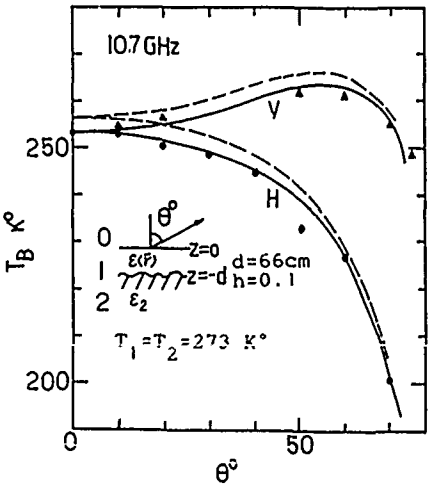


Fig.1. Brightness temperature vs angle $\varepsilon_s=(3.2+i0.01)\varepsilon_0$, $f_s=0.4$, $\varepsilon_2=(6+i0.6)\varepsilon_0$.

are used favorably to match the experimental data (ref.8) of snowpack, which is modeled as a layer of random medium composed by dry ice, and air. We choose L_z=0.06cm, L_l=0.03 cm, which follow the experimental correlation lengths of snow grain sample (Vallese and Kong, J. Appl. Phys.,52(1981)). Other parameters are given by the experiment, and are listed in the figure. The dashed lines come from the calculation by using the conventional RT equation. We can see that the conventional RT overestimate the thermal emission. Generally, ε_m'' is always larger than ε_g'', ε_{gz}'', where " denotes the imaginary part. If the layer depth d is not too large, thermal emission evaluated by ε_m and D_ε is always higher thanthat by ε_g,ε_{gz}, and D_{11}, D_{33}, D_{13} .

For possible effect due to the small roughness of the subsurface z=-d, we simply modify the reflectivity r_{a12} (Choudhury, etc. J. Geophys. Res., 84(1979)) as

$$r_{a12}(\theta)=\exp(-h\cos^2\theta)r_{a12}^0(\theta) \ , \ a=v,h \quad (27)$$

where h is the effective roughness, r_{a12}^0 is the reflectivity of smooth surface.

5. Conclusion

We have applied the strong fluctuation theory to derive the phase matrix and scattering coefficient for the vector radiative transfer equation for strongly fluctuated, nonisotropic, continuous random media. The new permittivity $\bar{\bar{\varepsilon}}_g$ is then numerically determined. Our RT equation is applicable to both strongly and weakly fluctuated random media. Employing the Gaussian quadrature and eigenanalysis approaches, the vector thermal RT equation is solved. Numerical results are favorably matched with the experimental measurement, and are compared with the conventional RT equation, which overestimated the specific intensity.

Acknowledgment

This work is supported by National Natural Science Foundation of China (NNSFC), and Fok Ying Tung Education Foundation.

References

(1). V.I. Tatarskii, "Wave propagation in a turbulent medium", McGraw Hill, NY(1961).
(2). L.Tsang and J.A. Kong, Radio Science, 11, 599(1976).
(3). L. Tsang and J.A. Kong, Radio Science, 16, 303(1981).
(4). Y.Q. Jin and J.A. Kong, J. Appl. Phys., 55, 1364(1984).
(5). Y.Q. Jin and R.G. Isaacs, JQSRT, 37, 461(1987).
(6). A. Stogryn, "Electromagnetic properties of random media", Rep. NO.6140, Aerojet ElectroSystems Co., CA(1981).
(7). Van Blandel, IRE Trans. Ant and Prop., 9, 563(1961).
(8). J.C. Shine, etc., Proceedings of the 12th international symposium on remote sensing of the environment, ERIM, Ann Arbor, Mich., 2, 877(1978).

RADIOBRIGHTNESS OF PERIODICALLY HEATED, TWO-PHASE MEDIA

A. W. England

Radiation Laboratory
Department of Electrical Engineering and Computer Science
The University of Michigan
Ann Arbor, MI 48109
(313)-763-5534

December 19, 1988

Abstract

Soils that contain liguids or gases that freeze during diurnal insolation will appear radiometrically distinctive. That is, they will appear to have anomalously high thermal inertias caused by the latent heats of fusion or sublimation. The effect should be observable in the diurnal variation of the radiobrightness of freezing moist soils, or in the radiobrightness of Mars soils that are saturated with CO_2 ice.

The 1-dimensional, heat flow equation for moist soils,

$$\frac{\partial E(T)}{\partial t} = \frac{\partial}{\partial z} K(T) \frac{\partial T}{\partial z}$$

is non-linear because both the enthalpy, E(T), and the thermal conductivity, K(T), are non-linear functions of T at freeze/thaw phase boundaries. Furthermore, diurnal insolation may cause phase boundaries at more than one depth, z. The problem is particularly difficult because these phase boundaries propagate and, occasionally, cancel themselves, and because soils that contain clay freeze over a range of temperatures rather than at 0^o C--that is, they possess diffuse phase boundaries.

The problem of periodic heating of two-phase media has come to be known as Stefan's problem. There are several numerical techniques for its solution. The Chernous'ko method was most readily modified for diffuse phase boundaries, and was developed as a modeling tool for examining the radiobrightness of diurnally heated soils. These models exhibit diurnal radiobrightness spectral gradients similar to those computed from the 10.7, 18, and 37 GHz radiobrightness temperatures from the Scanning Multichannel Microwave Radiometer (SMMR) on Nimbus-7 (reported in a separate abstract, England, et al), and may explain the anomalously flat radiobrightness spectrum of Mars.

TEXTURE ANALYSIS IN MATCHING 3D TREE MODELS TO REMOTELY SENSED DATA

Don Jayasinghe*, Richard Fournier** and John Miller*,**

INSTITUTE FOR SPACE AND TERRESTRIAL SCIENCE*
CENTRE FOR RESEARCH IN EXPERIMENTAL SPACE SCIENCE**

York University
4700 Keele St.
North York, Ont. M3J 1P3
Telephone: 416-736-2100 ext. 3510

Abstract

A geometrical-optical model has been developed to represent the radiative properties of a known forest Canopy for a given illumination and viewing geometry in terms of the physical Canopy parameters. Elements of this model for an individual tree will be presented. The parameter expression used in producing the three dimensionel tree shapes will be outlined. Presented also, is a comparison of the radiances of the model and the original, at similar conditions of illumination, viewing geometry with respect to their gray level frequency distribution and patterns. A comparison of the texture measures such as entropy, inertia, correlation, contrast, sum of squares, inverse difference moment, sum variance, difference variance, difference entropy and maximal correlation coefficient which were calculated within spacial gray level dependence method (SGLDM) for the original trees will be presented. Statistical tests such as student t-test and correlation coefficient analysis have been conducted with the texture measures obtained for the model and the original to show how closely they resembled each other.

A Set of Surface Emissivity Simulation for Microwave Remote Sensing from Satellite-Borne Sensors

Ya-Qiu Jin

Department of Electronic Engineering, Fudan University
Shanghai, China

Abstract

A set of different surface modelings of discrete scatterer and continuous random media is combined into a subpackage for the emissivity simulation of distinct surfaces, and radiative transfer in microwave remote sensing. The wave analytic theory with fluctuation dissipation approach, the radiative transfer equation, the radiative wave equation, and approaches for scattering from rough surface are applied to the simulations. Numerical results are matched with the available measurements.

1. Introduction

The microwave radiance received by satellite-borne sensors is contributed by the atmosphere (precipitation, cloud, etc.), the earth terrain, and their interaction, which may be understood as a system. The radiative transfer is significantly governed by the emissivity of the distinct surfaces. Simulation for this microwave system effectiveness requires the capability to determine the functional dependence of reflectivity and emissivity due to changes in the surface type, and relevant geophysical and hydrological parameters. This information was usually provided via available empirical data sets for pratical use. However, the range of available fieldmeasurements is seriously limited. In order to facilitate the evaluation of sensor scenorios, a subpackage of the earth terrain modelings for dry and wet snow, sea ice, calm and rough sea surface, vegetation and forestry, dry and wet soil, etc., and respective numerical approaches are developed. The geometry of these models are illustrated in Fig.1. They are one or multi-layer of discrete scatterers or continuous random media with possible rough subsurface. The wave analytic theory with fluctuation-dissipation approach, the radiative transfer theory, the radiative wave theory, and approaches to scattering from rough surface are applied to developing numerical algorithm to determine the emissivity and reflectivity of the models. Numerical results are favorably matched with the available measurements by means of truck-mounted, airborne, and satellite radiometers

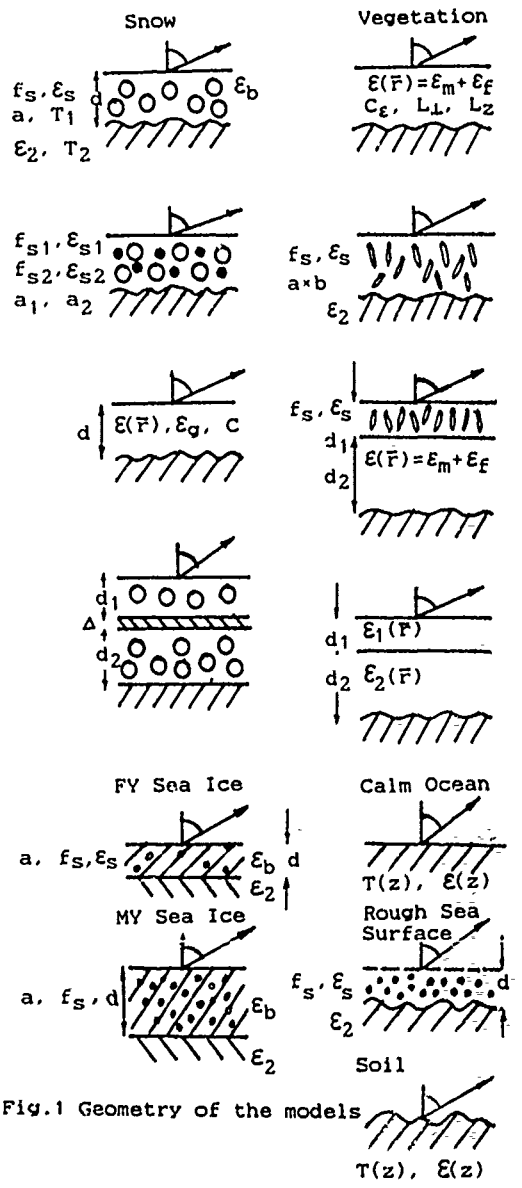


Fig.1 Geometry of the models

(e.g. DMSP SSM/I data).

This study also provides a theoretical basis for understanding the sensitivity of the microwave sensor data to variations in relevant parameters, and comparing the consistency of different theories and approaches for microwave remote sensing.

2. Surface Models and Approaches

(1). Snow:

Snowpack is modeled as one or multi-layer of discrete scatterers (Fig.1). Dry snow is composed of ice particles embedded in air, and the fractional volume f_s of ice particles is generally larger than 0.2 ($f_s=\rho_g/0.91$, here ρ_g is the density gm/cm^3). The radius of the effective ice grain is around 0.4~0.8 mm. The snowpack with refrozen laminas is modeled as multi-layer media. Wet snow consists of small water drops (radius a~0.2mm), ice particles, and air. The fluctuation-dissipation theorem is applied to obtain the brightness temperature of the snowpack. The brightness temperature in zeroth-order approximation is

$$T^0_{B\beta}(\theta)=k_0^3\cos\theta(\varepsilon''_g/\varepsilon_0)\int_{-d}^{0}dz_1T_1(z_1)|A_1(\theta)_\beta|^2 \cdot|(R_{\beta 12}\exp(i2k_{1z}d)\hat{\beta}(-k_{1z})\exp(ik_{1z}z_1) +\hat{\beta}(k_{1z})\exp(-ik_{1z}z_1))/k_{0z}|^2 +k_0^3\cos\theta(\varepsilon''_2/\varepsilon_0)\int_{-\infty}^{-d}dz_2T_2(z_2)|A_2(\theta)_\beta|^2 \cdot|(R_{\beta 12}\exp(i2k_{1z}d)\hat{\beta}(-k_{2z})\exp(ik_{2z}z_2) +\hat{\beta}(k_{2z})\exp(-ik_{2z}z_2))/k_{0z}|^2 , \quad (1)$$

where $\hat{\beta}$ denotes the vertical or horizontal polarization, $A_1(\theta)_\beta$, and $A_2(\theta)_\beta$ represent the transmission coefficients from the bounded region 1 and the underlying region 2, respectively. k_1 is the effective propagation constant of the region 1. R denotes the reflection coefficient. The permittivity ε_g is obtained by the strong fluctuation approach (ref.1-3) as follows

$$f_s\frac{\varepsilon_s-\varepsilon_g}{\varepsilon_s+2\varepsilon_g}+(1-f_s)\frac{\varepsilon_b-\varepsilon_g}{\varepsilon_s+2\varepsilon_b}=0 \quad (2)$$

The high-order solution of $T_B(\theta)$ also has been obtained (ref.1). Essentially, $T_B(\theta)$ is governed by the value of ε''_g, which is the imaginary part of the zeroth-order approximated effective permittivity of the random medium, region1. The emissivity from dry and wet snowpack are shown in Fig.2. There is a significant decrease in the emissivity due to the presence of snow, especially for vertical polarization. The emissivity of wet snow is generally higher than dry snow. The emissivity versus snow wetness is given in Fig.3. It can be seen that the increase of snow wetness initially increase the emissivity, then will cool the thermal emission, especially at high frequency.

The lamina structure due to refrozen surface will darken and cool thermal emission. Numerical results and comparisons with the measurements are shown in Fig.4.

Since the fractional volume of ice particles is large, the conventional radiative transfer equation with independent scattering assumption has to be modified. The radiative wave equation (RWT, ref.4,5) has been deve-

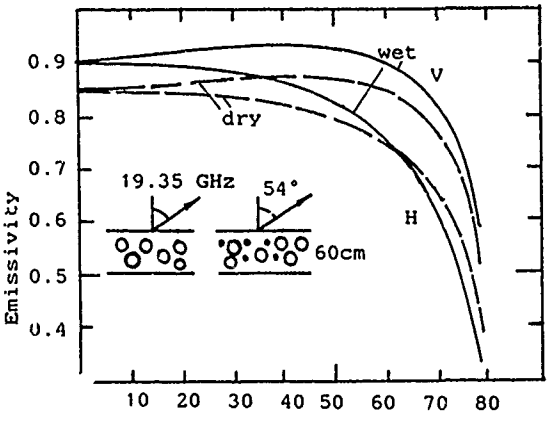


Fig.2 Emissivity vs Observation Angle for dry and wet snowpack

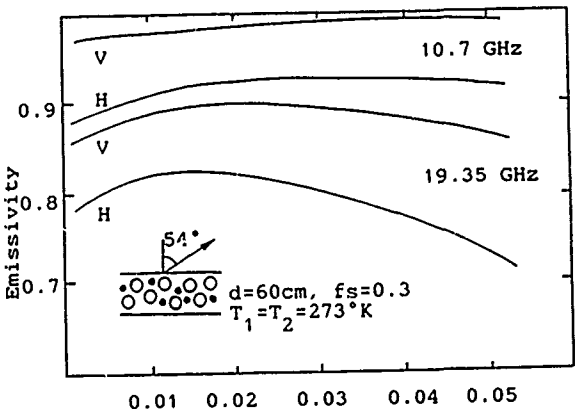


Fig.3 Emissivity vs Wetness of Snowpack

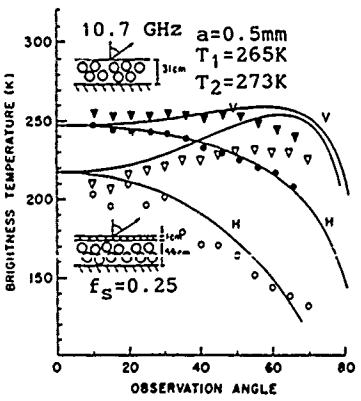


Fig.4 Brightness temperature vs Angle for two cases of snowpack

loped to take account of the scattering coherence of densely-distributed scatterers. The RWT equation follows the conventional radiative transfer equation (RT) in form, however the scattering coefficient and the phase matrix has been modified. The RWT equation is solved by the discrete ordinate method. A numerical example is shown in Fig.5. RWT and RT is compared.

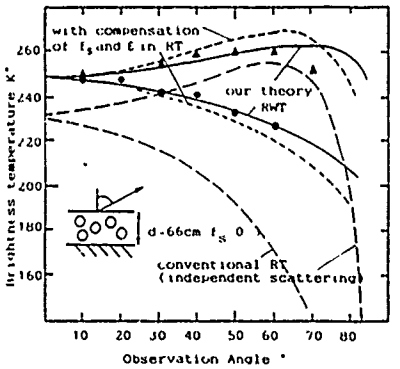


Fig.5 Brightness temperature vs Angle from RWT and RT approaches

Snowpack may also be modeled as a layer of continuous random medium (ref.6). The correlation function of the permittivity perturbation is generally Gaussian or exponential (ref.7). The correlation lengths from experiments of snow grain sample (ref.8) are about 0.3~0.8 mm. Note that this value is roughly equal to, or little smaller than the effective size of ice particle in discrete scatterer model.

(2). Sea ice:
The model for sea ice is a layer of ice with air pockets. The emissivity of sea ice depends on age, thickness, salinity, density, composition and surface characteristics. First year sea ice (FY) has a layer of thin congelation ice (d~20cm, ρ_s~0.85 gm/cm ; equivalently f_s~0.008) over a thick ice layer (ρ_s ~0.91 gm/cm). Multi-year sea ice (MY) has a recrystallized ice layer (d~50cm, or more) over a ice layer with thickness up to 500 cm. My sea ice has a large fractional volume of air pockets (f_s~0.25), which causes more scattering, and darken the thermal emission. By using the fluctuation dissipation approach, the example of the emissivities of FY and MY sea ice is shown in Fig.6.

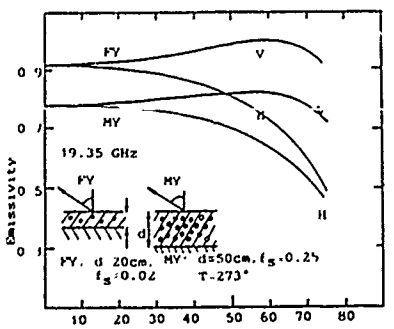


Fig.6

(3). Calm and rough sea surface:
Ocean with smooth surface is modeled as a simple half-space dielectric medium, whose permittivity is calculated by Debye formula. The emissivity of a rough sea surface is greater than that of a smooth sea because the wind-driven rough sea surface is covered with a layer of foam and white caps. The effect of white water (foam and white caps) is modeled by a layer of random discrete scatterers with a randomly rough surface, as shown in Fig.1. Both the wave analytic theory and the RT theory can be applied to the calculation of the emissivity. The Kirchhoff approach for wave scattering from randomly rough surface with large mean slope, or small perturbation method, two-scale rough surface approach (ref. 9,10,11) are readily applied to modify the reflectivity and transmittivity appeared in the boundary conditions. A numerical example of the brightness temperature from the atmospheric precipitation over rough sea is given in Fig.7. The emissivity increases due to the foam layer, which corresponds to a wind speed of about 10 m/sec.

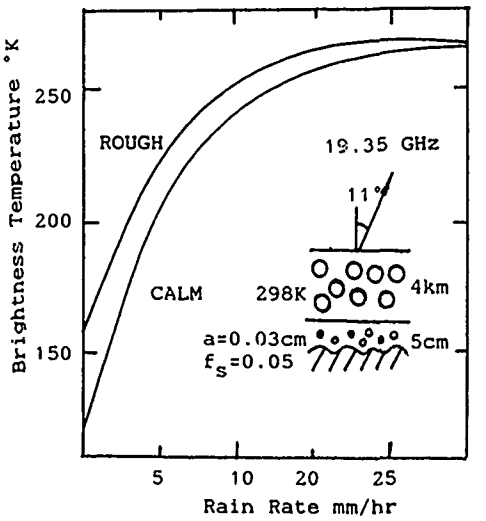


Fig.7 Brightness temperature vs Rain rate for the models of calm and rough sea surface

(4). Vegetation:
Vegetation canopy is modeled as a layer or multi-layer of discrete scatterers (more usually nonspherical, e.g. spheroid, cylinder, disk, etc.) (ref.12), or a continuous random medium (ref.13) with Gaussian or exponential correlation function, and noisotropic correlation lengths. The several examples in the modeling are also shown in Fig.1. The fractional volume of scatterers in vegetation is generally smaller than 0.1. The RT equation is applied to solve multiple-scattering and thermal emission from vegetation. For the two-distinct layers model, two RT equation with coupling boundary conditions have to be solved. The permittivity of vegetation canopy (leaf, grain, etc.) follows the empirical formulations of the four-phase refractive model (ref. 16). Then, the mean permittivity of whole layer (vegetation and air) is obtained by the

mixture formula, which is related with the scatterer shape. The correlation lengths (around the order of mm and cm) correspond to the physical structure of the vegetation, and are choosen to optimally match with the experimental data. Fig.8 shows a numerical result and matching with the data of full grown corn at 1.4 GHz. The vertical and transverse correlation lengths are 1 cm and 0.1 cm, respectively.

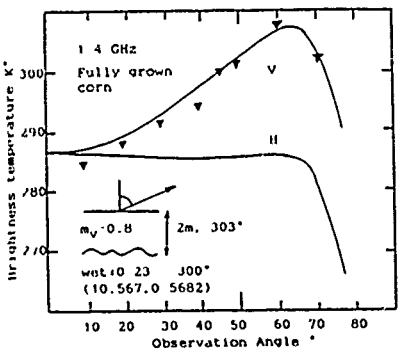


Fig.8 Brightness temperature vs Angle for a layer of full grown corn

Increasing the correlation lengths is, equivalently, to increase the size of scatterer, and will initially increase the brightness temperature, then will significantlly darken the emission due to strong backscattering.

(5). Soil:
Theoretical analysis of thermal emission from dry and wet soil is usually restricted to finding an effective permittivity for soil. By using the fluctuation dissipation approach the emissivity from wet soil with temperature and wetness profiles can be obtained. It can be seen also from Eq.(1). The mean permittivity of wet soil is based on the empirical mixture formula (ref.14). The effect from the randomly rough surface can be properly taken into account by employing the approaches of scattering from rough surface. Soil land is often modeled as a half-space underlying homogeneous or inhomogeneous medium for remote sensing of snow, vegetation, etc..

(6). Simulation of Atmospheric Radiative transfer
Simulated SSM/1 brightness temperature is calculated by the radiative transfer

$$I(\theta)=e_s(\theta)B(T_S)\exp(-\int_0^P k dp'/\cos\theta) + [1-e_s(\theta)]\exp(-\int_0^P k dp'/\cos\theta)\int_0^{P_S} B(T(p))d\tau' + \int^0_{P_S} B(T(p))d\tau' \quad (3)$$

where $\tau'(p)=\exp(-\int_p^{P_S} k dp'/\cos\theta)$. $I(\theta)$ is the microwave radiance, B is the Plank function, P_S is the surface pressure, k is the gas absoption coefficient, τ' is the downward transmission function. Appropriate atmospheric profile can be obtained as a input data. The emissivity of the surface $e_S(\theta)$ is calculated by the above modeling simulations. DMSP SSM/I data (ref.15) was collected over the ocean, moderate and heavy vegetation, and snowpack. The similations provide the spectral properties of the SSM/I brightness temperature, and is favorably matched with DMSP DMSP SSM/I data collected over the selected areas. The results are given in Fig.9.

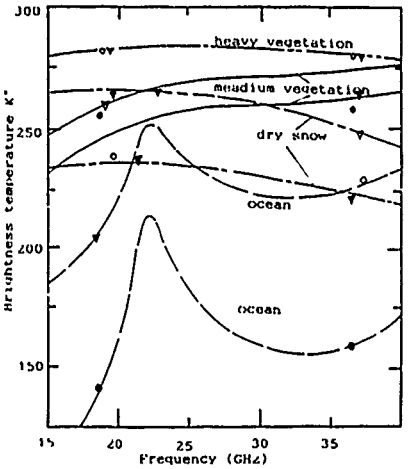


Fig.9 Brightness temperature vs Frequency for comparision with DMSP SSM/I data over selected areas

Acknowledgment

This ongoing work is supported by National Science Foundation of China (NNSFC), and Fok Ying Tung Education Foundation.

References

[1]. Y.Q. Jin, Electromagnetics, 4(1984), 323.
[2] L. Tsang and J.A. Kong, Radio Science, 16(1981), 303.
[3]. Y.Q. Jin and J.A. Kong, J. Appl. Phys., 55(1984), 1364.
[4]. L. Tsang and A. Ishimaru, J. Elect. Wave & Appli., 1(1987), 59.
[5]. Y.Q. Jin, J. Quant. Spectrosc. Radiat. Transfer, 39(1988), 83.
[6]. L. Tsang and J.A. Kong, Radio Science, 11(1976), 599.
[7]. A. Stogryn, "EM properties of random media", Rep. N0.6140, Aerojet ElectroSystems Co., CA(1981).
[8]. F. Vallese and J.A. Kong, J. Appl. Phys., 52(1981), 4921.
[9]. P. Beckmann and A. Spizzichino,"The scattering of EM waves from rough surface (Pergamon, New York, 1963).
[10]. A.K. Fung and M.F. Chen, Radio Science, 16(1981), 289.
[11].A.K. Fung and H.J. Eom, IEEE Trans. Ant Prop., AP-29(1981), 899.
[12]. M.A. Karam and A.K. Fung, Radio Scienc, 18(1982), 752.
[13]. S.L. Chuang and J.A. Kong, J. Appli. Phys., 51(1980), 5588.
[14]. J.R. Wang and T.J. Schmugge, IEEE trans. Geosci. & Rem. Sens., GE18(1980), 289.
[15]. R.G. Isaacs, Y.Q. Jin, R.D. Worsham, G. Deblonde, and V.J. Falcone, to be published.
[16]. F.T. Ulaby and R.P. Jedlicka, IEEE Trans. Geosci. & Rem. Sens., GE22(1984), 406.

REDUCTION OF INTERFERING SIGNALS IN SWEPT-FREQUENCY RADAR MEASUREMENTS USING COMPLEX WIENER FILTERING

David E. Weissman
Dept. of Engineering
Hofstra University
Hempstead, New York 11550

Brian Jersak
Dept. of Electrical Engineering
University of Texas at Arlington
Arlington, Texas 76019

Leo D. Staton
Antenna and Microwave Research Branch
NASA Langely Research Center
Hampton, Virginia 23665

A new approach to the reduction of scattered, interfering signals that corrupt measurments of the signal backscattered from radar targets of interest is being developed. It is being tested with sphere measurements in an indoor microwave radar range. This method is based on the concept of Wiener filtering (which minimizes the difference between the signal plus noise and the desired signal in the time domain). In contrast to the traditional Wiener filter, in which the time domain difference between two sequences are minimized, the approach reported here employs the frequency domain phasor amplitudes of a swept frequency signal. It minimizes the difference (least-mean-square-magnitude) between the signal-plus-noise and the signal complex phasors, across the entire spectrum. The Wiener filter therefore has complex coefficients which act on the phasors associated with the sequence of frequencies. The design of the filter is based on the complex autocorrelations of the signal-plus-noise and the noise. This approach is attractive in radar range measurments where the support structures create interfering backscattered waves. Initial tests with a 48 coefficient filter applied to 401 complex phasor amplitudes indicate successful results.

A HIGHRESULOTION FULLY POLARIMETRIC C-BAND RADAR

Anders Stjernman, Tommy Sturk, Gudmund Wannberg

Swedish Institute of Space Physics, Box 812, 981 28 Kiruna, Sweden
Telephone: 46 980 12244, Telex: 8754 IRF S, FAX: 46 980 15465

ABSTRACT

In preperation for interpretting ERS-1 data for land use, a C-band radar has been developed at the institute. The radar consists of two 20 dB standardgain horns, placed with orthogonal polarizations in the top of a 27 m high transportable telescopic mast and connected to the two ports of the S-parameter testset in our HP 8510B network analyser system with two 35 m long lowloss cables. The horns, which are fully steerable in azimuth and elevation, are both used both for transmitting and receiving and the builtin twelve-errorterm-correction of the network analyser is used to calibrate the measurements to get the four complex S-parameters simultaneously over a number of discrete frequencies. The data is then transformed to time domain where it is corrected for the antenna gain and for the range dependence of the illuminated beam area and of the spreading loss.The complex scattering matrices from individual scatterers within the mainlobe can thereby be determined with a range resolution downto approximately 10 cm (max bandwith of the horns is 2.05 GHz). In the future it is planned to expand the system with horns in the L-band and X-band or with a broadband (1-18 GHz) crosspolarized antenna depending on future grants.

1. INTRODUCTION

As a part of a national program the Swedish Institute of Space Physics has initiated a project that will investigate the possibility of using ERS-1 data for land resources overwiev. The reasons for locating this activity at Kiruna was partly because of the location of the ESA-ground station that is going to receive data from and to some part control the European ERS-1 satellite.

In order to meet future demands on interpretting data from polarimetric satellites it was decided that we should not only do measurements in the same polarization and frequency as the ERS-1 but we should also do fully polarimetric measurements at several frequency bands. As a first step towards this capability we applied for and received grants from the Wallenberg foundation to purchase a network analyser to be used as a transmitter and receiver as well as a data processing unit in the whole frequency range of interest.

2. SYSTEM DESCRIPTION

To be able to do measurements of fairly high trees in a geometry that is similar to that of ERS-1 with an incidence angle of 22 degrees it is necessary to place the radarantennas at a height of nearly 30 m. It is also essential to be able to make measurements at many different locations to collect data from a great number of different vegetation types. The choice we did was to purchase a mast that is mounted on a trailer and made of five telescopic sections which allows the mast to be raised to a height of 27 m. The permissable antenna area is 1 m^2 and max load is 1000 N. We also decided that we should use the S-parameter testset, the calibration capability and the large dynamic range of the networkanalyser to make polarimetric measurements with only two linearily polarized pyramidial horns. The frequency band of the horns was chosen to be compatible with that of ERS-1 i. e. 5.3 GHz. The networkanalyser and the hostcomputer is placed in a van which can tow the trailer.

In order to make the system even more flexible we also purchased a computer controlled antennarotor steerable in 180 degrees in elevation and 360 degrees in azimuth. The cables that connect the horns with the networkanalyser are 35 m long and has only 10 dB transmisson loss at 5.3 GHz and are usable up to a frequency of 10.2 GHz.

In purpose of getting better documentation over the measured data we also have placed a remotely controlled videocamera between the horns looking in the same direction as them.

3. CALIBRATION TECHNIQUE

The reflections and the attenuation in the cables and connectors are numerically compensated for by the 12-errorterm calibration model that is implemented in the network analyser. In order to compute the right errorterms we are using the calibration kit that we purchased for our dielectric measurements and we are making the calibration with TRL-calibration technique. In that way we can calibrate at the waveguideside of coax-to-waveguide adapters that connect the coaxcables with the horns. By viewing the data in timedomain we immediately get an indication on how well our calibration

succeeded. The horns themselves are low reflective and practically lossless why in a first approximation their influence can be neglected and merely gated away The antenna gains are rather readily calculated from their approximate aperture functions and can be compensated for in the frequency domain, afterwards the range dependent terms in the radarequation can be corrected for in the time domain after an FFT of the data. By converting back to frequency domain one should get a good approximation of the scattering matrices as seen from the satellite and there should also be a fading effect in frequency which makes it possible to calculate the necessary means in order to determine the scattering coefficients.

4. MEASUREMENT STRATEGY

We are planning to select a number of locations in the neighbourhood of Kiruna which represent different vegetationtypes such as different kinds of forest, clearcuts, swaps and so on. In order to get temporal information we are planning to revisit those locations at different seasons and under varying weather conditions. Beside the cross-section measurements we are planning to measure the dielectric constant of some key materials that could effect the radar cross-section.

5. CONCLUSION

With our radarsystem we are filling the gap between airborne radarsystems and truckmounted radarsystems. We have the possibility to control the exact position of the radar and to revisit the same spot many times in order to make comparisons of the measured data under varying temporal conditions and at the same time we have the possibility to do measurements of rather high forest from above the canopies to resemble satellite measurements. By using the S-parameter testset of our network analyser we are able to do polarimetry with only two antenna elements so long they are low reflective and connected to the ports with lowloss cables.

6. REFERENCES

1 HP 8510B NETWORK ANALYSER Operating and programming manual, part number.HP 08510-90070 Hewlett-Packard, 1985.

2 A. Stjernman, T. Sturk, G. Wannberg, A WAVEGUIDE DEVICE FOR MEASURING THE COMPLEX DIELECTRIC CONSTANT OF HOMOGENEOUS MATERIALS IN C-BAND; Submitted to IGARRS'89

3 E. V. Jull, APERTURE ANTENNAS AND DIFFRACTION THEORY; IEE Electromagnetic waves series 10,1981 ISBN 0-906048-52-4

ROLE OF SPACE OBSERVATIONS IN THE CANADIAN GLOBAL CHANGE PROGRAM

J. Cihlar[1], E. LeDrew[2], H. Edel[3], W. Evans[4], G. McKay[4], L. McNutt[1], A. Royer[5]

[1] Canada Centre for Remote Sensing, Ottawa, Ontario
[2] University of Waterloo, Waterloo, Ontario
[3] Department of Fisheries and Oceans, Ottawa, Ontario
[4] Atmospheric Environment Service, Downsview, Ontario
[5] University de Sherbrooke, Sherbrooke, Quebec

ABSTRACT

This paper describes a proposed program of space observations as part of the Canadian Global Change Program. The proposed program was developed by the Remote Sensing Technical Group established by the Royal Society of Canada. In addition to areas of emphasis and guidelines for the selection of projects, requirements for satellite data and data handling systems (including data integration, communications) are discussed. The full program is contingent upon additional funding.

RÉSUMÉ

Cet article décrit un programme d'observations spatiales faisant partie du Programme canadien de la transformation du globe. Le Groupe technique de la télédétection constitué par la Société Royal du Canada a développé ce dit programme. Les champs d'action et les directives pour la sélection des projets sont discutés; ainsi que les spécifications pour les données satellitaires et les systèmes de manipulation de données (incluant l'intégration des données, les communications). Le programme général en lui-même est conditionnel à des financements complémentaires.

1. INTRODUCTION

In recent years the scientific community, decision makers and the population at large have come to realize that the global environment is undergoing substantial changes. The crucial importance of the state of the environment for the well-being and future prosperity of humankind has been brought to focus at the political and economic levels as well (Lecuyer and Aitken, 1987; Anonymous, 1988). Many of the present or anticipated environmental problems are continental or global in nature, and the traditional data gathering as well as processing techniques are no longer sufficient for either monitoring or process studies.

Global observations of the Earth with satellite sensors are required simply because no other approach permits measurements to be made at the required spatial and temporal resolutions. Satellite sensors can acquire information about the geosphere, biosphere, hydrosphere, cryosphere, and atmosphere. In addition to monitoring changes as they occur, this information makes possible new insights into how the planet functions (Malone, 1988). Many of the changes in the global environment may be monitored on a world-wide basis with remote sensing satellite systems. Following several remote sensing satellites planned for the early 1990s, the Earth Observing System (Eos) will be launched by NASA, the European Space Agency (ESA) and the Japanese NASDA starting in 1996. Eos platforms will carry a large array of sensors designed to monitor the Earth as one system, and will provide the first comprehensive coverage of global processes from a complementary suite of remote sensing instruments.

Realizing the serious challenge posed by the various environmental changes, the International Council of Scientific Unions initiated an International Geosphere-Biosphere Programme (IGBP) in 1986. In preparation for the program of global change studies, the Royal Society of Canada initiated a planning process to define a Canadian Global Change Program (CGCP). A Remote Sensing Technical Group was set up to assist in this planning process by defining a remote sensing contribution to the program. This paper provides a brief overview of the main findings of this Group; a full report will be published separately (Cihlar et al., 1989).

2. OBJECTIVES AND SCOPE

Given the long-term nature of the global change program (measurements over a 20-year period are envisioned); the early stages of the planning process; and the research that needs to be done before space observations can routinely be used for monitoring global change, it is evident that the program should be planned in two stages. The objectives can realistically be defined only for the first stage during which preparations for the routine monitoring of the earth should be carried out. It is proposed that the following goals be achieved during the preparatory phase (i.e., by 1996):

1. Complete the development of procedures, models, and data handling systems which are required use satellite data for:

* detecting global changes over Canadian territory with emphasis on the boreal forest, grasslands, agricultural regions, the coastal zone, sea and lake ice, snow, permafrost, and the atmosphere;

* assessing the magnitude of the changes; and

* determining its current and future impact of these changes.

2. Develop next-generation satellite data processing and modelling technologies with emphasis on the integration of image analysis, geographic information systems, and digital networks.

3. Contribute to the definition, development, and implementation of international satellite observation programs for global change studies.

The above goals are motivated by the economic and social self-interest of Canada; its responsibility both as the second largest country in the world and a technologically advanced space nation; and its commitment to environmental preservation and sustainable development. The achievement of these goals will place Canada in an excellent position to establish a long-term, effective system for using space observations for global change as well as resource management purposes.

In order to develop the procedures and models for using the data from future satellite systems, investigations should be planned on a case by case basis. Where possible, advantage should be taken of other programs underway or planned, nationally or internationally. Some projects have been started or are under active consideration, others need to be developed. The range of investigations that should be undertaken in Canada include atmospheric correction models, albedo and net radiation mapping for climate modelling, vegetation monitoring and modelling, monitoring of atmospheric constituents, monitoring cryospheric variables and oceans monitoring. Details concerning these investigations are provided elsewhere (Cihlar et al., 1989). A final list of possible projects will result from further discussions within the CGCP, as well as in the international scientific community.

3. SATELLITE DATA AND ANALYSIS SYSTEMS NEEDS

3.1 Project selection criteria

The IGBP provides many challenges and opportunities. The human and financial resources that Canada may be able to devote to the program will likely not be sufficient to cover all the important questions, and choices will clearly be necessary. Three criteria are proposed to make these choices in the CGCP:

* The problem/issue selected should be important for global change in the Canadian or world-wide context. Importance could be economic or scientific, or preferably both.

* Problems/issues that are unique to Canada should receive priority. The main reason for "uniqueness" will be the geographic location of Canada on the globe and the consequent role of this northern latitude region in exhibiting, or affecting the trend of, global changes.

* Other conditions being equal, problems/issues where Canada has strength should receive priority.

3.2 Satellite data needed

Historical satellite data are kept in two basic types of archives, national and global. National archives usually maintain data received by that country through direct downlink to a ground receiving station. Generally, these are data either provided by sensors (often experimental) launched by that country, or data from foreign satellites which are considered important in the national context. Existing national archives typically contain data from sensors that provide images at high resolution and relatively high data rates. The Canada Centre for Remote Sensing (CCRS) archive of LANDSAT data received by Canada since 1972 is an example of this type.

Numerous satellite sensors gather data that need to be collected around the world to provide meaningful information (e.g., atmospheric measurements). These data can be efficiently handled by one agency in a global archive. Such an agency may have national mandate, or may serve the needs of the international scientific community (e.g., the World Data Centers). Global archived data sets will be crucially important for the success of global change studies, and the organization of these facilities therefore warrants special attention.

The Canadian program of earth observations for global change must assure access to both types of satellite data. It is important that data received in Canada be stored in permanent archives so that they are readily accessible and furthermore, Canadian scientists must be able to access relevant data from other satellites archived outside of the country.

During the operational period of the IGBP, the following satellite data should be available:

* Broad coverage of the entire territory (including the coastal zone) using sensors that measure atmospheric or surface variables. These could be coarse-resolution scanning sensors or profiling sensors.

* High spatial and spectral resolution coverage of selected sites on a sampling basis. In general, these sensors yield large amount of detailed local information but they cannot provide global coverage with a sufficiently short repeat cycle.

Both types of data described above will be collected by various satellites launched in different programs, and only some of these will be directly received by Canadian ground stations. This is why access to satellite data stored in foreign archives is important for the Canadian Global Change Program.

The multinational polar platform EOS program will require a linked network of earth observation data centres by the international partners to facilitate data access and exchange. The USA plans include the development of a data management system (EosDIS) that will serve the international scientific community studying global change. So far, there is only limited demonstration project experience in archiving and management of global data sets for supporting

scientific research. Canada will need to plan its participation in the global data archive and management by building on existing strengths in government data centres and in industry. A Canadian link to the international data exchange and archive network will need to be established to facilitate the access to data for global change which is necessary for active participation by Canada.

Data archived in Canada must be stored in a manner which will assure their preservation and retrieval for an essentially indefinite period of time. This requirement implies the permanency of the storage medium itself as well as the upward compatibility of the data storage and retrieval technology as the state-of-the-art develops in the future. With new storage media such as optical disks and optical tapes, the first problem is becoming effectively resolved. However, the question of recording data in a format that will be readable in the future will require continuing attention.

3.3 Data handling systems requirements

Research and applications carried out within the IGBP will be typified by extensive use of digital data which must be processed through a variety of inversion, descriptive, and predictive models. The data will reside in data banks managed by different agencies, primarily in Canada but data from other countries will also be required. It will be critically important that individuals involved in the IGBP have the means to determine what data are available, to obtain the required data sets from the data bases, and to combine and process these using appropriate models.

At the present time, the formats of individual digital data bases of potential importance to the CGCP are different, and access to the data is often difficult and time-consuming. They are accessed via computer tapes and usually involves reformatting and the resolution of compatibility problems. The data and information system needed for the CGCP should have the following features:

* Capability to determine from a remote terminal what data are available in the various data national or international data bases (regardless of the physical location of these data);

* Capability to ascertain (on - line) those characteristics of the data which will allow deciding whether the data should be acquired for the purpose at hand;

* Capability to transfer, or place an order for, a subset of a data base;

* Ability to combine data from various data bases and forms (point, line/polygon, raster) into a unifying framework, most often on a geographical basis;

* Capability to manipulate elements of the integrated data set;

* Capability to input the digital data into various models;

* Capability to run the models as part of development, verification, validation, or application activities, and to exchange the results.

Conceptually, the above capabilities can be provided by networking data bases (featuring a standardized access) with image analysis/geographic information systems; this configuration is termed GCNet. GCNet would take advantage of existing computer network capabilities. Its unifying theme will be the common interest of the GCNet participants in the problems of global change, and the joint development and use of tools needed for global change activities.

The geographic information system (GIS) concept has been developed principally for combining and manipulating various location - specific data types with the intended purpose of obtaining a new insight into a problem or deriving new information. For global change work, the GIS should have the following capabilities:

* To handle point, line, polygon, tabular, raster (image) and vector data easily and efficiently;

* To rapidly convert between raster and vector formats without loss of spatial accuracy;

* To answer queries concerning data content; to combine data layers; to run models using the data; and generally to apply various numerical spatial operations;

* To display data graphically and to provide various output formats (including cartographic outputs and simulated model results);

* To facilitate interactions with digital data bases listed above.

A number of systems are presently available commercially, in Canada as well as internationally. However, the existing systems do not have all the capabilities required for global change applications. This is due to the combination of demanding requirements and the continually developing state-of-the-art in GISs. The next generation of GISs is under development, and attempts are being made to build in some of the desired features (e.g., Smith et al. (1986).

The discrepancy between the requirements suggested above and present capabilities highlights the need to conduct further research and development in a focused way so that advanced GIS systems become available for global change activities. It should be emphasized that in comparison with existing GIS products,the advanced systems will also be of superior value for the management of natural resources (which currently represents a major use of GISs in Canada).

The development and implementation of a data and information systems described here will be a complicated process that will require time and the active cooperation of numerous agencies, in Canada and internationally. Clearly, it will not be possible to do a wholesale replacement of the existing facilities with new, compatible ones. However, it is possible to agree on a common set of standards as a goal, and to adopt a policy of gradual transformation of the existing systems to this standard (Office of Space Science and Applications, 1987). In this way, a distributed system with progressively increasing capabilities can be developed.

Developments with high relevance for advancing the GCNet concept are already taking place at the

Department of Fisheries and Oceans; the Atmospheric Environment Service; the National Research Council; the Department of Energy, Mines and Resources; and elsewhere.

3.4 Observatories

An integral component of a space observations is verification of image-based algorithms or interpretation procedures against ground data. Consequently, it is important that, in the IGBP planning process, provision for state-of-the-art field observatories be made that address the remote sensing requirements. The planning of such observatories is underway in Canada and internationally. A list of observational requirements for different components of the environment has been prepared (Cihlar et al., 1989).

4. CONCLUSIONS

The IGBP provides an opportunity for Canada to develop better understanding of the environment in which we live, and the way in which our way of life may be affected in the future by possible large - scale changes in this environment. With a more complete knowledge of the environment, we can implement economic and social policies in tune with the environmental imperatives and thus achieve the goal of sustainable economic development.

Since the monitoring and understanding of global change necessitates that repeated observations of a variety of biophysical parameters be made over very large areas, remote sensing - particularly satellite observations - must make a key contribution to ensure success for this program. This need places considerable responsibility on those involved in remote sensing research and applications as well as in the development of programs and policies for the future. It is clear that the potential for a pivotal remote sensing contribution to the CGCP exists. Its realization necessitates that much progress be made in the areas of satellite data acquisition, archiving and access; the extraction of biogeophysical information from satellite observations using inversion models; the establishment of systems that will facilitate global change research, including technological developments in certain key areas; and the formulation and implementation of a plan of activities that focuses the various techniques available for global change studies - laboratory, remote sensing, field, modelling - on problems that are of critical importance to the Canadian portion of the globe. From the viewpoint of remote sensing applications, it is very important to bear in mind that the techniques and knowledge developed through global change research will be directly applicable to the more established resource management uses of satellite data.

For remote sensing to make a contribution to global change studies, it is critical that the data be accessible at a low cost. This is because of the large volumes of data required and the fact that funding agencies do not have the funds that would be needed to purchase the data if policies such as those presently applied to high resolution satellite optical data applied.

The proposed remote sensing program is very ambitious. In our view, Canada has the necessary scientific and technological capabilities. However, significant new resources will be needed to achieve the intended goals. It should be remembered that these investments will provide payoff not only in understanding and assessing the impact of global changes but also in resource management and in the development of advanced information handling technologies.

5. REFERENCES

Committee on Space Research. 1987. Information Bulletin No.108, April 1987. 138 p.

Cihlar, J., E. LeDrew, H. Edel, W. Evans, G. McKay, L. McNutt, and A. Royer. 1989. Contribution of space observations to the Canadian Global Change Program. Report to the Royal Society of Canada by the Remote Sensing Technical Group. In print.

SOLAR SURFACE ALBEDO ESTIMATION USING REMOTELY SENSED SPECTRAL DATA

David L. Toll

Hydrological Sciences Branch
NASA/Goddard Space Flight Center
Greenbelt, MD 20771 USA

ABSTRACT

A solar albedo model based on converting narrow band satellite derived reflectance to four major spectral regions (ultraviolet, visible, near-infrared and shortwave middle-infrared) and weighted by the relative supply of global solar radiation is studied and further developed. Narrow band to broad band conversions within visible and near-infrared regions are shown to be accurate; however, the transformations are indicated to be surface feature dependent. Atmospheric aerosol and illumination effects are indicated to be near insensitive to spectral region integrations for solar albedo estimation, but are shown to be significant factors affecting the spectral and canopy albedo. The erectophile and spherical canopy albedos were sensitive to atmospheric and illumination condition, whereas the planophile was relatively insensitive. The contribution of a shortwave middle-infrared reflectance to the canopy albedo is shown to be an important component, accounting for albedo changes to 16% with increasing green vegetation.

Key Words: Albedo, Atmospheric Effect, Canopy Modeling

1. INTRODUCTION

The remotely sensed estimation of a surface, solar albedo, or the percentage of solar radiation reflected at the surface, is associated with numerous problems. In addition to basic remote sensing problems of sensor radiometric calibration, atmospheric attenuation with scattering, and emission corrections, image to map registration and sensor autocorrelation, there are problems associated with estimating the entire hemispherical surface exitance from the directional only satellite derived radiance, and from converting the sensor band pass derived reflectance to a broad solar band reflectance. There is an acute need to solve these problems for the determination of a solar albedo to study many land and climate processes.

Surface albedo (or the transpose, surface absorptan accounts for nearly three-fourths of all absor solar radiation in the Earth-atmosphere syst Several studies (Ref. 1) evaluated changes in the l to atmosphere reflected shortwave energy fluxes on vertical atmospheric motion and associa precipitation processes. Further, seve investigators have used surface albedo in surf energy balance estimation to solve for latent h flux or evapotranspiration as a residual (Ref. 2).

In this paper, initially, a solar band estimation model from satellite narrow band derived reflectance is given and discussed. Then specific processes and problems related to solar, surface albedo are addressed: 1) the conversion of narrow band derived reflectance to solar band albedo; 2) the effects and dynamics of the spectral composition of the incident solar radiation; 3) the canopy and energy interactions from changes to the angular distribution of the incident solar radiation; and 4) the ultraviolet and shortwave middle-infrared derived reflectance contribution to the solar albedo.

2. SOLAR ALBEDO ESTIMATION MODEL

Brest and Goward (Ref. 3) derived a solar, albedo model using narrow band Landsat MSS spectral data. A more general form of the model is given next.

$$\rho_{solar} = k_{vis}*\rho_{vis} + k_{nir}*\rho_{nir} + k_{smir}*\rho_{smir} \quad (1)$$

ρ_{solar} = solar albedo
ρ_{spec} = derived surface reflectance for the spectral regions (visible (0.38-0.72 μm), near-ir (0.72-1.30 μm) and mid-ir (1.3-3.0μm)).
k = proportion of surface, solar irradiance by spectral region

The spectral region derived reflectance is weighted by the average proportion of surface, solar irradiance for the specified spectral region. In this approach, the spectral band derived reflectance representing a satellite bandpass is converted to a broad band of "uniform" spectral reflectance. For vegetation they assumed there are three major spectral regions in the solar band where the spectral reflectance is near uniform. These three spectral regions are visible (0.38-0.72 μm), near-infrared (0.72-1.30 μm), and shortwave middle-infrared (1.3-3.0 μm). Gausman (Ref. 4) indicates these three spectral regions are uniquely characterized in terms of the leaf optical properties of reflectance, absorptance and transmittance.

3. SATELLITE NARROW BAND SENSING

The satellite waveband spectral response function is typically selected over narrow wavelengths in the solar region for inclusion within atmosphere windows and for sensing specific processes such as chlorophyll absorption and mineral type discrimination.

3.1 Satellite Narrow Band to Solar Band Conversion

A high spectral resolution field spectrometer (0.38-1.1μm) was used to collect spectra for simulating satellite waveband (Table 1) derived reflectivities for two surface covers (soybeans and orchard grass) at the Beltsville Agricultural Research Center in Greenbelt, Maryland on October 8, 1986. The surface reflectivity was derived from concurrent reference panel spectra following procedures given in Kimes and Kirchner (Ref. 5). The data were collected between 0900 and 1200 ldst at solar zenith angles ranging between 47° and 70°. The field spectrometer derived spectra with a spectral resolution to 0.01 μm, were radiometrically calibrated and integrated over the satellite wavebands and for the total visible (0.38-0.72μm) and the near-infrared (0.72-1.10μm) spectral regions. The spectral integrations were completed for comparing the satellite simulated reflectivity to the total reflected radiation by major spectral subregion. The objective is to determine if satellite derived narrow band reflectance can be extrapolated within the total visible or total near-ir spectral region.

Linear regression analysis was conducted between the satellite narrow band derived reflectances (the independent variable(s)) and the total spectral region (i.e., visible and near-infrared derived reflectance). For example, a satellite waveband sensing in the visible such as band 1 to NOAA AVHRR (0.570-0.699 μm) is compared in a linear regression model to the derived reflectance for the total visible reflectance (0.38-0.72 μm). Analyses of the correlation coefficient indicates the utility of the satellite system to estimate a broad band reflectance for the visible, near-infrared, or combined visible/near-infrared spectral region. A slope coefficient not equal to one indicates that a linear transformation may be required before the satellite spectral reflectance may be used to directly represent a spectral region reflectance.

Examination of Table 1 indicates all sensors with wavebands in both the visible and near-infrared spectral regions (i.e., MSS, Thematic Mapper (TM), and Advanced Very High Resolution Radiometer (AVHRR)) significantly represented the surface reflectance for the total visible and near-infrared radiation. This generally supports several studies indicating a high relationship of reflectance within both the visible and near-infrared regions. The visible derived reflectance of each satellite bandpass is highly correlated with the total visible reflectance by at least 0.964 r^2 for soybeans and 0.920 r^2 for orchard grass (Table 1). There is however a range in value of the linear slope coefficients for soybeans, ranging from 0.899 (the closer to one the better is the representation of the broad spectral region without incorporating a correction) for Landsat MSS-1 to 0.761 for NOAA AVHRR-1, and for orchard grass ranging, from 1.03 for MSS-1 to 0.705 for MSS-2 (Table 1). Similarly, for near-infrared reflectivities, correlations are generally high for all sensors with bands sensing within the near-infrared. For soybeans the correlation coefficient is 0.983 r^2 and higher, and for orchard grass the correlation is 0.980 r^2 and higher. However, the range in linear slope coefficients is higher for near-infrared radiation than for visible radiation, ranging from 0.909 for MSS-4 to 0.681 for TM-1. The high correlation but changing slope coefficients indicates the satellite wavebands may be used to estimate a solar albedo, however, a transformation of the satellite reflectance is probably needed. Since there is a difference between linear slopes for soybeans and orchard grass, analysis of results indicate a simple empirical model may not work for a range of surface feature conditions and hence future work should study a more robust procedure.

Table 1. Regression analysis statistics of the multiple correlation coefficient (R^2), intercept (b_0) with standard error, and slope coefficient (b_1) with standard error for the satellite simulated spectral reflectance (independent variable) to the associated spectral region (visible, near-infrared or total 0.38-2.3 μm).

	Soybeans			Orchard Grass		
	R^2	b_0(s.e.)	b_1(s.e.)	R^2	b_0(s.e.)	b_1(s.e.)
AVHRR1-VIS	.964	.008(.002)	.761(.031)	.865	.003(.004)	.791(.069)
AVHRR2-NIR	.983	-.008(.008)	.748(.020)	.994	.005(.003)	.743(.012)
AVHRR12-TOT	.988	-	-	.982	-	-
VISSR-VIS	.967	.008(.002)	.776(.030)	.939	.000(.002)	.798(.045)
VISSR-TOT	.527	.062(.004)	1.423(.287)	.696	.008(.014)	1.562(.230)
MSR-VIS	.613	.010(.008)	.298(.051)	.830	.007(.004)	.382(.038)
MSR-NIR	.970	-.009(.011)	1.847(.069)	.978	-.010(.006)	1.762(.057)
MSR-TOT	.999	.000(.007)	.945(.004)	.999	.000(.001)	.951(.006)
MSS1-VIS	.948	.008(.003)	.899(.044)	.983	.000(.001)	1.032(.030)
MSS2-VIS	.965	.007(.002)	.755(.031)	.900	.001(.003)	.705(.052)
MSS3-NIR	.920	.001(.018)	.720(.045)	.9[illegible]	.011(.005)	.768(.024)
MSS4-NIR	.995	.002(.004)	.932(.012)	.998	-.000(.001)	.909(.007)
MSS1/4-TOT	.997	-	-	.993	-	-
TM1-VIS	.830	.008(.004)	1.401(.135)	.966	.001(.002)	1.436(.059)
TM2-VIS	.946	.010(.002)	.859(.043)	.971	.000(.001)	.981(.037)
TM3-VIS	.964	.008(.002)	893(.036)	.961	.000(.002)	.938(.041)
TM4-NIR	.987	-.007(.007)	.697(.016)	.989	.007(.004)	.681(.015)
TM1/4-TOT	.991	-	-	.980	-	-

4. SOLAR RADIATION SUPPLY EFFECTS

To study the spectral composition and angular distribution of the solar radiation at the surface on the solar albedo the diffuse atmosphere data set of Dave' (Ref. 6) was evaluated. The Dave' data sets from Models 2 , 3 and 4 were analyzed (Table 2). The data set is simulated at 77 wavelengths between 0.31 μm and 2.40 μm.

Table 2 gives the percentage of solar irradiance incident at the surface in terms of diffuse, direct and global (direct + diffuse) radiation integrated over the four major spectral regions for Models 2, 3 and 4 at selected Sun zenith angles. The global radiation percentages are the spectral region solar irradiance coefficients (i.e., k) in Equation (1). Both an increasing Sun zenith angle (i.e., 0 to 70 θ_s) and an increasing aerosol loading (Model 2 to 3 to 4), contribute to increased scattering, producing less direct radiation and more diffuse radiation. For direct radiation there is proportionally less energy in the ultraviolet and visible regions with a shift of energy to the longer near-ir and middle-ir wavelengths. As can be seen by the small changes in the global radiation proportions between illumination and model conditions, there is nearly a complete compensating effect of an increased diffuse radiation when there is reduced direct radiation. Hence, analysis of results indicates a single average set of spectral proportions for the global radiation may be used in Equation (1) with only a small introduced error over many clear sky and illumination conditions. In addition, Table 2 shows the ultraviolet and shortwave middle-ir contribute to approximately 5% and 10% of the solar albedo. Table 2 also includes the percent diffuse radiation to global radiation. For increased scattering either from a higher Sun zenith angle or additional aerosols there is an increase in the percentage of diffuse radiation.

Table 2. Proportion of spectral Solar irradiance at surface by spectral region. Also given is percent diffuse to global. Analysis of Dave' (1978) atmosphere diffuse data set. Model 2 Gases with no aerosols. Model 3 gases plus low loading (τ'_a=0.09 @ .5 μm). Model 4 gases plus moderate aerosol loading (τ'_a=0.45 @ .5 μm).

	UV(0.31-0.38μm)	VS(.38-.72μm)	NR(.72-1.3μm)	MR(1.3-2.41μm)
Exoatmosphere	.063	.432	.350	.155

Sun Zenith Angle

	0°				45°				70°			
Model 2	UV	VS	NR	MR	UV	VS	NR	MR	UV	VS	NR	MR
Direct	.039	.471	.385	.105	.031	.468	.395	.106	.014	.474	.431	.110
Diffuse	.313	.635	.051	.001	.288	.658	.054	.001	.220	.711	.850	.001
Global	.053	.479	.368	.100	.048	.480	.373	.098	.039	.460	.387	.097
Diff(%)	29.6	6.6	0.7	0.1	39.8	9.2	1.0	0.1	67.6	18.0	2.4	0.2
Model 3												
Direct	.038	.466	.386	.110	.031	.460	.397	.112	.014	.428	.434	.124
Diffuse	.154	.568	.242	.034	.142	.579	.245	.033	.111	.600	.257	.033
Global	.052	.478	.370	.101	.048	.478	.374	.099	.040	.474	.386	.099
Diff(%)	34.4	13.8	7.6	3.9	45.6	18.7	10.1	5.1	74.3	33.8	17.8	8.8
Model 4												
Direct	.037	.444	.390	.129	.029	.429	.402	.140	.012	.360	.434	.194
Diffuse	.077	.528	.340	.056	.070	.528	.345	.056	.058	.521	.359	.062
Global	.050	.472	.373	.104	.046	.471	.378	.105	.042	.464	.385	.108
Diff(%)	50.8	37.4	30.5	17.8	63.9	47.5	38.6	22.7	89.9	72.8	60.4	37.1

4.1 Diffuse vs. Direct Radiation Effects

Analysis of the diffuse versus direct radiation effects on solar albedo was studied using canopy reflectance simulation modeling. The extended Suits model, SAIL (Ref. 7), with the addition of a complete leaf angular distribution was used. The SAIL model has shown to be in reasonable accuracy against "in-house", field collected bidirectional reflectance data. Results may be determined at any relative azimuth and view angle for a specific Sun zenith angle. For comparison between conditions the view zenith and relative azimuth angles of 60° and 90° as recommended by Kimes and Sellers (Ref. 8) were used. These angles were found to represent a hemispherical albedo for a wide range of conditions to within 11%, working optimally between 30° and 45° Sun zenith angles. The canopy optical properties (Ref. 4) were estimated from the average green leaf optical properties averaged over the ultraviolet, visible, near-infrared and middle-infrared wavelengths. A soil reflectance was estimated for a moist clay soil at 35-40% soil moisture. The percent diffuse radiation was estimated using the data in the Dave' Models 2, 3 and 4 for the four major spectral regions.

Figures 1-3 illustrate the change in solar albedo as a function of Sun zenith angle (θ_S) for each of the three Dave' models. Albedos are given for the common spherical (Fig. 1) and erectophile (Fig. 2) leaf angular distributions (LAD) and for contrast results are given for the planophile LAD (Fig. 3), each at a leaf area index of three. At low to moderate Sun zenith angles (to 60° θ_S) the solar albedo increases with increased diffuse radiation (Model-4). The increased albedo is from a decrease in the amount of direct radiation incident at the canopy relative to diffuse radiation. The canopy reflectance to the direct beam radiation is lowered because of the increased penetration to the less reflective, moist soil. In the canopy modeling, the diffuse radiation incident at the canopy is isotropically distributed and has a lower probability of interacting with the soil and hence has a higher reflectance, similar to the leaf reflectance. Hence, as the probability decreases for solar radiation to interact with soil, the canopy reflectance will be higher from the increased leaf reflectance and transmittance. All three figures illustrate a reversal in solar albedo between atmospheric models at high Sun zenith angles (i.e., higher than 65° θ_S). At high Sun zenith angles little direct solar radiation is reaching the soils where there is a lower reflectance than the leaves. However, the scattered diffuse radiation is isotropically incident at the canopy top (as input to canopy model), having an increased likelihood for interaction with the soils. Hence, at high Sun zenith angles, Model-4 with increased aerosol scattering has a lower reflectance. The erectophile derived canopy albedo (Fig. 2) has a greater probability for light to penetrate to the canopy bottom and therefore is shown to be the most sensitive to atmospheric condition. On the other side, the planophile leaf angle distribution permits the least amount of light to penetrate the understory and hence the solar albedo is least sensitive to atmospheric and illumination changes.

4.2 Shortwave SMIR and Ultraviolet Reflectance Effects

Papers reporting on solar albedos previously, have emphasized the incorporation of using only a visible and near-ir reflectance without the estimation of a ultraviolet or middle-ir reflectance (Ref. 9). However, research has indicated the unique contributions of middle-ir reflectance in comparison to visible and near-ir reflectance (Ref. 4). Middle-ir is affected both by liquid water absorption and scattering from the leaf's internal structure.

Table 3 gives the solar albedo for different spectral region combinations. Solar albedos are compared as a function of Sun zenith angle, LAD, leaf area index (LAI), single green leaf, bare soil, and atmospheric aerosol scattering (Dave' Models 2, 3, and 4). Inspection of Table 3 indicates a substantial difference of over 20%, when using the 50% visible versus near-infrared approach in comparison to using all four spectral regions weighted by the proportions given in Table 2. Differences from not including an ultraviolet reflectance are small. However, the middle-ir contribution is significant with a 15.4% difference for the base-line parameterization. This is 5% more than for the single leaf derived solar albedo difference (Table 3) and represents 5% more than the total supply of middle-ir solar radiation (Table 2). The larger difference may be attributed to the low transmittance (i.e., and high leaf absorptance) relative to near-ir yielding a lowering of the middle-ir canopy reflectance with increasing vegetation. Table 3 gives data indicating the large differences in albedo without a middle-ir as a function of increasing LAI, changing from 7.1% at a 0.5 LAI to 16.7% at a 7 LAI.

5. SUMMARY

A solar band estimation model was discussed based on a near uniform spectral reflectance of vegetation for four major spectral regions (ultraviolet, visible, near-infrared, and shortwave middle-infrared). Analysis of field derived spectra indicates a narrow-band visible and a near-infrared derived reflectance may be used to represent approximately 85% of the total shortwave solar radiation. However, simply using the satellite derived reflectance to represent a broad visible or near-infrared spectral region may introduce differences to 30%. In these situations a simple linear transformation of the satellite derived narrow band reflectance is required to correspond to a broad spectral region reflectance for solar albedo estimation.

Table 3. Comparison of canopy, leaf and soil solar albedo estimates with a relative azimuth of 90° and a view-angle of 60°. Base-line spherical LAD, Dave' Model-4, LAI=3, and Sun Zenith (θ_s) at 45°. Data in parentheses are the percent difference from #1 [ABS(#(2,3 or 4)-#1)/#1].

Column Summary

#1. All 4 major spectral regions weighted by the relative proportion of the surface, global radiation.
#2. Visible and near-ir weighted at 50% each.
#3. No ultraviolet, represented by visible. Wtgs. as in #1.
#4. No middle-ir, represented by near-ir. Wtgs. as in #1.

SOLAR SURFACE ALBEDO

PARAMETERIZATION	#1	#2	#3	#4
Single leaf	25.56	28.5(11.5)	25.6(0.2)	27.4(7.4)
Bare soil	15.48	16.0(3.4)	15.6(1.2)	15.7(1.9)
Base-line (θ_s=45°, LAI=3, Dave'-4, spherical)	21.76	25.9(19.1)	21.8(0.3)	28.9(14.9)
Base-line changes:				
LAD-Planophile	22.51	26.9(19.7)	22.5(0.2)	25.6(13.8)
LAD-Erectophile	23.22	27.4(18.2)	23.2(0.3)	26.6(14.6)
θ_s-0°	19.63	23.6(20.5)	19.6(0.0)	22.6(15.5)
θ_s-75°	25.12	29.1(16.0)	25.2(0.3)	28.9(14.9)
Dave'-2	20.10	24.2(20.7)	20.1(0.0)	23.0(14.4)
Dave'-3	20.48	24.6(20.4)	20.5(0.3)	23.4(14.6)
LAI-0.5	16.28	17.9(10.0)	16.4(0.8)	17.4 (7.1)
LAI-1	17.58	20.1(14.1)	17.6(0.6)	19.5(10.8)
LAI-7	24.00	28.9(20.4)	24.1(0.2)	28.0(16.7)

The supply of global radiation by spectral region as a function of atmospheric scattering and Sun zenith angle was shown to be relatively insensitive with the diffuse and direct radiation compensating for one another. However, although the global radiation was relatively constant, the changes to the proportions of direct and diffuse radiation, changing with illumination angle and aerosol loading, were shown to substantially change the solar albedo. The erectophile LAD was shown to be the most sensitive to atmospheric condition with the planophile LAD the least sensitive.

The shortwave middle-infrared reflectance was indicated to be a significant part of the solar albedo, with differences to 16% when using a near-infrared reflectance to represent a total shortwave reflectance. The ultraviolet reflectance was less sensitive with errors introduced at approximately 0.3%. Conventional solar albedo estimation using a 50% visible to near-ir reflectance weighting may lead to errors up to 25% in comparison to using all four major spectral region reflectances weighted by the spectral, global radiation incident at the surface.

6. REFERENCES

1. Sud, Y.C. and M. Fennessy, 1982, A study of the influence of surface albedo on july circulation in semi-arid regions using the glas GCM, Journal of Climatology, 2:105-125.

2. Taconet, O., R. Bernard, and D. Vidal-Madjar, 1986, Evapotranspiration over an agricultural region using a surface flux/temperature model based on NOAA-AVHRR data, J. Climate Appl. Meteorol., 25:284-307.

3. Brest, C. and S. Goward, 1987, Deriving surface albedo measurements from from narrow band satellite data, Int. J. Remote Sensing. in press.

4. Gausman, H., 1985, Plant leaf optical properties in visible and near-infrared light, Grad. Studies, No. 29, Texas Tech, Lubbock, TX.

5. Kimes, D. and J. Kirchner, 1982, Irradiance measurement errors due to the assumption of a Lambertian reference panel, Remote Sensing Environ., 12:141-149.

6. Dave, J., 1978, Extensive datasets of the diffuse radiation in realistic atmospheric models with aerosols and common absorbing gases, Solar Energy, 21:361-369.

7. Verhoef, W., 1984, Light scattering by leaf layers with application to canopy reflectance modeling, the SAIL model, Remote Sens. Environ., 16:125-141.

8. Kimes, D.S. and P.J. Sellers, 1985, Inferring hemispherical reflectance of the earth's surface for global energy budgets from remotely sensed nadir or directional radiance values, Remote Sens.Environ., 18:205-223.

9. Dickinson, R., 1983, Land surface processes and climate-surface energy balance, Adv. Geophys., 25:305-353.

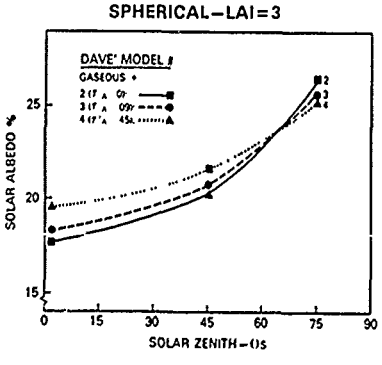


Figure 1.

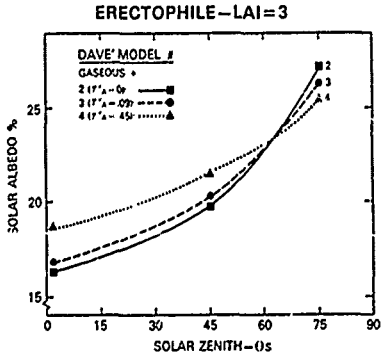


Figure 2.

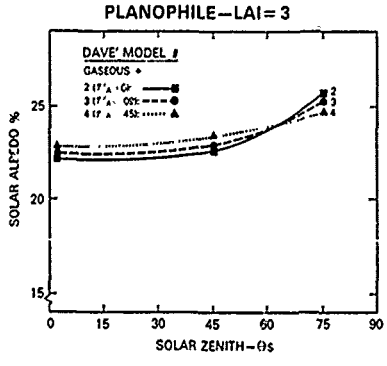


Figure 3.

Figures 1, 2 and 3. Solar albedo versus Sun zenith angle for three atmospheric-solar radiation Models of Dave' (1978). Albedos were derived using canopy modeling analyses over three leaf angle distributions.

CLIMATE CHANGE EFFECTS ON THE SNOWMELT HYDROLOGY OF WESTERN NORTH AMERICAN MOUNTAIN BASINS

A. Rango and V. van Katwijk

USDA-ARS Hydrology Laboratory
Beltsville Agricultural Research Center
Beltsville, MD 20705

ABSTRACT

Several effects of increasing atmospheric CO_2 content on climate are considered in this study, namely, increasing atmospheric temperature, changing snow water equivalent at the beginning of snowmelt, and changing precipitation during the snowmelt runoff season. The effects that these changes would have on runoff from mountain basins were evaluated by use of the Snowmelt-Runoff Model (SRM). The major effect on runoff was caused by increasing temperature. An increase in total seasonal runoff resulted, but the more interesting effect was a redistribution of runoff to the months April and May. Depending on whether snowmelt season precipitation and winter snow accumulation increased or decreased, the effects on the hydrograph could be significant. The model results indicated a potentially serious problem in western North America where climate change could widen the gap between water supply and water demand as well as causing existing water storage and distribution systems to be ineffective.

Keywords: Climate Change, Greenhouse Effect, Snowmelt Runoff, Hydrological Modelling, Mountain Basins, Water Supply

1. INTRODUCTION

As early evidence of the greenhouse effect starts to accumulate, there is a definite need to project what the long term consequences of such a major change would be on the hydrological cycle. An increase in global CO_2 in the atmosphere has become well accepted with most projections indicating that major effects of the increase will be felt in the middle part of the next century. It has been widely speculated that there will be a major climate change in response to the atmospheric CO_2 increase. There is general agreement that atmospheric temperature will rise dramatically in response to a doubling of atmospheric CO_2 content. Citations from the literature indicate a rise generally between 1.0°C and 5.0°C [1] and [2], although recently even a 6.0°C increase has been mentioned [3].

There is also agreement that precipitation amounts will be affected, but the amount and direction of change is more uncertain than the temperature response. Increases or decreases in precipitation are projected depending upon locality and the type of method being used to estimate the change. Both general circulation models and development of climate scenarios have been used to estimate precipitation response [1] and [2]. The effect on precipitation is in doubt even for a specific location, however, the most commonly cited changes range between a 10% decrease and a 10% increase.

2. PROCEDURE

Any of the climate changes in response to a CO_2 increase that have been mentioned should have an effect on hydrological processes. In mountainous basins where the accumulation and melt of seasonal snow cover is important for runoff production, changes in temperature and precipitation regimes could be particularly important. Because controlled experiments to evaluate such changes are impossible and accumulation of actual data on such changes will take many years, the most effective way to evaluate the hydrological effects of climate change is through well organized model experiments. A simple snowmelt-runoff model (SRM) used for simulating and forecasting flow from mountain basins during the snowmelt season [4] was adapted to allow evaluation of the simulated effects of climate change. In SRM, there are only three major input variables, and all three will respond to climate change. These variables are temperature, precipitation, and snow-covered area. Because the three major input variables are physically interrelated, simulating the end effect of the climate change is complex. For the moment the model parameters of SRM will be assumed to be unaffected by the climate change. These parameters are the degree-day coefficient, snowmelt runoff coefficient, rainfall runoff coefficient, temperature lapse rate, critical temperature (dividing rain and snow events), time of snowpack ripening, lag time, recession coefficient, and the basin area. Future studies will be directed at determining how climate change may affect the magnitude and timing of some of these parameters.

The adapted SRM was tested on two mountain snowpack basins in the western part of North America. The Rio Grande basin is located in the Rocky Mountains of southern Colorado and has an area of 3419 km^2. The Illecillewaet basin is located in the Canadian Rocky Mountains of British Columbia (1155 km^2). In both basins snowmelt contributes the major portion of streamflow. In the Rio Grande basin the average annual precipitation is 700 mm and the average annual runoff depth is 220 mm. The Illecillewaet basin has

an average annual precipitation of 1890 mm and an average annual runoff depth of 1455 mm. The Canadian basin is considerably more humid than the U.S. study basin.

Three climate changes were attempted on each of the basins to simulate the possible effects of a CO_2 increase in two years, 1983 and 1984. These changes were first run separately and then in combination. For temperature, a progressive warming of 1°, 2°, 3°, 4°, and 5°C was input to the model and a snowmelt runoff hydrograph was generated. Subsequently, changes in snowmelt season precipitation of -10% and +10% were input to SRM. The final climate change simulation was a change in winter snowpack accumulation. The April 1 snow water equivalent was changed by -10% and +10% and the effect on snow-covered area evaluated. After separate computer runs were made for each changed variable, various combinations of changes were input to SRM, and the resulting snowmelt-runoff hydrographs were evaluated.

3. RESULTS AND DISCUSSION

The major effect on the snowmelt runoff hydrograph of the two Rocky Mountain basins seems to result from an increase in temperature during the snowmelt season. The greater the temperature increase, the greater the runoff response. A major redistribution of runoff results in a concentration of increased flows in April and May coupled with a decrease in flow in the months at the end of the snowmelt season. This effect was noted by Martinec and Rango [5]. Because of this redistribution there is also an overall increase in the total snowmelt season runoff.

As one would expect, if the climate change has an effect on winter snow accumulation, a resulting effect will be felt during the snowmelt season. For example, on the Rio Grande basin a 10% increase in April 1 snow water equivalent will result in about a 6% increase in snowmelt season runoff. A similar decrease in runoff during the snowmelt season will result from a 10% decrease in April 1 snow water equivalent. The runoff response to the same changes on the Illecillewaet basin is not as pronounced as on the Rio Grande. Unlike the results of the temperature increase, however, the bulk of the increase or decrease in snowmelt season runoff occurs in June or July as a result of a change in winter snow accumulation.

Finally, an increase or decrease in snowmelt season precipitation has the least effect on the snowmelt hydrograph. A 10% change in the April-September precipitation only has a net effect of a 1-3% percent change in runoff. The effect on runoff can be expected to increase on basins where a significant amount of precipitation falls during snowmelt.

When more than one of the climate variables in SRM are changed at the same time, the results on runoff are magnified or diminished depending on the specific climate change in a particular area. For example, on the Rio Grande basin an increase in temperature, winter snow accumulation, and summer precipitation has the most drastic effect on snowmelt season runoff. Most model calculations and scenarios, however, call for a temperature increase and some sort of precipitation decrease. As an example of what can be expected in this situation on the Rio Grande basin, an annual temperature increase of 3°C was assumed along with a snow water equivalent decrease of 10% on April 1 and a 10% decrease in snowmelt season precipitation. The end result revealed a redistribution of the runoff to the months of April and May in response to the temperature increase combined with an overall decrease in seasonal runoff volume of about 2%. Figure 1 compares the 1983 simulated hydrograph using actual data and the simulated hydrograph resulting from the climate change outlined above.

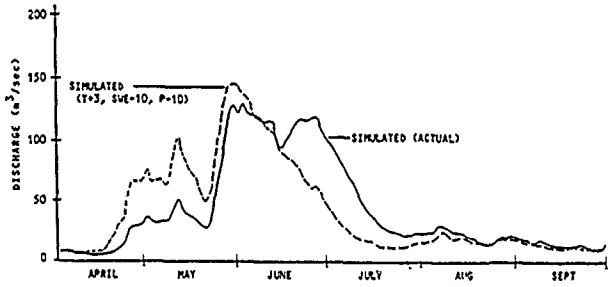


Fig. 1. Simulated change in runoff on the Rio Grande basin for 1983 as a result of a 3°C increase in temperature, a 10% decrease in April snow water equivalent, and a 10% decrease in snowmelt season precipitation (T+3, SWE-10, P-10).

The type of change shown in Figure 1 could have drastic side effects for water supply considerations in western North America. Water storage and distribution systems may be stressed considerably and not be able to handle the increased early season flows. Further, the decrease in seasonal volume will be very significant in most regions of western North America where total water supply is not even adequate today. The increasing demands for water are not synchronized with these decreasing water supplies that could result from climate change. A further intensification of the problem may be revealed as further experiments with the model are conducted. Model parameters, up to now, have remained unchanged as the model variables of temperature, precipitation, and snow cover have been perturbed. It is very likely that as temperature increases, and snow cover leaves the basin earlier, evapotranspiration losses will increase, and the time sequence of parameter values will be shifted. This shift will most likely cause an additional decrease of the runoff volume. Further experiments will take this effect into account.

Depending on the specific geographical location, an amazing range of effects on runoff can apparently take place. The major effect results from the temperature increase, but the effect of changed precipitation can magnify or diminish the runoff response to temperature to the point where there will be serious consequences for water storage and distribution systems. The results of this study show that there may be a need to construct new facilities or modify existing ones so that the growing gap between water demand and water supply does not become a disaster.

4. CONCLUSIONS

Although the effects on climate of increasing atmospheric CO_2 and the greenhouse effect are not firmly established, early indications are conclusive enough to attempt to evaluate the hydrological consequences of the climate change. Simulations produced by the Snowmelt-Runoff Model (SRM) on two snowmelt mountain basins in western North America indicate

that significant changes in the normal hydrograph may occur. Increasing temperature will change the timing of flow causing major increases in April and May runoff. Precipitation response to the changing climate when added to the temperature response can't change the effect of redistribution of flow, but it can magnify or diminish the effect of temperature. Depending on the geographical location of a basin, snowmelt season precipitation and winter season snow accumulation can increase, decrease, or remain unchanged. The most likely result is a decrease in both precipitation and April 1 snow water equivalent. The resulting hydrograph shows the possibility that there may be a significant widening between water demand and water supply. This results from both a decrease in flow as well as an occurrence of the maximum flow at a time when the water supply is not needed and when water storage systems typically have less storage capacity. If these effects are widespread, there may be a need for major expenditures to develop new water supply systems to upgrade existing facilities.

5. ACKNOWLEDGEMENTS

R. Roberts of the USDA, Hydrology Laboratory in Beltsville, Maryland did the programming required for both SRM and the algorithm required to change temperature, precipitation, and snow cover for the modified computer simulations.

6. REFERENCES

1. Gutowski, W. J. et al., "Surface energy balance of three general circulation models: Current climate and response to increasing atmospheric CO_2", Report to the U.S. Department of Energy, TR042, DOE/ER/60422-H1, Atmospheric and Environmental Research, Inc., Cambridge, MA, 119 pp., 1988.

2. Wigley, T. M. L. et al., "Scenario for a warm, high CO_2 world", Nature 83:17-21, 1980.

3. Schneider, S. H., "The greenhouse effect: science and policy", Science 243:771-781, 1989.

4. Martinec, J. et al., "The Snowmelt-Runoff Model (SRM) User's Manual", NASA Reference Publication 1100, 111 pp., Washington, D.C., 1983.

5. Martinec, J. and A. Rango, "Effects of climate change on snowmelt runoff patterns," Remote Sensing and Large-Scale Global Processes (Proceedings of the IAHS Third Int. Assembly, Baltimore, MD, May 1989), IAHS Publication No. 186, 1989.

USING REMOTELY SENSED DATA TO MONITOR LAND SURFACE CLIMATOLOGY VARIATIONS IN A SEMI-ARID GRASSLAND

Lee F. Johnson*, Nevin A. Bryant*, Anthony J. Brazel**, Charles F. Hutchinson***, and Robert C. Balling**

*Jet Propulsion Lab, California Institute of Technology, 4800 Oak Grove Dr., Pasadena CA, 91109, USA
**Arizona State University / Laboratory of Climatology, Tempe AZ, 85287, USA
***University of Arizona / Arizona Remote Sensing Center, Tucson AZ, 85719-4896, USA

ABSTRACT

Long term overgrazing in Mexico has caused a sharp discontinuity in surface properties along the international border in the semi-arid Sonoran Desert. The United States side, protected from overgrazing by the Taylor Act since 1934, exhibits longer, more plentiful grasses, less bare soil, and lower albedo than adjoining Mexican lands. The difference in surface properties affects evapotranspiration rates and evokes a temperature gradient between the two countries. The more exposed Mexican landscape tends to dry more rapidly than the United States following summer convective precipitation. Depletion of soil moisture causes higher surface and air temperatures in Mexico. This study used several satellite datasets, along with supporting ground and aircraft data, to detect and monitor differences between the two semi-arid grassland regimes.

KEY WORDS: remote sensing, hydrologic cycling, rangeland management, desertification

INTRODUCTION

Approximately 35 percent of the earth's land surface is characterized as arid to semi-arid, based either on vegetation or climate. In these regions, surface moisture and thermal energy are in a delicate balance, and the low amount of vegetative cover available for thermal inertia contributes to wide diurnal temperature fluctuations and dramatic changes in evaporative cooling after precipitation events. Temperature changes over short time periods probably have a large-scale forcing effect on regional energy budgets. As such, measurement and monitoring of surface temperatures and hydrologic cycling in arid and semi-arid zones is useful for scientific investigation of desertification processes. In the past, the large areal scale and sporadic weather episodes associated with these regions has made assessment of causal relationships difficult. The growing archive of satellite data, with its timely and synoptic coverage, promises to alleviate this difficulty.

Experiments designed to measure arid land surface variables and their possible effect on climate are complicated by the inability to establish "matched pairs" of study sites (Reining, 1978). Ideally, the site pairs would be located on either side of a boundary marking profound differences in land surface conditions, usually resulting from differing land management practices, in otherwise comparable environments. The border between the United States and Mexico in the Sonoran Desert presents such a boundary. The border has been marked by a fence since 1920, and the U.S. side has been protected from overgrazing since the Taylor Grazing Act of 1934. Over time, marked differences in land surface parameters have evolved in response to differing land management practices (Bahre & Bradbury, 1978). Bryant *et al.* (1985) estimate that Mexico currently has approximately 5% higher albedo and 29% more bare soil than adjoining United States lands.

A border swath extending from Nogales to Lochiel (approximately 40 km to the east) was chosen for investigation. In addition, three paired sites along the swath were selected for intensive in-situ analysis (figure 1). Each site-pair was ecologically distinct and accessible. Finally, a network of weather stations throughout the region was identified for analysis of long-term temperature and precipitation trends.

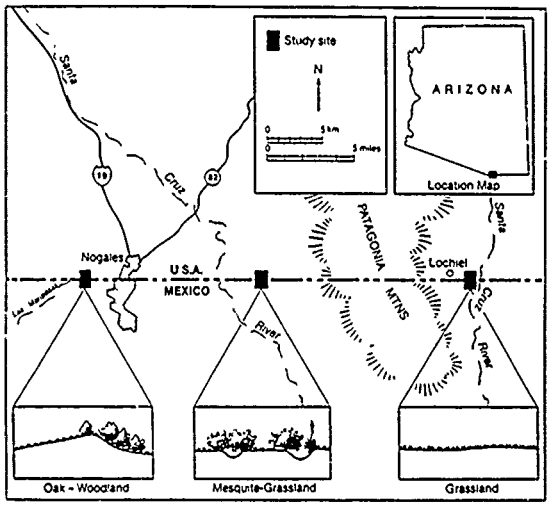


Figure 1. Study area map.

SUPPORTING INVESTIGATION

Balling (1988) analyzed archival records from the weather station network, and found summertime daily maximum air temperatures significantly warmer in Mexico by about 2.5 degrees (C) after accounting for elevation and latitude. Controlling only for elevation, the difference increased to 4 degrees. Evidence of warming in arid regions experiencing decreased vegetative cover agrees with classical agricultural meteorology and rangeland management literature (Dregne, 1983, Penman, 1956). No significant difference in precipitation amount was found between the two countries.

Balling (1989) assessed the affect of available soil moisture on a temperature gradient across the border. During the drying period following heavy precipitation across the network, air temperatures in Mexico were cooler than in the U.S. for the first few days, and thereafter warmer than the U.S.

Intensive field study was conducted during summer 1988 at the Lochiel site pair. Ground surface temperatures were found to be initially lower in Mexico than in the U.S. following rainfall. In subsequent days however, temperatures were higher in Mexico than the U.S. readings. Additional measurements show a more rapid depletion of soil moisture in Mexico (figure 2).

Field measurements reveal equal or higher ET rates in Mexico initially after rainfall. This, combined with higher albedo and greater measured subsurface heat flux evoke lower surface temperatures in Mexico immediately after rainfall. The ET rate in the United States is substantially higher than in Mexico beginning about 3 days into the drying cycle. This

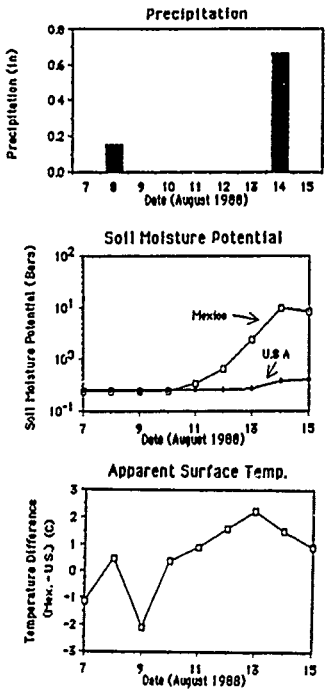


Figure 2. In-situ measurements at Lochiel, August 1988.

difference in ET, sustained over a period of several days, is associated with cooler air and surface temperatures in the U.S.

Two low-altitude aircraft flights were conducted along the border swath during summer 1988. Airborne thermal-infrared measurements on the first flight, which occurred one day after rainfall, showed Mexico on average cooler than the U.S. by one degree (C). Measurement on the next flight, which occurred five days after rainfall, showed Mexico warmer than the U.S. by about 1.25 degrees.

It is argued that changes in land surface parameters, by altering local and regional surface radiation budgets, are of sufficient magnitude to have a significant feedback on the prevailing climate. Many investigators have shown that spatial discontinuities in surface conditions may affect the rate of local convection in summer months. Miller *et al.* (submitted) assessed the impact of the Sonoran border discontinuity on summertime convective cloud development. Analog early- and late-afternoon Geostationary Operational Environmental Satellite (GOES) images were digitized and convective cloud formations were identified. Each image pair was compared quantitatively to calculate the rate of convective cloud formation. A higher growth rate (from 50% to 177%) was found over Mexico. It appears that higher surface and air temperatures of the overgrazed Mexican landscape may increase atmospheric

instability and thereby support a higher growth rate of afternoon cumulus than on the U.S. side of the border. These findings are consistent with other investigations (Mahfouf & Mascart, 1987, Anthes, 1984) which have shown that spatial discontinuities in surface conditions may affect the local convection rate in summer months.

SATELLITE DATA PROCESSING

The temperature and albedo gradient along the border are evident in the thermal and visible/near-infrared channels, respectively, of datasets from a variety of satellite-based instruments. Among these are the Heat Capacity Mapping Radiometer (NASA, 1978), the LANDSAT Thematic Mapper (NASA, 1978), and the Advanced Very High Resolution Radiometer (AVHRR) (Kidwell, 1981). Thus a large and diverse data archive exists with which to examine this and similar phenomena.

Thermal channel AVHRR data were used to observe the progression of the post-rainfall temperature gradient along the border swath and at the Lochiel site. The AVHRR is a polar orbiting satellite which provides daily thermal data in two channels (10-11 microns, and 11-12 microns) at a 1.1 kilometer spatial resolution. Sequential AVHRR datasets were acquired for the drying period following several of the rainfall events identified by Balling (1989). The data were calibrated (Lauriston *et al.*, 1979), subjected to the split-window atmospheric correction (Strong & McClain, 1984), corrected for emissivity (Price, 1984), and geo-registered. Estimated surface temperatures in Mexico were found to be consistently and significantly warmer than the U.S. beginning 4 days after rainfall and continuing for at least ten days. Prevailing cloudy conditions and other data acquisition problems prevented analysis of satellite data during the first few days of the drying cycle after any major event.

Data from the Scanning Multichannel Microwave Radiometer (SMMR) (Njoku *et al.*, 1980) were processed to estimate soil moisture in broad swaths (ie: 0.5 degrees x 2 degrees) on either side of the border. The SMMR is a five frequency, dual polarized radiometer with spatial resolution ranging from 30 km at 37 GHz to 150 km at 6.6 GHz. The SMMR TCT product, binned in .5 degree latitude by .5 degree longitude cells, was used to construct polarization ratios (PR) at 6.6GHz. The 6.6 GHz PR [(V-H)/V+H)] has been shown to be sensitive to soil moisture, particularly in areas of sparse vegetation (Kerr & Njoku, 1988). The PR in the border region varied with weekly precipitation for several rainfall events in 1982 (eg: figure 3).

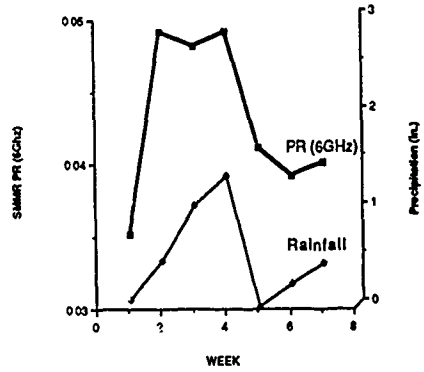


Figure 3. Comparison of SMMR polarization ratios with weekly rainfall totals; November - December 1982; USA.

It is believed that differences in PR's from several SMMR channels may be convolved to yield large-area model estimates of soil moisture content, surface roughness, and integrated atmospheric water content (Kerr, 1988). Such a model applied to TCT data for drying episodes in July and December 1982 showed a more rapid soil moisture depletion in Mexico.

Data from the LANDSAT Multi-spectral Scanner (MSS) were used to formulate the Normalized Difference Vegetation Index (NDVI) (Tucker, 1979). LANDSAT NDVI, [(CH_4 - CH_2)/ CH_4 + CH_2)], is a ratio formed with the red and near-infrared MSS channels. The NDVI is a greenness indicator which has been described by many investigators as an indicator of vegetative quantity and vigor, particularly in agricultural or other heavily vegetated regions. NDVI for the Nogales site pair was calculated for four years ('79, '82, '86, '87) at the annual vegetative maximum (autumn) and compared with the 45-day antecedent precipitation recorded at the nearby Nogales weather station. NDVI on each side was highly correlated (.98) with the amount of antecedent precipitation. Also, the NDVI was consistently and significantly higher on the U.S. side than in Mexico (figure 4).

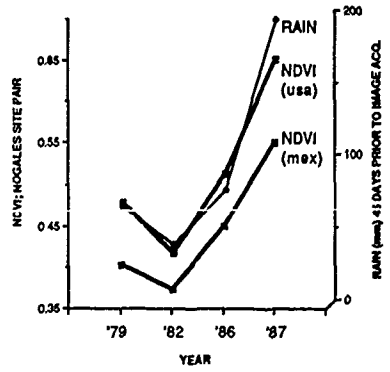


Figure 4. Comparison of MSS NDVI at vegetative maximum with rainfall at Nogales station.

Clearly, more observations are needed to better define the relationship between NDVI and cumulative precipitation for semi-arid regions. However, despite a relatively high proportion of exposed soil, the index appears to respond to vegetation cover in an encouraging fashion.

CONCLUSIONS & FUTURE STUDY

Preliminary conclusions from this study reveal that higher surface and air temperatures in Mexico may result from reduced vegetation cover and that latent and sensible heat flux after precipitation events is associated with near-surface soil moisture available for evaporative cooling. Satellite data appear to be a viable means of monitoring land surface variables at short and long wavelengths. Empirical results of this study may be incorporated into a general land surface climatology model of biosphere-atmosphere interactions similar to that of Verstraete (1988). Input of remotely sensed data may provide these models with information of finer spatial and temporal resolution than is currently available from existing meteorological sources.

ACKNOWLEDGEMENTS

The authors acknowledge Tom Miller, Mike George, and Matt Stoll of the Arizona State University/Climatology Lab, and Louisa Beck of the University of Arizona/Remote Sensing Center for contributions to the supporting investigation of this paper. The research described in this paper was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

REFERENCES

Anthes, R.A., "Enhancement of Convective Precipitation by Mesoscale Variations in Vegetative Covering in Semi-arid Regions," Journal of Applied Meteorology, 23:541-554, 1984.

Bahre, C.J., & Bradbury, D.E., "Vegetation Change Along the Arizona-Sonora Boundary," Annals of the Association of American Geographers, pp 145-165, 1979.

Balling, R.C., "The Climatic Impact of a Sonoran Vegetation Discontinuity," Climatic Change, vol. 13, no. 1, pp 99-109, 1988.

Balling, R.C., "The Impact of Summer Rainfall on the Temperature Gradient Along the United States-Mexico Border," Journal of Applied Climatology, vol. 28, no. 4, pp 304-308, 1989.

Bryant, N.A., Brazel, A.J., Hutchinson, C.F., "Land Surface Parameter Variability and Climate in Southwestern North America," NASA OSSA / Interdisciplinary Research Program in Earth Science, Proposal, 1985.

Dregne, H.E., Desertification of Arid Lands, Harwood Academic Press, 1983.

Kerr, Y.H., Personal Communication, 1988.

Kerr, Y.H., Njoku, E.G., "Monitoring Seasonal Variations of Soil Moisture and Vegetation Conver Using Satellite Microwave Radiometry," Proceedings 4th International Colloquium on Spectral Signatures of Objects in Remote Sensing, Modane, France, pp 403-406, 1988.

Kidwell, K.B., NOAA Polar Orbiter Data Users Guide, National Climatic Center, Washingto D.C., 1981.

Lauriston, L., Nelson, G.J., Ports, F.W., Data Extraction and Calibration of TIROS-N/NOAA Radiometers, NOAA Technical Memorandum NESS 107, Washington D.C., 1979.

Mahfouf, J.E., Mascart, P., "The Influence of Soil and Vegetation on the Development of Mesoscale Circulations," Journal of Applied Meteorology, 26:1483-1495, 1987.

Miller, T.A., Balling, R.C., Brazel, A.J., Lougeay, R.L., "Differential Cloud Cover Patterns Along the Border of the United States and Mexico," Journal of Applied Meteorology, submitted.

NASA, Heat Capacity Mapping Mission (HCMM) Data Users Handbook, 1978.

NASA, GSFC Specification: Thematic Mapper System and Associated Test Equipment, 400.8-D-210, 1978.

Njoku, E.G., Stacey, J.M., Barath, F.T., "The Seasat SMMR: instrument description and performance," IEEE Transactions on Oceanic Engineering, vol OE-5, pp 100-115, 1980.

Penman, H.L., "Estimating Evaporation," Transactions, American Geophysical Union, 1956.

Price, J.C., "Land Surface Temperature Measurements from the Split Window Channels of the NOAA-7 Advanced Very High Resolution Radiometer," Journal of Geophysical Research, vol. 89, no. D5, pp 7231-7237, 1984.

Reining, P., Handbook on Desertification Indicators, American Association for the Advancement of Science, Washington, D.C. 141pp., 1979.

Strong, A.E., and McClain, E.P., "Improved Ocean Surface Temperatures from Space - Comparisons with Drifting Buoys," Bulletin American Meteorological Society, vol. 65, no. 2, pp 138-142, 1984.

Tucker, C.J., "Red and photographic infrared linear combinations for monitoring vegetation," Remote Sensing of Environment, 8:127-150, 1979.

Verstraete, M.M., "Radiation Transfer in Plant Canopies: Scattering of Solar Radiation and Canopy Reflectance," Journal of Geophysical Research, 93:9483-9494, 1988.

ADAPTATION OF THE ISCCP CLOUD DETECTION ALGORITHM TO COMBINED AVHRR AND SMMR ARCTIC DATA

J. Key and R.G. Barry

Cooperative Institute for Research in Environmental Sciences
University of Colorado, Boulder
Boulder CO 80309-0449 USA

ABSTRACT

Cloud cover and surface temperatures in the polar regions are important indicators of global climate change, however, automated analyses of satellite radiance data have concentrated on low and middle latitude situations. The International Satellite Cloud Climatology Project (ISCCP) cloud detection algorithm is applied to Arctic data, and modifications are suggested. Both Advanced Very High Resolution Radiometer (AVHRR) and Scanning Multichannel Microwave Radiometer (SMMR) data are examined. Synthetic AVHRR and SMMR data are also generated. Modifications suggested include the use of snow and ice data sets for the estimation of surface parameters, additional AVHRR channels, and surface class characteristic values when clear sky values cannot be obtained. Greatest improvement in computed cloud fraction is realized over snow and ice surfaces; over other surfaces all versions perform similarly. Since the inclusion of SMMR for surface analysis increases the computational burden, its use may be justified only over snow and ice-covered regions.

Keywords: ISCCP, AVHRR, Arctic clouds.

INTRODUCTION

The important role that polar processes play in the dynamics of global climate is widely recognized (Polar Research Group, 1984). The variation of cloud amounts over polar ice sheets, sea ice, and ocean surfaces can have important effects on planetary albedo gradients and on surface energy exchanges (Barry et al., 1984). Cloud cover exerts a major influence over the amount of solar and longwave radiation reaching the surface, and is linked to the sea ice through a series of radiative, dynamical, thermodynamic and hydrological feedback processes (Saltzman and Moritz, 1980). Extent and thickness of sea ice influences oceanic heat loss and surface albedo which thereby influences global climate via the ice-albedo feedback (Budyko, 1969).

Current procedures for automated analyses of satellite radiance data have been developed for low and middle latitudes but their application to polar regions has been largely unexplored. These methods are reviewed in Key (1988), Rossow (1989), and Key et al. (1989). Those that have been applied to polar data often fail in the polar regions because snow-covered surfaces are often as reflective as the clouds and the thermal structure of the troposphere is characterized by frequent isothermal and inversion layers. Additionally, the polar darkness during winter makes that data collected in the visible portion of the spectrum largely unusable. A complex analysis method that can recognize and cope with these situations is therefore necessary (WMO, 1987).

The purpose of this study is to implement, test, and modify the basic International Satellite Cloud Climatology Project (ISCCP) algorithm for use with polar data. The International Satellite Cloud Climatology Project (ISCCP) to map clouds with satellite data began in July 1983. Its goal is to provide a uniform global climatology of satellite-measured radiances and from these to derive an experimental climatology of cloud radiative properties. As a basis for developing the ISCCP algorithm, Rossow et al. (1985) compared six cloud algorithms. The current state of the project is such that there is no single version of the algorithm which can be applied to all areas of the globe. The algorithm is currently operational globally, but performs rather poorly at high latitudes (Rossow, 1987), where it was found that the method in general detected too much cloudiness, in part because it does not distinguish between open water/sea ice and snow-covered/snow-free land, and because thresholds were not "tuned" for the small temperature differences and generally low IR radiances common in the polar regions.

DATA

The Advanced Very High Resolution Radiometer (AVHRR) on board the NOAA-7 polar orbiting satellite measures radiance in five channels encompassing the visible, infrared, and thermal portions of the electromagnetic spectrum (1: 0.58-0.68μm, 2: 0.73-1.0μm, 3: 3.55-3.93μm, 4: 10.3-11.3μm, 5: 11.5-12.5μm) with a nadir resolution of 1.1 km. Global Area Coverage (GAC) imagery is a reduced-resolution product created through on-board satellite processing, with each pixel representing a 3 x 5 km field of view. Channels 1 and 2 were converted to approximate spectral albedo and corrected for solar zenith angle; channels 3, 4, and 5 were converted to brightness temperature (NOAA, 1984). The typically low water vapor content in the polar atmosphere and the low physical temperatures reduce most atmospheric effects to a point where they may be neglected for these analyses.

The Nimbus-7 Scanning Multichannel Microwave Radiometer (SMMR) senses emitted microwave radiation in five channels: 6.6, 10.7, 18.0, 21.0, and 37.0 GHz, with two polarizations (horizontal and vertical) per channel. At these frequencies, passive microwave data is relatively unaffected by clouds and provides useful data year-round independent of solar illumination. The 18 and 37 GHz vertical polarization channels are employed here for surface parameterization, with

fields of view of 55x41 km and 27x18 km, respectively.

In order to study both clouds and surfaces beneath clouds, it is worthwhile to combine the AVHRR and SMMR data into a single image set. These are merged in digital form and mapped to a polar stereographic projection. This projection yields equal-area pixels true at 70° latitude with a five kilometer pixel size. SMMR data were converted to the five kilometer cells by duplication of pixels. Further details are given in Maslanik et al. (1988).

Three areas of the Arctic are examined (Figure 1); a seven-day summer series (July 1-7, 1984) of areas 1 and 2, and a winter series (January 6-12, 1984) of area 3. These data are part of an ISCCP test data set and include representative samples of all surface types found in the Arctic. snow-covered and snow-free land, sea ice of varying concentrations, open water, and permanent ice cap.

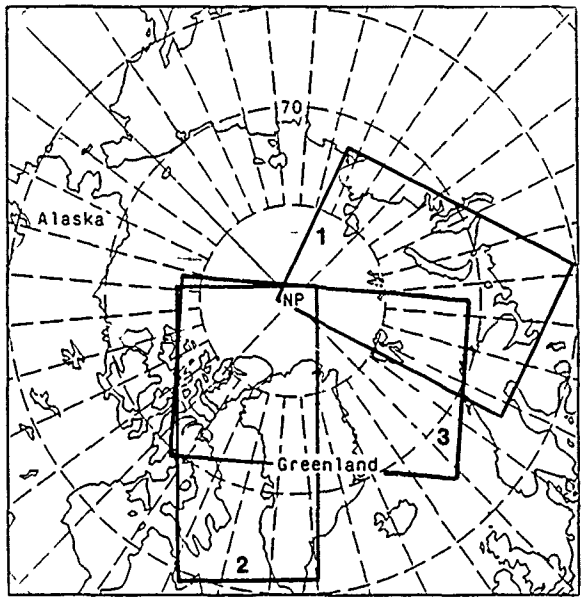


Figure. 1. The three study areas within the Arctic, one centered on the Kara and Barents Sea and the other two covering much of the Canadian Archipelago and northern Greenland.

Ancillary data include surface temperatures from the European Centre for Medium Range Forecasting (ECMWF) and sea ice albedo derived from a combination of DMSP imagery and the NOAA/Navy ice charts.

Spectral features examined are AVHRR channels 1, 2, and 3 albedos, channels 3, 4, and 5 brightness temperatures, and the differences between channels 3 and 4 and 4 and 5. Four surface and three cloud classes are analyzed. Surface types are snow-free land, snow-covered land, open water, and medium to high concentration sea ice. Cloud classes are low, middle, and high as defined by brightness temperature in AVHRR channel 4, assumed to represent temperatures at the top of optically thick cloud layers.

In order to test the sensitivity of the various cloud algorithms, a control data set with known characteristics was needed. A synthetic data set was developed which consists of seven days of AVHRR and SMMR data, SMMR-derived sea ice concentrations, and a land mask. The surface and cloud type maps for each day of the seven day period are first generated, where the minimum and maximum allowable sizes of surface and cloud "objects" are specified. Object dimensions are randomly chosen within the restricted range, and the class of the object is randomly assigned. Regions are then filled with normally-distributed data for each AVHRR and SMMR channel and for sea ice concentration using empirically-derived statistics.

THE ISCCP ALGORITHM

The ISCCP cloud algorithm has three major components: cloud detection, radiative analysis, and statistical analysis (Rossow et al., 1985). Of concern here is the cloud detection step. The algorithm assumptions are that cloud scene radiances are more variable in time and space than clear scene radiances and cloudy scenes are associated with larger visible channel and smaller infrared radiances than clear scenes. The major steps of the basic algorithm are summarized in Figure 2a and include a spatial variation test for the warmest pixel in a subregion (assumed clear), a temporal variation test, five-day clear sky compositing (requiring seven days of data), and a final bispectral threshold test of each pixel on each day based on clear sky composite values. Although this algorithm is undergoing constant change, the basic steps described here remain valid.

The algorithm has been adjusted here in order to deal with the problems cited earlier. Major modifications suggested include the use of snow and ice data sets for the estimation of surface parameters, elimination of the spatial test for the warmest pixel, the use of AVHRR channels 1 (0.7μm), 3 (3.7μm), and 4 (11μm) in the temporal tests, statistical tests for compositing, and the final multispectral thresholding, and the use of surface class characteristic values when clear sky values cannot be obtained. Additionally, the difference between channels 3 and 4 is included in temporal tests for the detection of optically thin low cloud and cirrus. The major steps of the modified algorithm are shown in Figure 2b.

TESTING AND ALGORITHM COMPARISON

Three versions of the ISCCP algorithm are compared. The original algorithm developed for low latitude summer conditions recognizes only two surface types: land and water. No SMMR or sea ice concentration data are employed. Spatial/temporal tests in the initial classification step are thermal only (AVHRR channel 4), and a bispectral threshold test (channels 1 and 4) is used as the final classification. This version with a thermal-only threshold test was also used to simulate winter applications. The algorithm with modifications described in the previous section is the third version tested. Four regions from the AVHRR imagery and four regions from the synthetic data sets are used as test data. Each region is 50x50 pixels or $(250km)^2$ and differs in surface and cloud types and proportions. Surface proportions changed in both data sets over the analysis period by up to 20%. These changes are due to sea ice movement and melting.

All versions of the algorithm perform best over land and open water. Snow and ice remain the problem areas although the modified versions performed best under these conditions. When cloud amounts are high (more than 80%), all versions compute cloud fraction to within approximately 5%. When cloud amounts are low, the modified version is more accurate, although cloud fraction often appeared to be too high. In the actual data, this is at least in part due to errors in the manual interpretation. In the synthetic data, this is probably due to the fact that clear sky areas are filled with values in the range of the mean plus or minus three standard deviations (following a Gaussian probability function), so that extreme values may be beyond threshold cutoffs and will consequently be labeled as cloud.

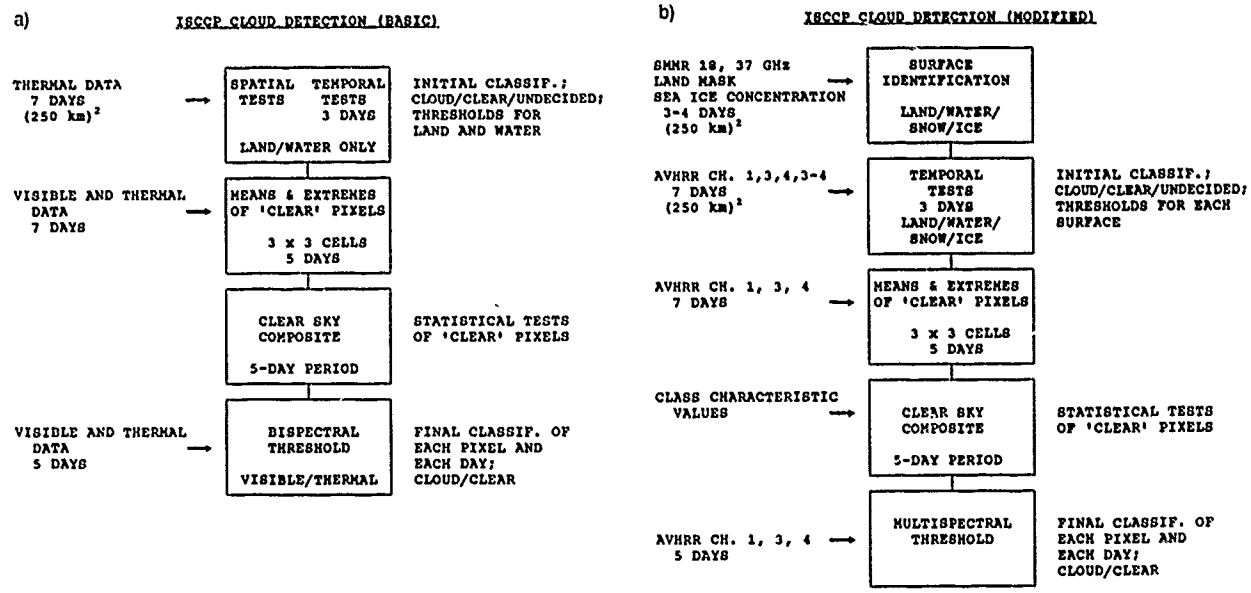


Figure. 2. Flow chart of (a) the basic ISCCP cloud detection algorithm and (b) the algorithm modified for use with Arctic data. Input are shown on the left; additional details are given on the right.

The basic versions of the algorithm often overestimate cloud amount by up to 20%. This is common over ice where, in the bispectral threshold test, the threshold for water is used. This albedo threshold is too small to account for variation in sea ice albedos, and consequently many clear pixels were mistaken as cloud. Similar observations were made by Rossow (1987). A related situation is that the basic version often makes an accurate assessment of cloud fraction, but for the wrong reason. For example, one sea ice region was covered by a very thin cloud layer. Channels 1 and 4 alone did not detect this condition, yet the cloud amount determined by the original algorithm version is similar to the manually-interpreted amount. It appears that the algorithm is labeling cloud what it sees in channel 1 as sea ice. The snow and ice data sets used in the modified versions solve these problems by providing appropriate thresholds.

Root mean square (RMS) errors illustrate that the modified version was most accurate in computing cloud fraction for both data sets. The thermal-only version performed reasonably well with the synthetic data set, at least in part for reasons explained above.

APPLICATION

The modified version of the algorithm is next applied to the Arctic study areas. Surface albedos determined over the five-day compositing period for the two summer study areas, which overlap the winter area, are shown in Figure 3. Values are averages over each quarter region. Sea ice albedo derived from DMSP imagery and NOAA/NAVY ice charts is in general agreement for sea ice. Composited surface temperatures are shown in Figure 4 for AVHRR channel 4. Since surface emissivities in channel 4 (11 m) are near unity, the temperatures presented are considered to be close estimates of physical temperatures. These are in close agreement with the ECMWF data.

Cloud fraction for the middle five days of the period was also computed with the modified ISCCP algorithm (not shown), and is similar to the mean cloud amount for each month. Cloud amount tends to be lowest over northern Greenland and the Canadian Archipelago, and highest over the Greenland Sea.

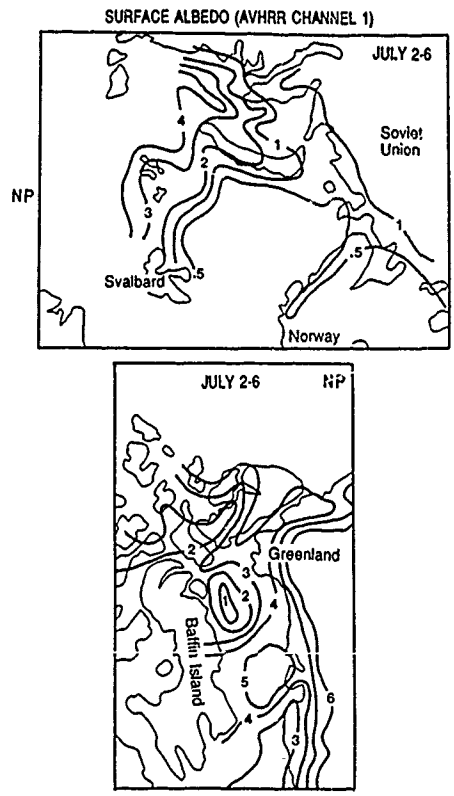


Figure. 3. Surface albedos (AVHRR channel 1) in study areas 1 and 2 for the compositing period July 2-6, 1984, in tenths.

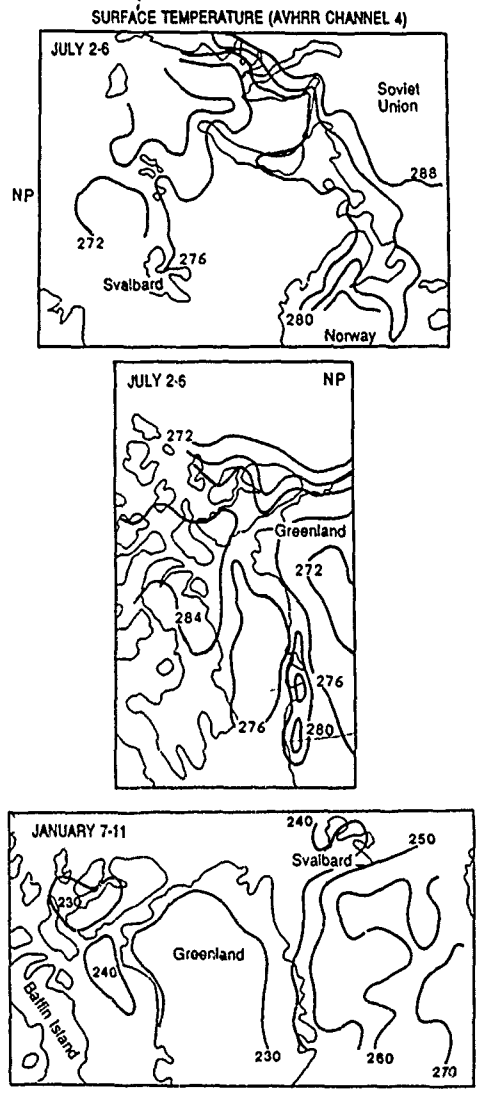


Figure. 4. Surface temperatures (AVHRR channel 4) for the compositing periods, July 2-6, 1984 (top and middle) and January 7-11, 1984 (bottom), Kelvins, for all three study areas.

CONCLUSIONS

The cloud detection step of the International Satellite Cloud Climatology Project algorithm has been adapted for use with Arctic AVHRR and SMMR data. Based on test data for summer conditions in the Arctic, the modified algorithm is expected to yield an average improvement of 5-10% in computed cloud amount over the original version, depending on surface type and cloud proportions. All versions of the algorithm perform best over snow-free land and open water, so that improvement will be greater than this figure over snow, ice cap, and sea ice, but less over open water and snow-free land. Synthetic data sets have proven useful in testing and validation.

The best method of cloud detection with Arctic AVHRR data includes first an accurate identification of surface types and changes. This allows thresholds to be set appropriately. Passive microwave data is useful in this step. Next the temporal variability of pixel radiances must be examined, using AVHRR channels 1, 4, and the reflected component of channel 3 during summer and the difference between channels 3 and 4 in conjunction with channels 4 or 5 winter analyses. Temporal changes are most important in winter when surfaces may be colder than cloud layers and spectral information alone is inadequate. Compositing over a 5-day period, using 30-day values where necessary, provides the clear sky information for the multispectral thresholding of the daily data. This method provides a basis for future cloud detection algorithms for the polar regions.

ACKNOWLEDGEMENTS. This work was supported under NASA grant NAG-5-898. Thanks are due to W. Rossow and E. Raschke for providing AVHRR GAC data.

REFERENCES

1. Barry, R.G. A. Henderson-Sellers, and K.P. Shine, "Climate sensitivity and the marginal cryosphere", in Climate Processes and Climate Sensitivity, J. Hansen and T. Takahashi (eds.), Geophys. Monog. 29, Amer. Geophys. Union, Washington, DC, pp221-337, 1984.
2. Budyko, M.I., "The effect of solar radiation variations on the climate of the Earth", Tellus, Vol. 21, pp611-619, 1969.
3. Key, J.R., "Cloud analysis in the Arctic from combined AVHRR and SMMR data", Ph.D. dissertation, Department of Geography, University of Colorado, Boulder, 180 pp., 1988.
4. Key, J.R., J.A. Maslanik, and R.G. Barry, "Cloud classification using a fuzzy sets algorithm: a polar example", Int. J. Rem. Sens., 1989 (in press).
5. Maslanik, J.A., Key, J.R., and Barry, R.G., "Merging AVHRR and SMMR data for remote sensing of ice and cloud in the polar region", Int. J. Rem. Sens., 1989 (in press).
6. NOAA, NOAA polar orbiter data user's guide, U.S. Department of Commerce, National Oceanic and Atmospheric Administration, NESDIS, February, 1984.
7. Polar Research Board, The Polar Regions and Climatic Change, National Research Council, National Academy Press, Washington, D.C., 59 pp., 1984.
8. Rossow, W.B., "Application of ISCCP cloud algorithm to satellite observations of the polar regions", in WMO, 1987. Report of the ISCCP workshop on cloud algorithms in the polar regons, National Institute for Polar Research, WCP-131, WMO/TD-No. 170, Tokyo, Japan, 19-21 August 1986, 1987.
9. Rossow, W.B., "Measuring cloud properties from space: a review", J. Climate, Vol. 2, pp201-213, 1989.
10. Rossow, W.B., F. Mosher, E. Kinsella, A. Arking, M. Desbois, E. Harrison, P. Minnis, E. Ruprecht, G. Seze, C. Simmer, and E. Smith, "ISCCP cloud algorithm intercomparison", J. Clim. Appl. Meteor., Vol. 24, pp877-903, 1985.
11. Saltzman, B. and R.E. Moritz, "A time-dependent climatic feedback system involving sea-ice extent, ocean temperature, and CO_2", Tellus, Vol. 32, pp93-118, 1980.
12. WMO, "Report of the ISCCP workshop on cloud algorithms in the polar regons", World Climate Research Programme, WCP-131, WMO/TD-No. 170, Tokyo, Japan, 19-21 August 1986, 1987.

Tropical Deforestation In Pará, Brazil: Analysis with Landsat and Shuttle Imaging Radar-A

Thomas A. Stone, George M. Woodwell, Richard A. Houghton

The Woods Hole Research Center
P. O. Box 296
Woods Hole, Mass. 02543 USA

Abstract

Deforestation globally is proceeding at rates far in excess of previous experience. The amount of carbon being released from soils and forest as a result is uncertain. It is generally thought to be in the range of 1-3 X 10^{15} g annually, but the data on which the estimate is based are from 1980 and the current surge of deforestation may have pushed the rate of emissions beyond these estimates. The most satisfactory measurements will be obtained in due course from satellite imagery. The need is immediate.

We have examined forests in the Brazilian Amazon Basin state of Pará with Landsat and Shuttle Imaging Radar-A (SIR-A) data. The area of primary forest there can be defined with either the combined SIR-A/Landsat data or with the Landsat or SIR-A data alone. It appears that SIR-A data helps distinguish a greater number of classes within the primary forest. It does not appear to help similarly for clearings. The highest SIR-A data values comes from areas that appear to be degraded primary forest.

Secondary forest has replaced many of the sites originally cleared. They are easily distinguishable from primary forests with Landsat data. Secondary forests had NDVIs (0.60 to 0.69) characteristic of rapidly growing vegetation. Landsat was sufficient to classify this type of secondary forest.

L-Band radar data would be most helpful in determining primary forest composition but are insufficient in areal coverage to help with the determination of the areas or rates of deforestation.

I. Introduction and Objectives

Our recent work has been in the forests of the Amazonia because of its high rates of deforestation. For this work we sought sites where both Landsat and radar data were available and where successional or secondary forest stands were common. The state of Pará in the eastern Amazonia was chosen because the imagery was available and there seemed to be a higher probability of finding secondary forests.

Our hypothesis was that the 23.5 cm wavelength Shuttle Imaging Radar-A (SIR-A) data added new information fundamentally different from that available from optical satellites for the analysis of forests. For instance, Hoffer et al. (1986) has shown strong correlations between SIR-B L-Band backscatter and biomass for pine forests and Sun and Simmonett (1988) have found the major contributor to L-Band backscatter was the trunks of pines. It is clear that radar have the potential to add information on the woody biomass of the forests to the data on leaf activity available from optical satellites. But, the pines studied by others and the tropical wet forests are intrinsically different in canopy volume, structure and moisture content.

II. Methods

A. The Paragominas Region: Eastern Pará

The state of Pará is the second largest Brazilian state in Amazônia, covering 1,248,000 km^2. By 1975 8,650 km^2 had been cleared; by 1978 clearings had increased 160% to 22,445 km^2 (Tardin et al. 1980). By 1980 the area cleared had increased another 50% to 33,900 km^2 (fide Fearnside 1985), about 3% of the total area of forests of Pará. There are no more recent estimates.

More than 90% of clearing is for pasture. There has been sufficient time for abandoned pasture to revert to secondary forest, if the pasture was not severely degraded. Hecht (1982) estimated that 50% or more of the pasture in Paragominas was degraded.

B. Imagery

We acquired Landsat images of the Paragominas region for 1981 and 1986 from the Brazilian Space Agency, MCT/INPE. We sought Landsat data from dates as close as possible to November 15, 1981, the date of the shuttle overpass. The best match was a Landsat Multispectral Scanner (MSS) image from August 6, 1981. A Landsat Thematic Mapper (TM) image of the same region dated July 17, 1986 was acquired. A part of the optical film of SIR-A Data Take #31 was digitized at the Jet Propulsion Laboratory (JPL). SIR-A data cover a swath 50 km wide, were acquired at an incidence angle of 50° +/-3 and are HH polarized (Cimino and Elachi [eds.] 1982).

C. Data Entry and Setup

The 1981 MSS and radar data were used with an ERDAS-PC system. Each file of radar data was smoothed with a 3 X 3 low pass filter to reduce speckle and registered to the 4 bands of MSS data to create a 5-banded image. The 1981 MSS and 1986 TM data were co-registered and covered a common area of 15,750 km^2.

D. Clearings and Forests

The 5-banded data set allowed interpretation of the SIR-A and MSS data within two broad categories, primary forests and cleared land. The second category also included secondary forest younger than 26 years.

The first approach was analysis of the correlation of the MSS and SIR-A data of areas that were obviously primary forest or clearings. We selected 55 clearings various sizes from the 5-banded dataset. We then examined the distribution of MSS normalized difference vegetation index (NDVI) data. The NDVI was calculated by subtracting Landsat MSS5 from MSS7 and normalized; NDVI=(MSS7-MSS5)/(MSS7+MSS5). The NDVI is correlated

with green-leaf biomass and leaf-area (Tucker et al. 1980, Holben et al. 1980). We also examined the distribution of the NDVI data against the radar data for a 1,160 km^2 block that was half forested.

A second approach to the examination of the 5-banded dataset was an unsupervised classification of large blocks (1,160 km^2) and a similar classification of the same blocks with only the 4 MSS bands.

A third approach was to analyze the data of a 1,160 km^2 block with principal components.

E. Change Detection

We used change detection techniques with Landsat to determine the amount of clearing and forest regrowth from 1981 to 1986. We subtracted the spectral values of MSS5 and MSS7 for 1981 from 1986 to create a difference image which highlighted the areas of greatest change such as forest to cleared or from pasture to forest. The final step in the change detection process was unsupervised classification of the change image.

III. Results

A. MSS NDVI and SIR-A Data

The mean NDVI for all "clearings" was 0.40 and for forests was 0.57. The clearing category could include everything from bare soil to secondary forest, 21 to 26 years old, as old as the first clearings in the region.

There was a strong correlation of MSS data between the two visible and the two near-infrared bands (Table I). The radar data were positively correlated to the near-infrared bands and negatively to the visible bands.

The mean NDVI for forests was 0.57 which is typical. There was little variance in the MSS bands; more variance occurred in the radar data of the same sites. The correlations between the four MSS bands and the one radar band (Table II) were similar to the clearings.

The radar and NDVI data for the 1,160 km^2 block were examined. Forested sites had the strongest radar response, consistent with the observations of others (Wu 1983, Brisco et al. 1983) and unlike Mato Grosso where some cleared areas had stronger SIR-A backscatter than the surrounding forest (Stone and Woodwell 1988). However, the largest contiguous areas of highest radar return were primary forests immediately adjacent to very large clearings (10 to 60 km^2) with the lowest NDVIs (Fig. 1). That this forest had the strongest radar response was probably due to degradation of the forest due to selective logging, which may damage up to 50% of the remaining trees, clearing fires which have escaped into the forest, or microclimate changes which make the primary forest all "fringe". It appears there was greater radar response from degraded primary forest due to a combination of canopy loss, increased soil response, and enhanced response from the trunks of trees.

Another dissimilarity with the Mato Grosso data were the presence of numerous areas with significantly higher NDVIs and lower radar backscatter than the primary forest sites, and with higher backscatter than the majority of clearings. High NDVI sites are thought to have high photosynthetic capacity. One hypothesis is that these were secondary forests, vigorously growing yet lower in standing woody biomass and therefore lower in radar backscatter than primary forests. This hypothesis was supported by the observation that areas defined in the field as secondary forest (C. Uhl pers. comm. 1988, also had NDVIs greater than 0.60.

As we proceed from the lower to higher NDVIs (Fig. 2), there was little increase in radar backscatter until an NDVI of about 0.50. At this point the NDVIs of the primary forest began to dominate. Areas that appeared to be secondary forest had the highest NDVIs and slightly less radar backscatter than the primary forest. Other areas with the highest radar backscatter and the same NDVI as primary forest are immediately adjacent to the largest clearings and appear to be degraded primary forest.

B. Unsupervised Classification of Merged Data Set

Unsupervised clustering of five blocks (1,160 km^2) was done for both the MSS/radar data set and for the MSS data alone.

For the MSS data alone the variability was greater among the clearings as opposed to the forest sites. The greater variability resulted in identification of several clusters of similar plots among the clearings as opposed to a single cluster among the forest plots (Table III).

When the merged MSS and SIR-A data were clustered the variability in forests was increased sufficiently to add another forest type without affecting the number of classes of clearings. This observation confirms that radar data will be more helpful in understanding forest types than types of clearings.

C. Principal Components

Principal components analyses (Table IV) of the mean MSS and SIR-A values for clearings revealed that 76% of the variance could be described with one principal component, 91% could be described with two and, 99.9% could be described with three. In remotely sensed imagery the first principal component (PC1) usually describes the overall brightness variance in the data. Information on other aspects of the image generally resides in the other components (Forster 1985).

PC1 for the clearings was almost equally distributed among all 5 bands of data. The second principal component (PC2) described 15% of the variance and was dominated by the radar coefficient and by the two near-infrared bands. The majority of the variance was in the radar data. The third principal component (PC3) was divided almost equally among all five bands of data. The results suggest that there is little additional information to be gained from use of L-Band radar in these clearings.

Principal components analysis of the forests (Table V) revealed that 66% of the variance was described by PC1, 27% described by PC2, and 8% of PC3. If PC1 describes the overall brightness variance in the data, all the remaining differences within forest types were in PC2 and PC3.

PC1 in the forest analysis described 66% of the variance compared to 76% in the clearings. Because there was a greater range of brightness in the clearings this result was expected. PC2 described 27% of the variance in the forest sites compared to 15% in the clearings. As in the clearing sites, PC2 was dominated by the SIR-A data and by the first near-infrared band. Because of the greater amount of variance described by PC2 and PC3 for the forest sites compared with the clearing sites, it appears that radar data would be more helpful in distinguishing between forest types than between types of clearings. PC3 was also dominated by the SIR-A data unlike the clearing sites where the green band (MSS4) dominated.

Because the correlation between MSS and radar data was weaker in the forest sites than in the clearing sites there is more basis for looking more closely at L-Bands data to understand these forests.

D. Change Detection

The results of an unsupervised classification of the region covered by both the 1981 and 1986 Landsat data (15,751 km^2) showed the region to be 78.1% forest, 20.4% clearings and 1.5% clouds, shadows and bad data in 1981.

Santos et al. (1979) estimated that 18.4% of Paragominas had been cleared by 1978, similar to our estimate for 1981 of 20.4% cleared of the portion of Paragominas that we studied. The mean rate of clearing from 1960 to 1981 was 153 $km^2\ yr^{-1}$ or 0.97% yr^{-1}, the same rate found by Santos et al. (1979). In contrast, the statewide clearing rate was 0.37% yr^{-1} from 1975 to 1978 and 0.45% yr^{-1} from 1978 to 1980 (Tardin et al. 1980). Consequently, the deforestation rate in the Paragominas region was double the statewide average.

The unsupervised classification of the 15,750 km^2 change image showed that 69.2% of the area remained unchanged as forest. Because we knew that 78.1% of the region was forest in 1981, another 8.9% of the region could have been cleared. As 2.8% of change was due to clouds, shadows, and flooding, 70.2% of the area was estimated to have remained as forest.

The area of new clearing was 7.9% (78.1% less 70.2%). The rate of forest clearing from 1981 to 1986 was 1.6% yr^{-1}, an increase of 50% over the mean 1960-1981 rate of forest clearing.

IV. Discussion and Summary

We have explored here the contributions made using SIR-A data to supplement Landsat in detecting and measuring deforestation and reforestation in the tropics. Currently, the contributions are limited. Optical data from Landsat and other satellites are sufficient for an inventory of forest versus nonforest over large areas. Radar data offer the intriguing possibility of directly determining woody biomass and improving our understanding of forest types, regrowth, and degradation but at this point the quality and quantity of the data are insufficient.

A. Rates of Forest Clearing

The area of forest cleared from 1960 to 1981 was 20.4% of the area examined (15,571 km^2). The rate from 1960 to 1981 was 0.97%/year, double the statewide average. If the majority of clearing began in 1969 as suggested by Santos et al. (1979), then the rate of clearing has been 1.7%/yr from 1969 to 1981. This is a high rate by any standard.

The amount of clearing for 1981 to 1986 was an additional 7.9%. This is the same as the rate from 1969 to 1981. There was no indication that the rate of clearing was slowing because of the loss of incentives for deforestation. Rates of forest clearing are not declining in any of the areas of Amazonia we have examined to date; Rondonia, Mato Grosso, and eastern Pará.

B. Primary and Secondary Forests and Clearings

The area of primary forest can be defined with either the combined radar/Landsat data or with the Landsat or radar data alone. Based on clustering analysis the L-Band data helps distinguish a greater number of classes within the primary forest than in the cleared lands. SIR-A data appears also to provide information on degraded primary forest which is not available from optical data.

The results reported were affected by fading of the radar signal. But the 25% of the variance in the data related to factors other than total brightness may be affected by woody biomass differences or other differences in the vegetation such as moisture or canopy volume.

Secondary forest are distinguishable from the primary forest with MSS data alone. Secondary forests had a very high NDVI (0.60 to 0.69) which is characteristic of rapidly growing vegetation. The secondary forests have a small amount of woody biomass compared to the primary forest which may result in a diminished radar signal.

C. Radar and Optical Data Correlations

Radar backscatter did not vary significantly with NDVI for the areas of pasture and clearings. Therefore, factors other than NDVI and its correlate, green leaf area, were controlling the radar response. It may instead be controlled by remnant woody biomass, as suggested by Stone and Woodwell (1988) or by other factors, and therefore be unresponsive to changes in NDVI up to 0.45 or to leaf area.

Above an NDVI of 0.50 there were three well defined forest types, probable disturbed primary forest with the highest mean radar values and with NDVIs from 0.50 to 0.60, undisturbed primary forest with intermediate radar values and NDVIs from 0.50 to 0.60, and secondary forest with intermediate to lower radar values and NDVIs from 0.55 to 0.69. That the secondary forest has a lower radar return signal than the primary forest could be an indication that woody biomass was dominating the radar return signal from the primary forests.

The majority of the variance between different types of clearings is displayed along the NDVI axis and the majority of the variance between types of forests is displayed along the SIR-A axis. To determine the biomass classes of regions dominated by green biomass and woody biomass we need to use the both the visible and microwave portions of the spectrum.

L-Band data would be most helpful in determining primary forest composition but are insufficient in areal coverage to help with the determination of the areas or rates of deforestation. Optical data such as those from Landsat and AVHRR are sufficient for determining the area and rates of deforestation and reforestation in the tropics.

VI. Acknowledgements

All Landsat data was purchased from MCT/INPE. This work was performed under contract with Martin Marietta Energy Systems Inc. with the U.S. Department of Energy, and under JPL Contract #95811 with NASA.

V. References

Brisco, B. Ulaby, F. and M. Dodson, 1984. Spaceborne SAR data for land cover classification and change detection. IGARSS '83, PS-2:1.1-1.8.

Cimino, J.B., and C. Elachi, (eds.) 1982. Shuttle Imaging Radar-A Experiment. Publ. 82-77, JPL, Pasadena.

Fearnside, P., 1985. Deforestation and Decision Making in the Development of Brazilian Amazonia. Interciencia 10(5):243-247.

Forster, B., 1985. Principal and Rotated Component Analysis of Urban Surface Reflectances, Photogr. Eng.& Rem. Sens., 51(4):475-477.

Hecht, S., 1982. Deforestation in the Amazon Basin: Magnitude, Dynamics and Soil Resource Effects. Studies in Third World Societies 13:61-108.

Hoffer, R., Lozano-Garcia, D., Gillespie, D., Mueller, P., and M. Ruzek, 1986. Analysis of Multiple Incidence Angle SIR-B Data for Determining Forest Stand Characteristics. The Second Spaceborne Imaging Radar Symposium, JPL Publ. 86-26, pp.159-164.

Holben, B., Tucker, C., and C-J Fan, 1981. Spectral assessment of soybean leaf area and leaf biomass. Photogr. Eng. & Rem. Sens. 46(5):651-656.

Santos, A., Novo, E., and V. Duarte, 1979. Relatorio Final Do Projeto INPE/SUDAM. INPE-1610-RPE/085, Sao Jose Dos Campos, 148 pp.

Stone, T., and G. Woodwell, 1988. Shuttle Imaging Radar-A Analysis of Land Use in Amazonia. Inter. J. of Rem. Sens. 9(1):95-105.

Sun, G. and D. Simonett, 1988. Simulation of L-Band HH radar backscatter from coniferous forest stands: A

comparison with SIR-B data. Inter. J. of Rem. Sens.

Tardin, A., Lee, D., Santos, R., de Assis, O., Barbosa, M., Moreira, M., Pereira, M., and C. Filho, 1980. Subprojecto desmatamento convenio IBDF/CNP-INPE, Sao Jose Dos Campos, 44 pp.

Tucker C., Holben, B., Elgin, J., and J. McMurtrey, 1980. Relationship of spectral data to grain yield variation. Photogr. Eng. & Rem. Sens. 46(5):657-666.

Wu, S.T., 1983. Analysis of synthetic aperture radar data acquired over a variety of land cover. Digest IGARSS '83, 2(FP-5):1-6.5.

****

	B1	B2	B3	B4	SIR
B1	1.00				
B2	0.97	1.00			
B3	-0.67	-0.76	1.00		
B4	-0.78	-0.85	0.98	1.00	
SIR	-0.53	-0.50	0.37	0.43	1.00

Table I. Correlation between the Spectral data of Clearings.

****

	B1	B2	B3	B4	SIR
B1	1.00				
B2	0.99	1.00			
B3	-0.35	-0.39	1.00		
B4	-0.63	-0.66	0.93	1.00	
SIR	-0.65	-0.60	0.01	0.27	1.00

Table II. Correlation between the Spectral data for Forests.

****

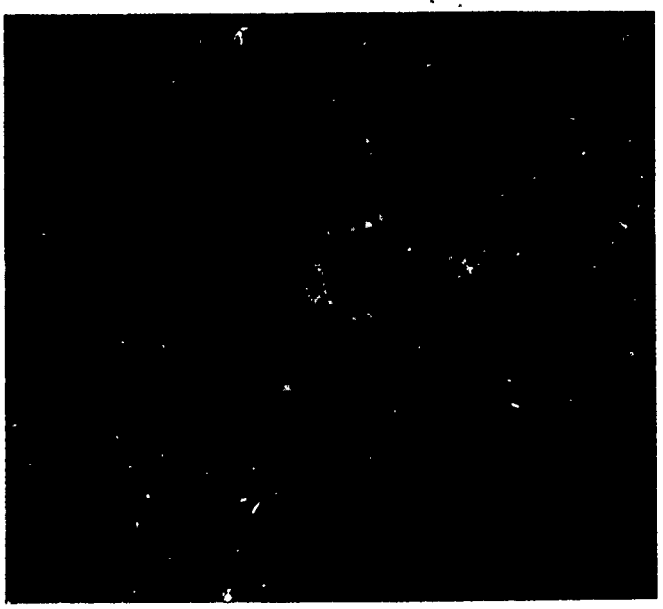


Figure 1. Smoothed SIR-A data of 1200 km² near Paragominas, Pará The large center field covers 57 km². The lower right of the image is the near side of radar data acquisition. Note the brightest radar backscatter is from forests immediately to the north, west, and south of the large central field. These bright tones represent areas of probable primary forest degradation.

	Landsat data only			Landsat + SIR-A		
block1	91.9%	for	1 clus	89.2%	for	2 clus
	6.8%	clr	3 clus	9.9%	clr	3 clus
block2	83.7%	for	1 clus	79.1%	for	2 clus
	16.2%	clr	7 clus	20.3%	clr	4 clus
block3	81.1%	for	1 clus	80.2%	for	3 clus
	18.0%	clr	6 clus	18.2%	clr	3 clus
block4	57.7%	for	2 clus	46.9%	for	3 clus
	42.1%	clr	7 clus	46.2%	clr	5 clus
block5	88.1%	for	1 clus	79.6%	for	3 clus
	10.4%	clr	2 clus	19.1%	clr	1 clus
mean forest	80.5%			mean forest	76.4%	
clearing	17.3%			clearing	22.7%	

Table III. Summary of results of unsupervised classification of clearing types and primary forest types using 1,160 km² blocks of MSS data and merged MSS/SIR-A data.

****

	Described Variance %	Cuml%	Eigenvector SIR-A	Green	Coefficients Red	NIR1	NIR2
P1	76.1	76.1	0.31	-0.47	-0.49	0.46	0.49
P2	15.1	91.2	0.89	-0.09	0.02	-0.37	-0.28
P3	8.4	99.6	-0.36	-0.58	-0.44	-0.51	-0.30
P4	0.4	99.9	0.03	0.59	-0.74	0.12	-0.31

Table IV. Clearings Principal Components.

****

	Described variance %	Cuml%	Eigenvector SIR-A	Green	Coefficients Red	NIR1	NIR2
P1	65.5	65.5	0.34	-0.51	-0.51	0.36	0.48
P2	26.6	92.1	0.55	-0.26	-0.22	-0.64	- 0.42
P3	7.6	99.7	-0.76	-0.39	-0.44	-0.25	-0.15
P4	0.2	99.9	-0.06	0.17	0.05	-0.63	0.75

Table V. Principal Components for Forests.

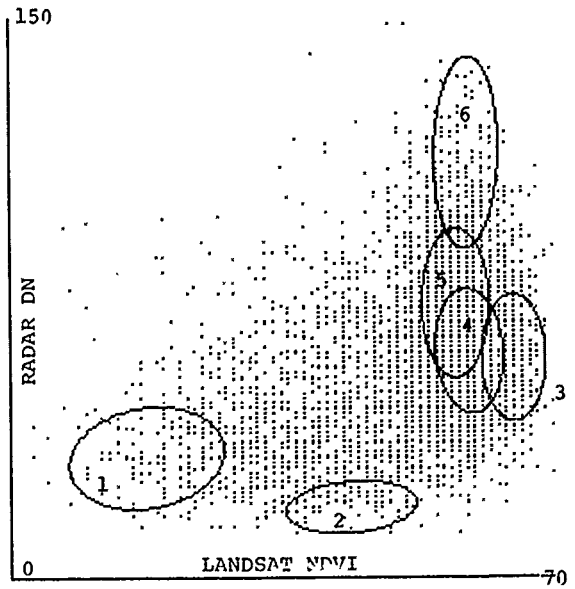


Figure 2. Radar DN and Landsat NDVI valucs for all pixels in a 1,160 km² block. Typical clearings are within ellipses 1 and 2. Areas of secondary regrowth with the highest NDVIs are in ellipse #3. Typical primary forests are ellipses #4 and #5 and probable degraded primary forests are #6, the region of strongest radar backscatter. Primary forests have higher radar DNs than clearings. Secondary forests had higher NDVIs than did primary forest or fields and a higher radar DN than fields. The majority of the information about different field types is along the NDVI axis and the majority of the information about forests is along the radar axis.

A COMPARISON OF AVHRR GAC AND HRPT DATA FOR REGIONAL ENVIRONMENTAL MONITORING

A. S. Belward and J-P. Malingreau

Commission of the European Communities, Joint Research Centre,
Institute for Remote Sensing Applications, 21020 Ispra (Varese) Italy.
Tel.(0)332 789830, Fax (0)332 789001, Telex 380042.

Over the last two decades interest in the condition of the Earth's environment and resources has grown considerably. Importance is now attached to all levels from specific populations to the processes of global change. As vegetation forms the base trophic level of all terrestrial ecosystems, the study of vegetation dynamics is a vital element of environmental monitoring. At the Institute for Remote Sensing Applications the analysis of vegetation dynamics is applied to the monitoring of renewable natural resources in Africa. Current topics are crop production, vegetation change in major river basins and degradation of the savanna - tropical forest ecosystem. Multispectral data from the Advanced Very High Resolution Radiometer (AVHRR) sensor on the NOAA series of satellites operated by the National Oceanic and Atmospheric Administration (NOAA) are used extensively in this work.

AVHRR data are available in full resolution (1Km) and reduced resolution (4 Km) formats. For the African continent the reduced resolution Global Area Coverage (GAC) data archive offers more complete coverage (both areally and temporally) than is available from full resolution data archives. There is near daily coverage of the entire continent at 4km resolution, or limited coverage at 1km resolution. The aim of this work is to establish the value of the low spatial resolution data through comparison with full resolution data sets.

Contemporaneous AVHRR imagery in reduced resolution GAC and full resolution High Resolution Picture Transmission (HRPT) data types are available for most of West Africa. Imagery from 25th January 1988 was found to be almost completely cloud free for the entire land mass. A study area was selected: 4°N to 20°N, and 18°W to 6°E - covering more than 4.5 million km^2. For this window all five channels of the HRPT and GAC data for 25th January 1988 were geometrically corrected using an orbital model plus satellite earth location points. Resampling was to a 1Km pixel by nearest neighbour for both data sets. A final image translation from manually derived control points ensured optimum image to image registration. The Normalised Difference Vegetation Index (NDVI) was calculated for both data sets, then difference images were generated for all channels plus the NDVIs. Mean value and variance were obtained on a range of cover types from original channels, NDVIs and difference images.

At GAC resolution detailed features such as small river valleys are lost. Regional scale features are preserved, providing information on vegetation dynamics over important cover types like the savanna - forest transition zone. GAC resolution data provide a synoptic view of the entire West African region in a single 512 pixel by 512 line window. This combined with the existence of good time series make these data well suited to environmental monitoring on a regional scale. The increased spatial resolution of the HRPT data is of value in providing more accurate spatial information, but the 16 fold increase in data processing volume plus the lack of good time series make these data less suitable for regional studies in the West African context.

RADARSAT: CANADA'S MICROWAVE SATELLITE

E. Shaw

Canada Centre for Remote Sensing
Energy, Mines and Resources Canada

E.J. Langham

RADARSAT Project Office
Canadian Space Agency

ABSTRACT

The overall characteristics of the RADARSAT remote sensing mission are described as an introduction to the papers that follow. Emphasis is placed on the features that make it unique from satellites that carry optical sensors including its sun-synchronous dawn-dusk orbit, cloud-free, real-time images, and electronic radar beam steering. These features lead to important differences in the availability of global data, geographic coverage and mission management.

INTRODUCTION

Now that agreements have been reached with all partners in RADARSAT, the Canadian Government will proceed with the construction and operation phases of this mission. This marks an important turning point in the life of a satellite project whose mission, design and configuration has evolved considerably. This paper reviews the present design in the context of program objectives and describes some of its special characteristics and the way in which the mission will be managed. Technical descriptions of the ground and space segments as well as the economic and commercial aspects of the mission are dealt with in more detail in the following papers.

MISSION OBJECTIVES

The RADARSAT project has its origins in the Canadian national requirement for information for resource management and environmental monitoring. In particular, during the 'oil crisis' of the late 70's when oil exploration was being vigorously pursued in the Arctic, it was apparent that sea state, ship position and ice information would be required on an unprecedented scale when oil tankers brought the oil to southern markets. Satellite remotely sensed data gives coverage on the scale and frequency required but poor illumination much of the year at high latitudes and persistent cloud cover make synthetic aperture radar (SAR) the only acceptable sensor technology for reliable, fast, information. Mission requirements for RADARSAT were based on national and bilateral studies with the U.S. and were updated following the results of an experimental program which had been conducted during and after the Seasat mission. These experiments showed that SAR data also had important land applications such as mapping of surface structure, soil moisture conditions, and crop and forestry monitoring in cloudy regions. Economic studies showed the associated benefits from land applications would be greater than those from sea and ice mapping.

The fact that RADARSAT is a Canadian project leads to the next mission objective which is to establish a Canadian mission control facility for polar orbiting satellites. Finally, RADARSAT data is to be collected and made available globally. This objective will be achieved by assigning the sole rights for data distribution to a private sector company.

PARTNERSHIPS

The RADARSAT project has partnerships with U.S. agencies, Canadian provincial governments and the private sector. Canada is responsible for the design and integration of the overall system, for construction of the radar, for the provision of the satellite platform, for control and operation of the satellite in orbit and for operation of the data reception stations in Prince Albert, Saskatchewan and Gatineau, Quebec.

- U.S. Agencies

NASA will provide the launch services and will operate a data reception station in Alaska in exchange for radar data for its research programs; NOAA will facilitate the participation of the American private sector in the distribution of data.

- Provinces

All Canadian provinces have participated in the planning of the RADARSAT Program. Quebec, Ontario, British Columbia and Saskatchewan will share in the costs in order to develop technology within their industries. An agreement has also been developed with the remaining provinces for their participation. They can make a pre-payment in order to receive data at cost of reproduction.

- Private Sector

A private sector company, RADARSAT International (RSI), has been incorporated to distribute the radar data. RSI is seeking industrial partners in the U.S. to provide the U.S. private sector with an appropriate role in the global marketing of SAR data. RSI has the right to distribute the data in excess of the international partners' governmental requirements. In return, they will invest in developing this market and provide royalty payments to the Government from sales.

SATELLITE MISSION

The RADARSAT launch is scheduled for mid-1994, using a medium-class, expendable launch vehicle. The satellite payload will consist of an advanced synthetic aperture radar (SAR), and its associated down-link transmitters, tape-recorders and its command computer. The satellite platform will furnish the electrical power, attitude stability, orbit maintenance, and housekeeping telemetry. The satellite will use space-qualified sub-systems and redundancy in order to attain its 5-year lifetime.

The orbit selected for RADARSAT is similar to the orbits for the LANDSAT and SPOT satellites; it is sun-synchronous with an altitude of about 800km, but the time of its descending and ascending equatorial crossings are about three hours earlier at dawn and dusk respectively.

- The Orbit

The orbit parameters for RADARSAT are shown in Table 1.

Geometry :	Circular, sun-synchronous
Equatorial crossing time:	6 a.m.
Altitude :	792km
Inclination :	98.6 |
Period :	100.7 minutes
Repeat Cycle :	16 days
Sub-cycle:	3 days

Table 1: Orbit Parameters

For microwave satellites, the dawn-dusk orbit offers several advantages because the solar illumination is more constant and the satellite is rarely in the earth's shadow. This leads to better thermal stability of the satellite and to smaller, more efficient solar arrays.

Also, because the satellite is in sunlight for most of the year, there is no need to carry large batteries to power the SAR in eclipse. As illustrated in Figure 1, there is a brief period of eclipse, maximum duration seventeen minutes, in the middle of the Antarctic Summer.

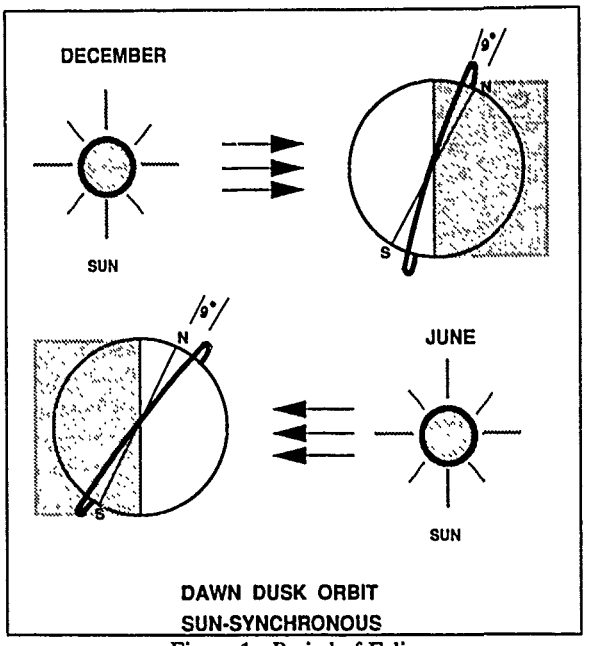


Figure 1 - Period of Eclipse

Perhaps the greatest advantage, however, is that the SAR can be turned on at any time in the orbit, which means that there is no distinction between ascending and descending passes from an operational point of view. This will greatly increase the opportunities for large scale geomorphological studies using different radar illumination directions.

Another advantage to operators of receiving stations for remote sensing satellites is that the data reception periods for RADARSAT do not conflict with the other optical sensor satellites. This allows more efficient use of the tracking antenna and recording equipment.

The normal configuration of the RADARSAT mission has the SAR pointing to the north so that there is almost complete coverage up to the North Pole. However, twice during the mission for a period of about 2 weeks, the satellite will be rotated 180 degrees about its yaw axis so as to direct the beam to the south. The purpose is to obtain a complete SAR map of Antarctica at the times of maximum and minimum ice cover.

THE SAR

The SAR is an advanced multi-mode instrument, that operates at 5.3 GHz (C-Band). It has a choice of three transmit pulses and numerous beam selections to give images with a variety of swath widths and resolution. (see Table 2).

Mode	Swath	Resolution	Incidence Angle
Standard	100km	28m x 30m	20 - 49 (4 looks)
High Resolution	55km	8m x 8m	20 - 49 (1 look)
Experimental	100km	28m x 30m	49 - 60
Scan SAR	500km	100m x 100m	20 - 49 (6 looks)

Table 2 - SAR Modes

The multiple beam feature is achieved by a beam-forming network with electronic phase-shifters, that can be changed in a fraction of a second. This enables both the beam width and the pointing direction to be changed very rapidly. Indeed in the SCANSAR mode the switching takes place continuously in effect creating two or more interleaved SAR swaths at the same time. These are then processed to create a single wide swath image.

These various configurations of the SAR each have applications for which they are best suited. Thus, for example, the wide swath mode will be that most used for ice surveillance, while the standard mode will be preferred for crop monitoring and the high resolution mode will be used where most detail is required. 'Stereo' images can be obtained by imaging the same area at different incidence angles. The experimental modes are a group of high incidence angle beams which are not subject to the same performance specifications as the other beams. These beams are useful for monitoring of shipping and icebergs where the reduced return from the sea compensates for the deterioration of the beam shapes.

The SAR can be turned on at any time in the orbit to a maximum accumulation of 28 minutes data per orbit. The radar coverage available using the 500km swath mode over North america is shown in Figures 2 (one day) and 3 (three days). The gaps in coverage will be filled in on following days.

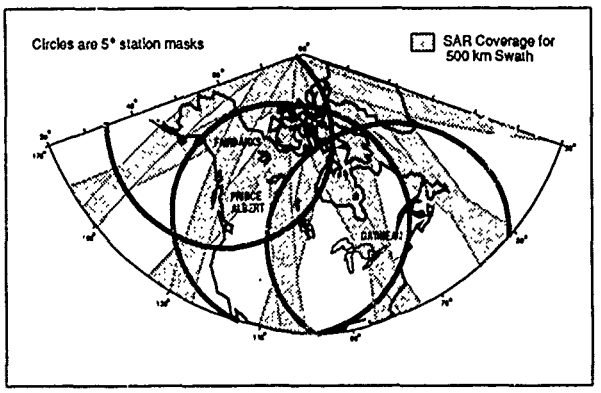


Figure 2 - One day coverage

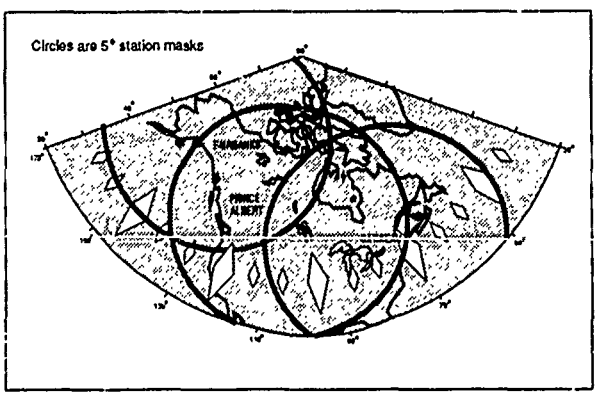


Figure 3 - Three day coverage

MISSION MANAGEMENT

Mission operations will be coordinated by a Mission Management Office (MMO) which will serve as the interface between the user community, the Mission Control Facility for commanding the satellite, and the ground reception and processing facilities. In Canada, the ERS-1 processing facility will be upgraded to handle the increased throughput of RADARSAT. In addition to coordinating and scheduling the national and international requests for SAR data acquisition, the MMO will monitor the entire distribution system. Thus, from the point of view of the project partners, it will be the executive office for the mission, responsible for implementing the provisions of the various agreements.

Because of the certainty of obtaining cloud-free images responding to user requests will be easier. However, because it now becomes practical to monitor real-time phenomena such as sea and ice conditions, the throughput demand of the delivery system is increased. Also, the number of modes and beam directions of the radar will need more attention to resolve conflicts. The satellite will carry two high speed tape recorders, each capable of storing 10 minutes of data, but to improve the chance of their surviving the five year mission in working order, their use will be limited to high priority data requirements.

The satellite has been designed so that it will have excess capacity to that required by the Governments who have participated in its construction. The excess capacity will be sold by the private sector distributor, RSI, who will market direct reception and data distribution agreements to interested parties.

Since interpretation of radar images is till in the development stage for some applications, the mission is also supported by a vigorous applications development program. This work is based on an airborne SAR experimental program and will be expanded to include data from the ESA ERS-1 satellite which is due to be launched in 1990, and it will continue beyond the end of the first RADARSAT mission. When applications technology reaches appropriate stage it will be transferred to the end users. In fact this process has already begun through agreements with Canadian regional remote sensing centres and special purpose information centres, such as the Ice Centre and the Crop Information Centre.

CONCLUSION

Canada has a record of success in pioneering communication satellite technology; this project is the first Canadian remote sensing mission. It will pioneer the use of the dawn-dusk orbit, the first multi-mode SAR on a satellite, the satellite radar mapping of Antarctica, and the commercial distribution of radar data. It is ambitious, but we have had the advantage of taking part in both the SEASAT and ERS-1 missions along the way. We have benefited greatly from international cooperation in the past and, through the RADARSAT project, we hope to contribute in our turn. One of our most important objectives is to make RADARSAT data available worldwide and in this way we are following the example of the optical satellite remote sensing projects. We are looking forward to joining with other countries and agencies and playing our part in making high quality remotely sensed data available to organizations working on global environmental problems.

RADARSAT: REVIEW OF PERTINENT ECONOMIC AND COMMERCIAL ISSUES

Marcel St-Pierre
Chief Economist

RADARSAT Project Office
Canadian Space Agency
Ottawa, Ontario

ABSTRACT

The expected availability of operational satellite synthetic radar data in the nineties, such as RADARSAT, offers an opportunity to fulfill the true commercial value of remote sensing from space. This opportunity will be enhanced by the high reliability of this source of information as an input to the decision making process affecting the management and monitoring of our natural resources and environment.

This paper reviews some pertinent economic and commercial issues relevant to RADARSAT, stresses the potential of the RADARSAT SAR system to meet commercial and operational user requirements and identifies the necessary conditions for the realization of the anticipated economic and commercial benefits identified for the RADARSAT project in Canada.

INTRODUCTION

Ever since the launch of the first Landsat satellite in 1972, Canada has been among the world leaders in the field of remote sensing. Recognizing the constraints imposed by its limited financial resources, Canada has, over the years, established a recognised worldwide competitive expertise by being selective in its choice and development of critical segments of the remote sensing sector. (EMR, Dec. 1988)

The development of an airborne synthetic aperture radar (SAR) and of systems to process the SEASAT data are concrete examples of this strategy which resulted in the Canadian remote sensing industry dominating both the world radar survey markets and the processing technology. Government approval of the RADARSAT project in the Spring of 1989 will reinforce the commercial and technical leadership of the Canadian industry in the field of remote sensing. RADARSAT will be an operational satellite carrying the most advanced civilian radar in the world. With RADARSAT, Canada will become one of the major world suppliers of operational satellite remote sensing data in support of the management of the Earth's resources and its environment.

Project approval also marked the beginning of a challenging and critical period which will determine the technological, operational and commercial success of RADARSAT. Between now and the launch of the satellite in 1994, key timely decisions will need to be taken to ensure that the Canadian remote sensing community, including the Governments, private sector and universities, will work in a coordinated and effective manner, to prepare users in various targeted market segments for the acceptance, use and integration of RADARSAT information data products in their decision making process.

The purpose of this paper is threefold: to review the factors and barriers which have affected the development of a viable commercial remote sensing market to date; to stress the potential of the RADARSAT SAR system to meet commercial and operational user requirements; and to identify the necessary conditions for the realization of the anticipated economic and commercial benefits identified for the RADARSAT project in Canada.

Review of the factors and economic issues which have affected the development of a viable operational and commercial satellite remote sensing information market.

Since the launch of the first Landsat earth observation satellite series in 1972, the remote sensing "information industry" has evolved considerably. Until recently, the Landsat system had remained primarily a research and development effort, although, in many circumstances, it has been presented and used as if it were an operational system. (KRS, 1988)

The Landsat technology was presented as a revolutionary cost effective tool to assist in the management, exploitation and monitoring of the Earth's resources and the environment. Large potential economic benefits were attributed to the use of that "technology" and, consequently, it generated high expectations among early users.

Unfortunately, being primarily R & D and not market driven, it became obvious that this system had not been designed to be cost effective and to meet the operational and commercial requirements of users.

As the technology failed to deliver all of its promises, the reality eventually caught up with the dream. Too much emphasis had been put on selling the technology and not enough on understanding and addressing user needs and requirements, and on delivering, in a cost effective and timely manner, the information contained in the imagery. The information provided was often not packaged and presented in a way that could be easily merged in the user decision making process. (KRS, 1988)

In a nutshell, the lack of commercial success in the early days of the remote sensing industry can be attributed to the failure to recognize that, at the user end, the information product must be usable immediately.

The Landsat experience has demonstrated that remote sensing information takes its full value, in most cases, only after it has been properly extracted and merged with other sources of information and integrated into other information systems.

Too many applications were developed without the target user in mind. Too often, remote sensing data suppliers failed to clearly demonstrate and prove the value and the cost effectiveness of the new or complementary source of information they were proposing. They failed, in many cases, to clearly explain to the user, in his own language, terminology and environment, how the proposed new source of information could be an asset to his business.

The early eighties saw the election of several Conservative governments. Reduction of the deficit became an urgent priority in most of these countries. These governments also held different views on what ought to be the respective role of government and industry. They strongly believed that governments should not unduly interfere in the market place and compete with the private sector. Directives were thus given to privatize and commercialize several Crown corporations and government agencies for which there were no longer a perceived public purpose.

Early in the process, remote sensing systems were identified as prime candidates for privatization. These programs involved large government expenses and moreover, the large economic benefits attributed to their use were mostly in the form of cost savings and increased profits to the users. Thus, it was only logical that the limited groups of users that benefit from these systems should pay full market value. Consequently, several governments came to the conclusion that such systems ought to be developed and operated by the private sector on a commercial basis.

In the mid-eighties a serious attempt was made in that direction. France created SPOT Image, a private company, to market and distribute the SPOT data products worldwide. The overall objective was to develop a system which would be commercially viable within a period of about ten years. The ultimate goal is for SPOT 4, by 1996, to be developed, built and operated on a full commercial basis without government subsidies. Until then, however, the French government is to pay for all the capital and operating costs for SPOT 1, 2, and 3.

The United States transferred the Landsat satellite system to a private consortium, EOSAT, in September 1985, following the enactment of the Landsat Commercialization Act of 1984 and the issue of a Request for Proposal (RFP) to the industry. The government agreed, on an interim basis, to subsidize the operation of Landsat 4 and 5 and the capital cost for Landsat 6. The goal is for the private sector to be able to operate Landsat 7, the follow-on system, on a fully commercial basis.

Unfortunately, the funding provisions negotiated with the private sector were never endorsed by Congress (Washington Remote Sensing Letter, March 1, 1987). This has created a climate of uncertainty regarding the interim funding for the operation of the Landsat system and its continuity, and, is most likely to result in a data availability gap between Landsat 5 and Landsat 6.

In view of the apparent lack of progress being made by EOSAT since 1986 to develop a market for remote sensing, and as part of the process of funding Landsat 6, the U. S. Congress directed, in 1988, the Department of Commerce to conduct studies for an Advanced Civilian Earth Remote Sensing System (ACERSS) to follow Landsat 6. KRS Remote Sensing, a Kodak Company, was awarded the contract to study the technical, financial, management and policy options for an ACERSS. The approach favoured by KRS was a market driven approach to support the selection and development of a system that would embody the ingredients that could ensure the commercial viability of the system selected.

The consultants reviewed and assessed the conditions under which an ACERSS could become commercially viable. One of the main conclusions of the study was that such a system would not be commercially viable before the year 2000 and, therefore, would need to be subsidized directly or indirectly by the Government on an ongoing basis.

In Canada, the possibility of the RADARSAT satellite system being developed, built, operated and financed completely by the private sector on a commercial basis was evaluated in 1986. It was concluded that in view of the current foreseeable state of the remote sensing market, a full commercial approach would not be feasible.

But as the project had been judged economically desirable for the country, the federal government decided to go ahead with it and to seek the involvement of international, provincial and private sector partners in its financing. On the other hand, the data distribution and marketing portion of the system being judged commercially viable, it was contracted out to a Canadian controlled consortium, RSI International Inc., which was incorporated in 1988.

RSI International Inc. was granted an exclusive license to distribute and market globally the RADARSAT SAR data and data products, with the exception of data collected for the use of the Canadian and United States Governments and for approved scientific research as defined in a Memorandum of Understanding between the two countries.

France, the United States and Canada have now realized that the market for remote sensing information needs to be further developed and as a consequence, is not yet ready to support a fully commercially viable industry. Government support is still necessary and is economically justifiable.

On hindsight, past experience leads to the two following conclusions:

(1) remote sensing systems were probably ready for "privatization" but not for "commercialization". A system can be economically viable from a public policy viewpoint without being commercially viable. The advantage of privatization, which does not necessarily exclude a government subsidy, is to bring more market and efficiency considerations into the development and operation of remote sensing systems and therefore improve the development of commercial by-products;

(2) the distinctions between the concepts of "economically" and "commercially" viable, and "privatization" and "commercialization" need to be better understood. In the case of the satellite remote sensing system, the market is not sufficiently developed to capture, through an appropriate pricing policy, all the economic benefits attributed to user groups. The level of benefits vary from one application to another and, to capture the consumer surplus realized by the different user groups, in different market segments, would require the use of a price discriminating policy, an approach which may not be permitted under existing USA laws and the United Nations Open Skies policy principles.

RADARSAT: A Turning Point in the Operational and Commercial Use of Satellite Remote Sensing Information data Products.

RADARSAT had been originally conceived as an end-to-end system to fulfill the needs for near real-time data information requirements such as ice, ocean, coastal and environmental monitoring. Since then, however, it has been modified to better cope with land application requirements. In short, RADARSAT is a practical solution to Canada's operational and regional monitoring requirements which call for a reliable and dependable system (not affected by cloud coverage and darkness) with fast turn around, frequent and timely repeat coverage.

RADARSAT is an earth observation satellite, carrying a powerful synthetic aperture radar (SAR) sensor, capable of imaging the earth through cloud and darkness. The radar will acquire data ranging from a fine-resolution and narrow-swath mode (10 meters, 50 km) to a lower-resolution wider swath mode (100 meters, 500 km). The standard mode will have a swath of 100 km and a resolution of 28 meters.

RADARSAT will transmit data from space to earth receiving stations for processing and transmittal to users in near-real time via communication satellites such as TeleSat's Anik. RADARSAT will cover Canada's Northwest Passage every day, all of Canada every three days and all of the world every 16 days.

The timeliness, accuracy and reliability of radar data is a distinct advantage over optical satellite sensors in operational and commercial use. It may be true that optical data from LANDSAT or SPOT are superior for many thematic mapping applications, but data acquisition is only possible on clear days, a major impediment to operational use in many geographic areas of the world.

In that regard, the following statistics are self explanatory: between 1986 and 1988 Canada has received and archived 245,559 SPOT images of which only 8.9% are completely cloud free images and 15.3% are called usable images (containing a limited amount of clouds). After three years of operation, there are still areas of Canada for which we have not yet been able to obtain at least a usable image. (EMR, update 1988)

In contrast, RADARSAT users can invest in systems to exploit radar data with the confidence that they can depend on a constant, accurate and predictable source of data supply. The probability of RADARSAT SAR being able to image a specific area at a predictable date and obtain a 100% quality image is 100%.

RADARSAT is not intended to replace the satellite borne optical sensors and/or the sensors available on airborne platforms . To the contrary, RADARSAT is there to fulfill specific market requirements that cannot be adequately provided by other sensors. In many cases, RADARSAT will turn out to be an excellent complementary source of information to other sensors and, at times, will be considered a reliable "second best" solution.

To optimize the potential for operational use of SAR data, several key concepts and systems have emerged during the RADARSAT planning phase. Several related key decisions were made prior to RADARSAT being approved and, some of them, have already been implemented or are currently being implemented. These are: the Ice Centre, the Crop Information System, the Ocean Information System and the Geology Information Centre.

(1) the Ice Centre located at Environment Canada is responsible for monitoring sea ice conditions in Canadian navigable waters and for producing daily forecast reports for different users . An Ice Data Integration and Analysis System (IDIAS) was developed under contract by MDA to merge and process in near real-time different digital data sets in support of sea ice analysis and sea ice forecasting conditions.

(2) the Crop information system was approved in 1986 and is currently being developed at the Manitoba Remote Sensing Centre. This system will have the capability to merge, integrate and pre-process different satellite data sets in a timely manner to allow improved crop forecasting accuracy. This system is currently under development in full cooperation with the end users (Canadian Wheat Board, Statistics Canada and the Department of Agriculture) to ensure that the system will meet their operational requirements.

(3) the Ocean Information System was approved in 1987. Its purpose is to allow access to different ocean related data bases and to facilitate the merging and analysis of the information through a GIS.

(4) finally, the concept of a Geology Information Centre is currently being developed. A contract has been awarded to assess the feasibility of merging and analyzing, through an appropriate information system such as a GIS, different geological related data sets.

In sum, RADARSAT has the potential for responding and offering a solution to many of the concerns and critics raised to date by users in the market place: predictability of data, repeat coverage, quick turn around, accuracy of information and higher spatial resolution. The instrument will be quite flexible and capable of meeting different user information requirements. RADARSAT's electronically steerable beam feature makes it possible to acquire information of the Earth's surface over different swath widths, different incidence angles and different spatial resolutions. On any orbit, it can access information within a swath up to 800 km (500 km nominal) with a resolution ranging from 10 meters to 100 meters depending on the swath width selected.

RADARSAT APPLICATION DEVELOPMENT AND INFORMATION SYSTEMS: Key to Operational and Commercial Uses of the Satellite SAR Data Information

Because RADARSAT has not been originally conceived as a pure R & D system but rather as a (pre) operational system, it

is imperative that an efficient, selective and practical approach to the development of the market for RADARSAT data products be elaborated and implemented to ensure both its operational and commercial success in Canada and abroad.

Operational and cost effective end-to-end systems need to be conceived and developed to respond, in a market driven manner, to the user needs and requirements for major potential applications. Although significant steps have been made in that direction since 1986 such as the Ice Centre, the Crop Information System and the Ocean Information System, it is important to remain alert and focus on our approach for meeting our objectives and goals regarding our ability to address properly user needs and requirements in a cost effective and timely manner.

Between now and the launch of the satellite in 1994/1995, the Canadian remote sensing community, including Government Departments and Agencies (at the Federal and Provincial levels), the private sector (RSI International, value added firms, software and hardware suppliers) and universities across Canada will need to work, in a timely, coordinated and effective manner, under the leadership of the Canadian Space Agency and the Canada Centre for Remote Sensing, to prepare users, in various targeted market segments, for the acceptance, use and integration of RADARSAT information data products to their operations. The merits of using the Canadian Advisory Committee on Remote Sensing (CACRS) structure as a coordinating framework should be given serious consideration. Existing relevant government programs at the federal and provincial levels such as the Radar Data Development Program (RDDP) of the Department of Energy Mines and Resources (EMR) need to be reviewed to ensure that they are fully in support of RADARSAT. Furthermore, maximum use should be made over the next few years of the SAR data availability from the EMR Convair 580, and the ERS-1 and JERS-1 satellites, to identify, develop or fine tune cost effective SAR data application in support of RADARSAT prior to its launch. Relevant pilot and convincing demonstration projects in support of RADARSAT major targeted market segments should be conceived as soon as possible.

More specifically, the following considerations will need to be kept in mind, during the development of RADARSAT data applications, to ensure the operational and commercial success of RADARSAT:

(1) priority for development of SAR data applications will need to be clearly defined both from an economic, commercial and scientific point of view;

(2) user needs and operational requirements in the targeted market segments will need to be fully considered and understood (alternative or current source of information, stock or flow type of information, cost, labour, reliability, frequency, investment, systems, timeliness etc..);

(3) the usefulness, value, advantages and cost effectiveness of the new source of information proposed for an operation compared to an existing one should be fully documented (reliability, quality, timeliness, ROI, pay back, cost-benefit, cost-effectiveness, product differentiation, references and champion cases); and,

(4) consideration as to whether or not the application envisaged would require the use or the development of a supporting information system such as GIS, or IDAIS.

The state of development of SAR applications is, in general, except for ice applications, far less advanced than that of optical sensors. But, based on experience and lessons drawn from Landsat on the need to always keep the end user in mind, it is nevertheless possible to meet the challenge of developing timely high quality RADARSAT SAR applications. However, this challenge will require an overall commitment by the Canadian remote sensing community at large. Scarce resources will need to be efficiently coordinated and focused on the development of selective applications which fully meet end user requirements.

CONCLUSION

The development of application products and services that fully meet end user needs and requirements for timely, cost-effective and easily usable information in the decision making process is the key to the establishment of a commercially viable remote sensing industry.

RADARSAT has the potential to satisfy adequately the operational and commercial requirements of several end user groups. This will only be feasible, however, through a concerted effort, by the remote sensing community at large, to increase industry awareness and understanding of how remote sensing can benefit industry information needs.

REFERENCES

Energy, Mines and Resources, Canada Centre for Remote Sensing, " SPOT Imagery Catalogue", 1988 update, Ottawa, Ontario, Canada.

Office of Technology Assessment,"Remote Sensing and the Private Sector, Issues for Discussion", Washington D.C., March 1984.

KRS Remote sensing, "Study of Advanced Civil Remote Sensing System", Vol. 2, Market and Financial Assessment, a report submitted to NOAA by Peat Marwick Main & Co. and Geodynamics Corporation with KRS Remote Sensing, A Kodak Company, August 1988.

Energy, Mines and Resources, Canada Centre for Remote Sensing, Radarsat: The Program after 1999, Ottawa, Ontario, Canada, Dec. 1988.

Washington Remote Sensing Letter, Vol. 6, March 1, 1987.

RADARSAT INTERNATIONAL INC.:

COMMERCIAL DISTRIBUTION OF RADARSAT PRODUCTS

D.R. INKSTER

INTERA TECHNOLOGIES LTD.
#2500, 101-6th Avenue S.W.
Calgary, Alberta T2P 3P4
Canada

Abstract

As a result of its all weather capability, RADARSAT offers a global imaging capacity that vastly exceeds that of the visual satellites. To ensure that this international capacity is translated into cost effective and timely products, the Canadian company "RadarSat International" (RSI) has been formed. RSI has been assigned an exclusive international license by the Canadian Space Agency (CSA) to market and distribute RadarSat products, excluding data required by the Governments of Canada and the U.S. for their own use. At present, RSI is organizing and preparing for its marketing, data production, and distribution roles.

This paper describes the initial organization of RSI, and the company's national and international plans for marketing, processing, sale and distribution of RadarSat products. An initial plan for the classes of product which will be available to users is given, and the mechanisms which will be developed to allow users to access the RSI system are described.

Key Words: RadarSat, RSI, commercialization, marketing

1. INTRODUCTION

When Phases C and D of the RadarSat program were approved in June of 1987, funding from the Canadian Government was subject to several conditions. In addition to finalizing relationships with Provincial and International partners, the RadarSat Project Office was required to develop an agreement with a private sector partner for the marketing and distribution of RadarSat products. It was argued that a partner with experience in the international remote sensing market could effectively deliver RadarSat products and services to that market. There was no suggestion that the RadarSat Program be a commercial program. That is, it was never intended that the initial capital costs of developing and launching the spacecraft and the on-going costs of operating the system be recovered from the sales of data, even using the most optimistic sales forecasts. However, partial cost recovery is possible. Therefore, it was proposed that a private sector partner could return some revenue to the Government to help offset the operational costs of the program, and, at the same time, develop a viable international business. Based upon these ideas, and the results of competitive proposals from several potential partners, negotiations began with the private sector. Over a two-year period, a business plan and agreement were developed, and the company RadarSat International Inc. (RSI) was established. At present, RSI is poised to take a major role in the international space-borne remote sensing business.

2. CORPORATE ORGANIZATION

RadarSat International is a new Canadian company, organized by the following Canadian aerospace and remote sensing companies.

SPAR AeroSpace Ltd.
INTERA Technologies Ltd.
DIGIM Inc.
MacDonald Dettwiler and Associates Ltd. (MDA)

With SPAR and MDA, this group includes much of the Canadian capability in space segment payload development and ground segment receiver stations, processing and display systems. INTERA and DIGIM are the two largest Canadian remote sensing data sales company, with extensive experience with radar data, multispectral data, and value-added services and products. These companies are committed to contributing funding, staff, and experience to RSI, and to making the company a technical and commercial success.

The Memorandum of Understanding between the Canadian Space Agency, NASA and NOAA calls for the identification of a U.S.-based distributor of RadarSat data. It is expected that such a distributor will be selected by RSI during the pre-launch period, and that that distributor will become a shareholder in RSI. Other partners may join the RSI team during the prelaunch period. To the extent that partners bring technology or marketing experience that is important to RSI business areas, they will be welcomed as equity participants in the company.

3. BUSINESS PLAN

RSI will build its business on the rights granted in the Memorandum of Understanding between the company and the Canadian Space Agency. These rights can be summarized as follows:

-an exclusive international license to distribute and market RadarSat data to all users, excluding the Canadian and U.S. governments. These governments have the right to acquire data for their own programs or for research efforts, provided that they do not sell or distribute data to others.

-access to about 35% of the data collection capacity of the satellite to collect data for its customers.

-an exclusive right to process RadarSat data required by Canadian Government departments and by other customers, using existing Government processing facilities and the company's own systems.

-an exclusive right to produce, market and sell to Canadian users, LandSat and SPOT data collected by Canadian receiver stations, commencing in 1989.

-a right to participate in other Canadian space-borne remote sensing missions and potential follow-on missions to RadarSat to the extent that they can be reasonably commercialized.

In return for these rights, RSI has obligations to conduct a vigorous and effective international marketing campaign, to produce data to meet market requirements, and to distribute data products and services according to a policy of non-discrimination. In addition, the company will return revenue to the Government in proportion to the level of data sales.

The success of the RadarSat program as measured by international sales of data will be determined by the perceived value of the products, and by competition from visual satellite programs such as SPOT and LandSat. These two visual satellite programs are both well established, with international networks of receiver stations and distributors, and with combined sales of approximately $30M U.S. during 1988. The principal markets for the visual satellites include the monitoring of renewable resources such as forestry and agriculture, the geology market, and the market which requires maps and mapping products. For example, Figure 1 shows recent sales of LandSat and SPOT products in the Canadian market. It is interesting that the market for the visual satellites does not include a significant element associated with monitoring of surface events that are rapidly changing, such as ice reconnaissance and ocean surveillance. Even in the crop monitoring area, LandSat and SPOT data is of limited use, since cloud cover often makes it impossible to collect the imagery that is required.

FIGURE 1
LANDSAT/SPOT MRKT (CANADA)

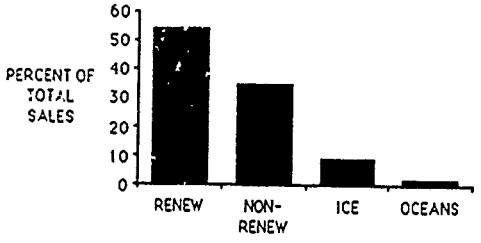
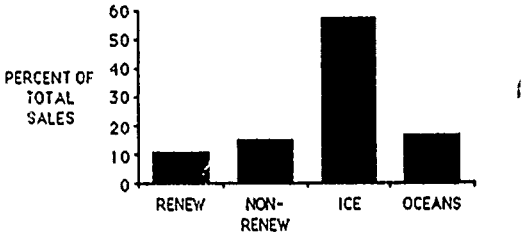


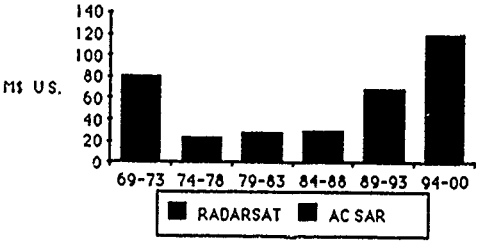
Figure 2 provides an interesting contrast with the market segmentation shown above. Here we see a forecast of the market for RadarSat products. The radar products have the advantage with respect to the visual products of being available regardless of cloud cover or darkness, hence will be useful in markets where routine monitoring is essential. On the other hand, the visual satellites with their richer spectral content will likely be relatively more successful than radar in the more traditional remote sensing market areas. These figures suggest that users will find data from RadarSat to be complementary, rather than competitive with data from LandSat and SPOT.

FIGURE 2
RADARSAT MARKET BREAKDOWN



Since the markets for radar and visual satellite data are so different, it is difficult to forecast the success of RadarSat based on forecast utilization of visual satellite products. A more realistic approach involves forecasting the market for RadarSat data using the experience of the international market for airborne remote sensing radar programs. Figure 3 shows the historical commercial market for airborne radar programs, including geology, mapping, ice reconnaissance, and ocean surveillance markets. Markets for research data, and programs conducted by national governments have been excluded from this figure. Viewed in this context, the RadarSat program is seen as a logical extension and growth of an existing market, rather than a new and untested market.

FIGURE 3
INTERNATIONAL SAR DATA SALES



The marketing strategy of RSI is based upon the unique market for radar data. During the period before 1993, an emphasis will be placed upon establishing agreements with international receiver station operators. These agreements must be in place well in advance of the launch of the satellite to ensure that, after launch, an efficient user interface is established, and that data can be received, processed and distributed effectively. During the same period, arrangements will be made with clients who require large volumes of data on a regular basis. By making pre-launch commitments, these clients will be assured of access to the data they require. During the two-year period before launch, a major effort will be made to develop the market for so-called "off-line" data. This market will include users who require data within a period of a few days or longer, for such applications as renewable and non-renewable resource monitoring and mapping. The key to success in this market will be to cooperate with existing vendors of visual satellite products. This cooperation could involve the development of a single network of agents and distributors, a common catalogue, and a common order system.

It is hoped that a system can be developed whereby potential users of RadarSat data around the globe can easily obtain information about the program, place their orders, and receive their data, without an unnecessarily large investment in marketing and infrastructure.

The distribution of LandSat and SPOT data in Canada in advance of the launch of RadarSat requires special attention in the RSI Business Plan. This function is very important, as it allows the company to develop a marketing, data production, and distribution system for remote sensing data well in advance of the launch of RadarSat. It is intended for RSI to take over this function during 1989. While the details of this new private sector distribution of remote sensing data are not yet finalized, the strategic approach is clear. First, it is the intent of RSI to minimize any impact on users associated with the change. During the transition, users will not see significant changes in the product line or level and type of service currently available from CCRS. Second, RSI intends to make a significant impact in the area of marketing. During the transition, a network of distributors will be established across Canada, and these distributors will eventually provide a new and, hopefully, more efficient user interface. Finally, using the distributors and a vigorous marketing effort, it is anticipated that sales of remote sensing products can be significantly enhanced. The company intends to use the existing CCRS production facility at Prince Albert to meet market requirements initially. As the market develops and as the product line evolves, this facility will be upgraded as required. Eventually, with the launch of RadarSat, it is expected that a fully integrated production, marketing and distribution system will be used to service the needs of users for all types of remote sensing data in the country.

4. RADARSAT PRODUCTS AND SERVICES

The details of the products and services which will be available to users from the RadarSat mission are currently being planned. For instance, RSI intends to conduct a User Requirements Study starting in 1989 and continuing until preliminary results are available from ERS-1. The intent of this study will be to define the types of products required by the market, and issues involving the timing of delivery of these products. Based upon the results of this work, the specifications will be set for the upgrades to the ERS-1 processor required for RadarSat.

Perhaps the simplest way to segment the products which will be available from RadarSat is to divide them into "on-line" products and "off-line" products. "On-line" refers to products required for near real-time monitoring applications, such as ice reconnaissance and ocean surveillance. These products will be ordered in bulk by large volume users, and will be rapidly processed and delivered by communication systems. For Canadian non-Government users of ice reconnaissance data, requirements for on-line data will most likely be integrated with AES data requirements. It is expected that AES will request acquisition and processing of most passes covering Canadian ice covered waters. These data will be processed and communicated to the AES IDIAS system within hours of collection. Rather than reprocess data for non-AES users, a system will be established whereby users acquire data from the AES archive on a commercial basis. International users will, in most cases, be served from their local receiver station. Where there is a significant on-line market, these stations will be capable of meeting the demand on a commercial basis.

Users can expect the availability of "off-line" products to be rather similar to current practices of LandSat and SPOT. In comparison to SPOT, RadarSat has even more flexibility in imaging geometry and product resolution Hence, users will have a choice of selecting the coverage specifically required for their application. A range of processing parameters will also be available. At least the following products will be available:

SGF : SAR geo-referenced fine resolution products, generated in swath section units.

SGC : SAR geo-referenced coarse resolution products, generated from SGF's in swath section units.

SSG - SAR systematically geocoded products, generated for user specified scenes.

SPG : SAR precision geocoded products, generated for user specified scenes.

The pricing policy for RadarSat products will be based upon market conditions for remote sensing products in the mid-1990's. Significant discounts will be offered to large volume on-line users, to ensure that RadarSat products are price-competitive with more traditional means of acquiring surveillance information. Off-line users will find RadarSat products to be available on a "Guaranteed delivery" basis at a price that is competitive with other satellite products. The intent of this policy is to establish a large and dedicated market which will support follow-on radar satellite missions.

RSI intends to establish a close and cooperative relationship with the international value-added industry. Value-added companies producing maps, forecasts, interpretations and other higher level products are considered to be part of the marketing function. They will have a greater ability to service local markets for all types of RadarSat products than RSI itself. Therefore, these companies will be supported with marketing and development efforts. More important, RSI will refrain from offering value-added products in competition with the rest of the industry.

5. SUMMARY

The decision of the Government of Canada to involve the private sector in the marketing, production and distribution of space-borne remote sensing data is an important one which will affect this area of technology dramatically for many years. Within Canada, all of the "stake-holders" in remote sensing will see changes. Starting with the LandSat and SPOT data streams, and continuing with RadarSat, users will notice a change in the marketing and delivery of products. Whether they deal with a local distributor or with RSI, they will deal with a party with a commercial interest in their application. The value-added industry, and hardware and software vendors will benefit from the presence of RSI. RSI will use value-added companies as part of the marketing organization, thereby enhancing their ability to win projects both nationally and internationally. The impact on equipment vendors will be based upon initial sales to Canadian clients followed by access on a preferred basis to the international market. Canadian companies which have historically enjoyed success in this market will find new markets for reception, processing, distribution, interpretation and analysis hardware and software as a result of the RadarSat program. Finally, the Government of Canada will be affected. Direct commercial feedback will be available on the value of different areas of remote sensing technology; this can be used to guide internal research and development. Equally important, a revenue stream will be returned to the Government in proportion to the sales of RSI to help off-set the operating costs of the RadarSat mission, and other remote sensing activities. While this revenue stream is a small portion of the total cost of the space and ground segments of RadarSat, it is expected to grow as remote sensing from space becomes more commercially viable.

If RSI is to succeed, it is essential that the stake-holders mentioned above maintain the proper perspective of the commercial viability, or rather the lack of commercial viability, of remote sensing from space. As noted earlier, RadarSat is not a commercial program, in the sense that it is not intended to recover all costs from the commercial sales of data. The relationship between RSI and the Canadian Space Agency (CSA) reflects this

reality. As the major partner, the CSA maintains control of the design, launch and operation of RadarSat, as well as the reception, archiving and cataloging of the data collected. In addition, through the Canada Centre for Remote Sensing, product development and application development research is funded and conducted. RSI, through its contribution of less than 5% of the total program costs, is a minor partner involved in the activities described above. RSI will only succeed if a strong partnership with the Government can be established. This partnership must involve an RSI role in research and development of RadarSat applications, to ensure that products can be developed that are suited to the international market. Further, it must include an RSI role in ground segment activities associated with tasking the system, and with reception, cataloging and archiving to ensure that RSI has the expertise necessary for its international marketing activities. This partnership, currently being negotiated with the CSA, will further enhance Canada's role as a major member in the international remote sensing community.

LAND APPLICATIONS OF RADARSAT

R. J. Brown
Canada Centre for Remote Sensing
Ottawa, Ontario

ABSTRACT

For most agricultural monitoring applications timeliness is by far the most important data characteristic. Landsat has proven to be inadequate for this task with its 16-day repeat cycle and vulnerability to cloud cover. Within Western Canada, where the time from crop emergence to harvest is typically of the order of 2 months, even a 16-day repeat cycle is marginal for any crop condition monitoring role. A satellite-borne synthetic aperture radar (SAR), with its all weather capability, holds great promise to meet this frequent data requirement. The proposed Radarsat SAR, in addition to having this all-weather capability, has also several important imaging parameters which greatly enhance the potential for using this sensor within agricultural applications. This paper will identify these features alone with some examples of the significance.

Keywords: Radar, Agriculture, SAR, RADARSAT

INTRODUCTION

Through the Radar Data Development Program (RDDP), within Canada, the national research and development efforts into the use of microwave imagery for the different application areas is coordinated. These different application areas include Ice, Oceans, Non-Renewable Resources and Renewable Resources. This paper will deal primarily with the use of synthetic aperture radar (SAR) imagery within Renewable Resource applications development and, in particular, within agriculture.

There are four distinct agricultural applications areas which are:

a. Landuse identification including crop type and area estimation
b. Vegetation vigour estimation which includes crop and forage yield prediction.
c. Landuse change monitoring which includes monitoring marginal landuse, soil conservation practices, changes in irrigation practices, etc.
d. Characterization of catastrophic events such as floods, tornadoes, hail storms, etc.

This paper will address how radar data and, in particular, Radarsat imagery may be used to meet the functional requirements identified above.

Through the RDDP our research program is a cooperative effort involving universities, government agencies and industry which include the University of Sherbrooke, Laval University, University of Guelph, University of Waterloo, University of Saskatchewan, the College of Geographic Sciences, the Ontario Centre for Remote Sensing, the Manitoba Centre for Remote Sensing, the Canada Centre for Remote Sensing, Agriculture Canada, Statistics Canada, the Prairie Farm Rehabilitation Administration, the Canadian Wheat Board, Intera Technologies Ltd., Terrain Resources Inc., Ducks Unlimited Canada, and the Saskatchewan Research Council. Each of these agencies is looking at specific aspects of how to use SAR imagery in agriculture so that Radarsat data can be used effectively when these data become available in the mid 1990's. Presently, the research program uses a combination of data from a ground-based scatterometer facility, which is operated by the University of Saskatchewan, and from the CCRS airborne SAR and scatterometers. These will be followed up by an analysis of data from the European Space Agency (ESA) Earth Resources Satellite (ERS-1), Japanese Earth Resources Satellite (JERS-1), the Shuttle Imaging Radar (SIR-C). Hence, by the time Radarsat is launched in 1994 we will know the capabilities of SAR for agricultural application and will be able to use the imagery from this satellite for resource management. Figure 1 lists the satellite SAR's which will be available in the next decade along with the expected launch dates and sensor\satellite characteristics. It can be seen from this table that there will be a large number of different SAR sensors in space during the next decade. This is vital for applications development, since any operational program using remotely sensed data, needs data continuity.

RATIONALE FOR RADARSAT DATA

In most agricultural application timeliness is an important and driving data characteristic. Consequently, SAR sensors, which have the capability to image under virtually any atmospheric conditions, offer the potential to play a major role in the remote sensing of vegetation. Visible and near infrared (VIR) imagery has been shown to provide good information for many applications but these sensors are restricted in their utility because it is necessary that cloud-free conditions exist at the time of data acquisition. Consequently, the Landsat Multispectral Scanner (MSS) and Thematic Mapper (TM) have not been used extensively in monitoring applications because of the uncertainly of data, which is unacceptable within operational programs. One notable exception, in this VIR wavelength region, is the use of AVHRR data for large area vegetation condition assessment. This sensor, with a swath width of approximately 2500 km, can image any agriculturally significant area of the earth daily. Hence it is possible to overcome the cloud problems with this frequency of coverage coupled with processing procedures to generate multi-date composite images which have been partially corrected for atmospheric effects (Holben, 1986). It is perceived that an agricultural monitoring system, using remotely sensed data would use a combination of imagery from VIR and SAR sensors at low and high resolution. For example, the low resolution AVHRR imagery, along with other data such as meteorological and field reports, could be used to identify those areas where there may be a problem. The higher resolution SAR or VIR would then be used to look at the areas of concern in more detail. In some cases, however, the SAR has a unique role to play which cannot be fulfilled using VIR imagery. For example, in the estimation of soil moisture (an input to crop yield models) the SAR can play a major role, In particular, it is felt that the Scansar mode of Radarsat will be of great value in the estimation of soil moisture over a large area.

One important aspect of the use of remotely sensed data is often overlooked. That is, a system to preprocess the raw imagery, to correct for radiometric and geometric distortions, and to make these data available to the user agencies in a timely fashion. This has been addressed already and a remote sensing Crop Information System has been installed at the Manitoba Remote Sensing Centre as a joint project between CCRS (federal government) and the provincial Department of Natural Resources. Presently this system is using primarily data from the AVHRR sensor but has been designed to handle SAR imagery when these become available and to integrate the SAR and VIR data sets.

The ensuing sections will present some of the results from the current research programs and illustrate the capabilities of Radarsat SAR for agricultural applications.

RADARSAT CHARACTERISTICS AND APPLICATION AREAS

Radarsat will be in a 792 km high, dawn-dusk polar orbit. The SAR sensor (C-band, horizontal polarization transmit and receive) has five main operational modes which are given in Table 2. One important feature of Radarsat is the 500 km accessibility swath. In the Standard Mapping mode, with a spatial resolution of 28 m, data may be collected over a range of incidence angles from 20 to 49 degrees. Actually, if the Experimental Beam mode is included data may be collected to incidence angles of 55 to 60 degrees. Thus, for applications areas where changes in incidence angles do not have an important effect upon the information content of the imagery, data may be acquired of an area of interest virtually when requested. For example, these data could be used for land use identification and area estimation. At C-band, crop type determination is relatively insensitive to incidence angle and hence for this application data could be collected over the full extent of the 500 plus accessibility swath.

Not only can these data be acquired regardless of atmospheric conditions, but for some applications, SAR data acquired earlier than would be the case for VIR data can be used. For example, canola can be reliably separated from the grains using SAR imagery about three weeks before it is possible using Landsat MSS VIR imagery (Brown et al., 1984). We expect similar results for the identification of potatoes. This is presently under investigation by one of the members of our cooperative research group. The ESA ERS-1 sensor will provide some useful information on the utility of SAR data, acquired at small incidence angles for crop identification The effect of incidence angle on crop type determination is very difficult to assess using aircraft data because of the rapid change of this parameter with ground range on these data. The only good satellite SAR data that can be used to assess the incidence angle effect was acquired with SEASAT (incidence angle = 23 degrees) but this sensor operated at L-band which is a wavelength about 3 times greater than the 5.66 cm wavelength of Radarsat. At the longer L-band, there will be more penetration of the vegetation canopy and consequently, depending upon the vegetation density, the SAR will be measuring a combination of radar backscatter from the vegetation surface, soil surface and vegetation structure between the soil and the top of the plants. Hence these data are not a good indicator of the expected performance of Radarsat.

The Scansar and Wide Swath modes, with reduced spatial resolutions of 35, 50 or 100 m, should be ideal for large area soil moisture estimation. Dobson and Ulaby (1986) have reported that the radar backscatter is most sensitive to soil moisture at incidence angles between 10 and 20 degrees and at C-band. We have found, using aircraft data, that at C-band that there is substantial sensitivity of the radar backscatter to soil moisture at incidence angles of 45 to 60 degrees. For example, we have imaged two adjacent fields that have recently been planted to potatoes and had not emerged yet. One of the fields had been irrigated while the other had not. The radar backscatter from the two fields was substantially different. The one with the greater soil moisture had a substantially larger value of the radar backscatter. It should be noted that, at C-band, the SAR is essentially responding to the soil moisture in the top 5 cm of soil. Thus, there is a need to develop the models to relate these surface moisture values to moisture at the root zone in order to use these data in yield models. In other cases, it is exactly the surface moisture that is required. For example, to determine whether there is sufficient moisture for emergence and the beginning of crop development the surface soil moisture is required. Part of our research program is examining the quantitative relationship between radar backscatter and soil moisture and assessing the contribution of surface roughness to the radar backscatter magnitude. This is

being done using a combination of the ground-based scatterometer facility, to get a precise relationship between radar backscatter and surface parameters, and airborne SAR to extend the analysis over a larger region (Pultz et al., 1989).

It may be possible to also use the Scansar data to monitor vegetation condition over large areas to complement that information supplied by the AVHRR sensor. It has been shown that there is a relationship between the radar backscatter and leaf area index (Ulaby et al.,1984, Le Toan et al.,1986). We are continuing work in this area, also with the ground-based scatterometer and airborne SAR to extend these measurements to other areas, crop types and soil conditions and are confident that SAR may be used in the future to assess vegetation condition. Another approach which is being pursued is to relate texture to vegetation conditions. We have found substantial different textural information in fields of different vegetation condition.

The SAR, operating in either the Standard Mapping or Fine Resolution modes could be used to monitor catastrophic events such as floods. The actual mode of operation would depend upon the spatial resolution requirements. SAR imagery is by far the best data source for this purpose because of the certainty of acquiring the data. Normally, in such situations imagery is required as soon as possible after a happening. Considering the 500-600 km accessibility swath most areas within Canada can be imaged within two days and in the worst case within three days.

CONCLUSIONS

The multiple modes of operation of the Radarsat sensor suggests that this SAR will be the most useful satellite SAR to date for many applications. The capability of acquiring data at resolutions from 7 to 100 m and over swath widths from 40 to 500 km covers application area from large area monitoring to detailed examination of small areas.

ACKNOWLEGMENTS

I would like to thank Mrs. G. Ferguson for setting up the paper on the output printer, Dr. S. Parashar for critically reviewing the paper and all of the members of the agricultural research team for their inputs to the papsr.

REFERENCES

Brown, R.J., B. Guindon, P.M. Teillet, and D.G. Goodenough, "Crop Type Determination From Multitemporal SAR Imagery", Proc. of Ninth Can. Sym. on Rem. Sens., St. John's, Nfld., pp 683-691, 1984.

Dobson, M.C., and F.T. Ulaby, "Active microwave Soil Moisture Research", IEEE Trans. Geoscience and Remote Sensing, Vol GE-24, pp 23-26, 1986.

Holben, B.N., "Characteristics of Maximum-Value Composite Images from Temporal AVHRR Data", Int. J. Rem. Sens., Vol. 7, pp 1417-1434, 1986.

Le Toan, T., A. Lopes, and M. Huet, "On the Relationships Between Radar Backscattering Coefficent and Vegetation Canopy Characteristics", Proc. of IGARSS'84, Strasbourg, pp 155-160, 1984.

Pultz, T.J., R. Leconte, R.J. Brown, and B. Brisco, "SAR Response to Spatial and Temporal Variations in Soil Moisture and Vegetation", Proc, of IGARSS'89, Vancouver, B.C., 1989.

Ulaby, F.T., C.T. Allen, G. Eger III, and E. Kanemasu, "Relating the Microwave Backscattering Coefficient to Leaf Area Index", Remote Sensing of Environment, Vol 14, pp 113-133, 1984.

Table 1: Past and Future SAR Sensors in Space

Satellite/ Sensor	Launch Date	Freq./ Pol.	Incid/ Angles (degrees)	Reso'n (m)	Swath Width (km)
SEASAT	1978	LHH	23	25	100
SIR-A	1981	LHH	50	40	50
SIR-B	1984	LHH	15-60	17-58	20-40
ESA ERS-1	1990	CVV	23	30	99
JERS-1	1992	LHH	35	18	75
SIR-C	1992/ 1993	X,L,C (Quad Pol)	15-55	10-60*	15-90*
RADARSAT	1994	CHH	20-60	10-100*	50-500*
EOS	1996, 1998	X,L,C (Multiple Pol)	15-55	25-500*	50-700*
ESA ERS-2	1990's (Late)	CVV	23	30	99

*depending on mode

Table 2: Radarsat SAR Modes

Mode	Resolution (Looks)	Swath Width	Incidence Angle Range
Standard Mapping	28 m (4 looks)	100 km	20-49
Wide Swath	35 m (4 looks)	150 km	20-49
Scansar	100 m (7 looks) 50 m (4 looks)	500 km 300 km	20-49 20-39
Fine Res.	10 m (1 look)	50 km	20-49
Experimental	28 m (4 looks)	100 km	49-60

RADARSAT: Ice Surveillance in an Ice Information System Context

by

Lyn McNutt
Canada Centre for Remote Sensing
Energy, Mines and Resources
O'Connor Street
Ottawa, Ontario
K1A 0Y9
(613) 998-5916
(fax) 995-5021

John Falkingham
Ice Centre
Environment Canada 110
373 Sussex Drive, Block E
Ottawa, Ontario
K1A 0H3
(613) 996-4297

ABSTRACT

As developments have occurred in the use of remote sensing information for ice surveillance, they have been incorporated into the ice information system at the Ice Centre, Environment Canada (ICEC). The key to effective use of remote sensing data is its integration into a complete ice information system. The Synthetic Aperture Radar (SAR) data to be provided by RADARSAT will provide one of the key sources of satellite ice surveillance data to the ICEC. This data, along with other information sources, will be incorporated into the Ice Data Integration and Analysis System (IDIAS), and will be used to prepare the ice analyses and to initialise and update ice forecast models. Algorithms designed specifically for the extraction of ice information from remote sensing data will provide guidance products for the forecasters to be used in preparation of the ice information products.

The ICEC currently uses both satellite and aircraft remote sensing data to help create their ice information products. As the capabilities of the sensors and platforms have evolved, the ICEC has commissioned the development and use of new data collection systems, and the development, testing, and implementation of expanded ice information systems such as IDIAS and the Berg Analysis and Prediction System (BAPS).

In conjunction with these advances at ICEC, the Canada Centre for Remote Sensing, as part of the Radar Data Development Program, has been working on algorithm development specifically related to the use of SAR data in preparation for the launch of RADARSAT. This includes the development of ice classification routines, an ice velocity extraction algorithm, and the merging of SAR data with ancillary data sources to facilitate the incorporation of remote sensing information into ice models.

In addition to the increased volume of ice surveillance information which RADARSAT will provide, the use of this operational spaceborne SAR information with routine aircraft ice reconnaissance and ice models has the potential to provide significant statistics on ice processes. This will form an integral part of an Arctic Environmental Data Base for climatological monitoring.

This paper describes the nature of the expanded ice service to be ready for use with satellite SAR data and delineates the algorithms currently under development in support of this system. It then describes the benefits to be derived from the incorporation of RADARSAT data into the IDIAS system and the upgrades necessary to accommodate the increased data volume, including a description of the improved products which are currently envisioned. Finally, the paper will present ideas for the generation of ice climatological information as part of the routine data analysis and information extraction function.

Key Words: Radar, SAR, Ice, Information Systems, satellite remote sensing.

THE RADARSAT SYSTEM

S. Ahmed & H.R. Warren

RADARSAT Project Technical Office

and

M.D. Symonds and R.P. Cox

Spar Aerospace Limited, Ste. Anne-de-Bellevue, Quebec

SUMMARY

RADARSAT is a Canadian-led cooperative program with the USA to launch and operate a remote sensing satellite with a Synthetic Aperture Radar (SAR). The spacecraft is scheduled for launch into a sun-synchronous dawn-dusk polar orbit in 1994 on an expendable launch vehicle for a five year mission. The orbit altitude and inclination are 792 km and 98.5° respectively.

The right-looking SAR, with the unique ability to shape and steer the radar beam over an 800 km accessibility swath will provide daily Arctic coverage, three day coverage of the Canadian land mass, and 16 day coverage of the globe. A wide variety of swath widths, incidence angles and resolutions can be selected. The SAR can also be oriented to the left of the flight path to enable coverage of central Antarctica.

The objective of the RADARSAT program is to generate data of both applications and research value related to global ice, oceans, renewable resources and non-renewable resources.

The expendable launch vehicle will be of the medium class in NASA terminology. The spacecraft will have an ability to store the instrument data which can later be transmitted to an appropriate data acquisition station, thus ensuring global coverage.

1.0 INTRODUCTION

RADARSAT will be the first radar-equipped multi purpose remote sensing spacecraft designed with both the research and operational user in mind. The SAR swath, side-looking to the right of the satellite's path, can be electronically altered in width and spatial resolution to suit particular applications. By combining two co-located images taken with different incidence angles, stereoscopic data sets can be made.

These multiple modes of operation offer, at one extreme, a high resolution beam of 45 km width and, at the other, a lower resolution 500 km swath. The latter, termed the SCANSAR mode, provides the unique ability to cover nearly the entire Arctic, from 70° latitude to the North Pole, every day. The cumulative coverage in a three-day period enables total access to anywhere on the globe north of 48°N or south of 48°S.

Recently approved by the Canadian Government, the project will go forward as an international endeavour, with the United States providing the launch on a NASA medium class vehicle. There is also a partnership between federal and provincial governments with contributions from British Columbia, Saskatchewan, Ontario and Quebec. All the Canadian provinces will, however, participate through making use of RADARSAT's data. A group of Canadian and US companies is forming a consortium called RSI as a data marketing and distribution company.

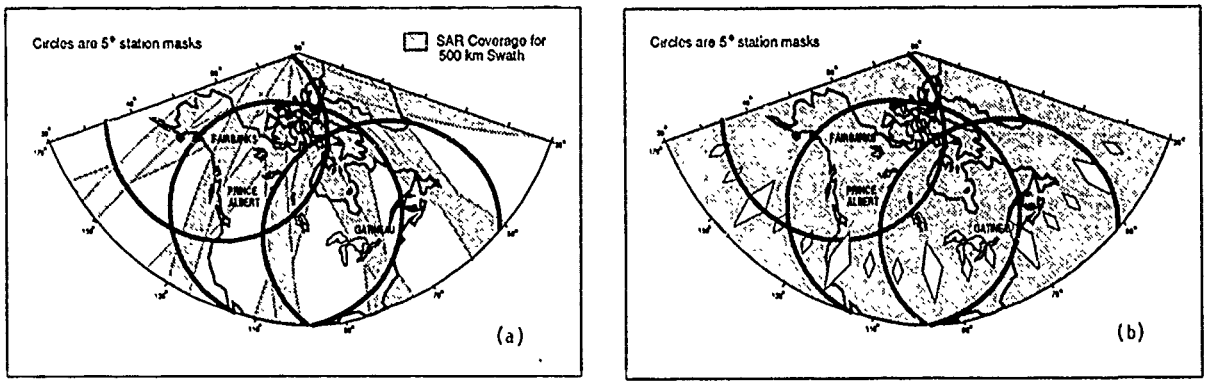
With its launch in 1994, the satellite's five-year design life is timed to follow the earlier ERS-1 and JERS programs [1], [2] and to precede the flight of a SAR on the NASA and ESA polar orbiting platforms.

2.0 MISSION CONCEPT

The current mission plan is to locate receiving stations at Fairbanks (Alaska), Prince Albert (Saskatchewan), and Gatineau (Quebec), with the mission control centre and main data processing and distribution centre in the Ottawa area. The ANIK communication satellites will be used for a high data rate back haul from the remote ground stations to the Ottawa centre. ANIK will serve also to forward the processed and interpreted imagery to the end user. For applications such as Arctic ice or sea conditions, where the data are urgently needed, the time interval between satellite overflight and transmission of the finished product to the user will be of the order of four hours.

A unique capability of the satellite is to swing the SAR beam from the right side to the left side of the flight path by a yaw manoeuvre of the spacecraft. The particular objective here is to provide complete coverage of the Central Antarctic land mass which would otherwise be inaccessible. This coverage will be obtained once during maximum ice conditions and once at minimum conditions.

The baseline orbit of RADARSAT is sun-synchronous at 792 km altitude with 98.5° inclination, and an ascending node of 1800 hrs. This is designed such that the SAR can provide the daily and cumulative three-day coverage as noted above and also has a 16-day global repeat cycle. The latter will provide SAR access to anywhere on the globe at least every 16 days. The coverage by the 500 km SCANSAR swath for one day and three days is illustrated in Figure 1. A



(a) 1 DAY COVERAGE and (b) 3 DAY COVERAGE

FIGURE 1.

high data rate tape recorder will be used to recover data from regions remote from the data acquisition stations.

The Canadian data acquisition, processing and distribution systems are expected to be based on those developed for ERS-1. Thus the RADARSAT program will benefit from experience gained during this and other preceding missions.

3.0 SYNTHETIC APERTURE RADAR (SAR) CHARACTERISTICS

The synthetic aperture radar (SAR), will operate at C-band (5.3 GHz) in a variety of commandable strip mapping modes, illuminating a swath to the right of the spacecraft nadir track. Operation of the SAR sensor will be allowed for a maximum of 28 min during each 101 min orbit. Capability will also be provided to record and store 15 min per orbit of SAR data aboard the spacecraft when it is not within view of a receiving station, for later transmission to ground.

In addition to the high duty cycle per orbit, and the very high reliability required to support its five-year operational usage, the RADARSAT SAR has three other unique features which distinguish it from previous or other presently planned spaceborne SARs. These include the ability to place and tailor each selected swath within an 800 km wide accessibility region through elevation beam steering and beam shaping; the flexible use of three radar bandwidths to achieve high and nominally uniform ground range resolution across the swath or to match its operation to special requirements; and the very high processor capacity which produces priority imagery at one-quarter real-time rates, and potentially achieves a zero backlog of unprocessed data over each 24 hour period.

The SAR is being developed in Canada by Spar Aerospace Ltd., with major subcontracts to Canadian Astronautics Ltd. and COM DEV Ltd. The 300 W average power radar transmitter will be based on the Dornier Systems/AEG ERS-1 high- power amplifier, with minor

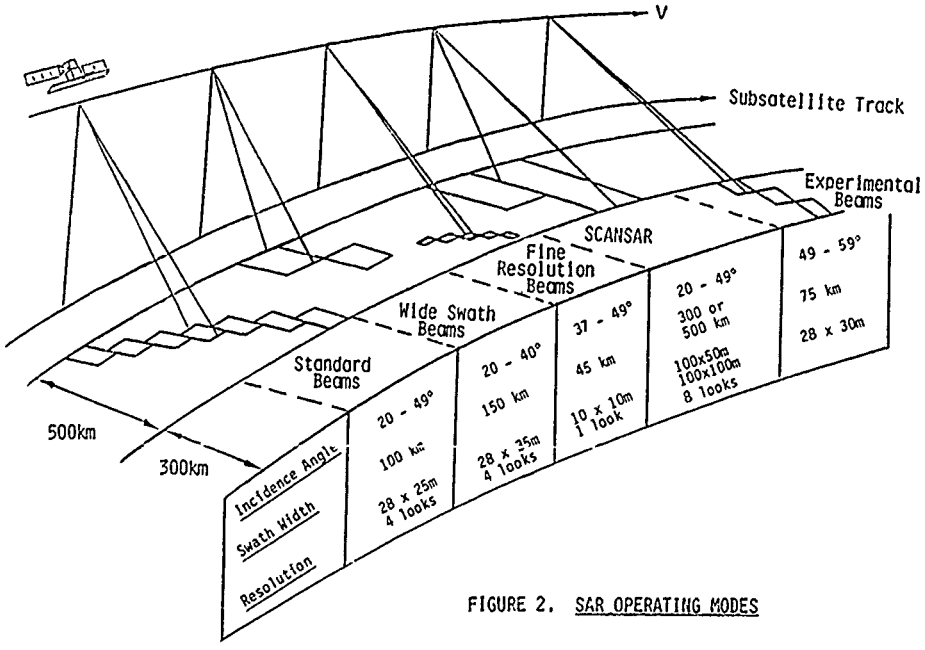


FIGURE 2. SAR OPERATING MODES

TABLE 1 Basic RADARSAT SAR Characteristics

Frequency	5.3 GHz
RF bandwidth	11.6, 17.3 or 30.0 MHz
Sampling rate	12.9, 18.5 or 32.3 MHz
Transmit pulse length	42.0 μs
Pulse repetition frequency	1270 - 1390 Hz
Transmitter peak power	5 kW
Transmitter average power (nominal)	300 W
Average radar data rate	73.9 - 100.0 Mb/s
Sample work size	4 bits each I and Q
Antenna size	15.0 x 1.6 m
Antenna polarization	HH
Antenna elevation phase quantization	8 bits

modifications. The antenna will be 15.0 m long x 1.6 m wide, and will be electronically steerable in elevation through a 29° primary region, and further 10° experimental region to form many different beams and beam shapes upon command. See Figure 2. These beams are realized using phase-only control of an elevation beam-forming network consisting of fixed power dividers to achieve a single aperture taper, but using 32 non-reciprocal ferrite phase shifters for excitation control. The radiating surfaces will be composed of slotted wave-guide panels formed in aluminum. Table 1 outlines the basic radar characteristics.

When the spacecraft is rotated in its Antarctic mode, both the thermal design and power raising capabilities may limit maximum SAR operation to 15 minutes in this mode. Analysis shows that this is adequate to provide coverage of the Antarctic.

The availability of the SAR for operation is constrained by the spacecraft power system, the payload thermal design and the hardware of the SAR. Typical continuous length of regular operation is 15 minutes and the maximum availability of data is 28 minutes in one orbital period.

4.0 THE SPACECRAFT SYSTEM

The RADARSAT spacecraft is about to commence its detailed design phase, with the design definition phase now completed. The spacecraft configuration, as seen in Fig 3 comprises two main modules, the Payload Module and the Bus Module.

The spacecraft has been configured to be launched by the NASA medium class launch vehicle. To fit the 15m SAR antenna within the 3.5m x 1.9m diameter shroud without compromising the electrical

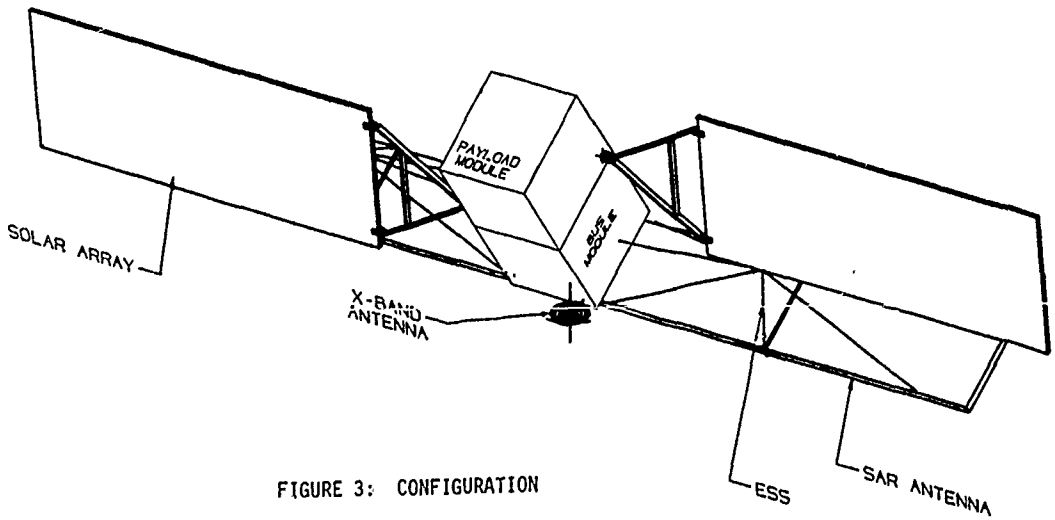


FIGURE 3: CONFIGURATION

performance, a four panel antenna design has been adopted, with the panels stowed parallel to the launch vehicle axis.

The launch vehicle injects the spacecraft directly into its polar orbit. After separation the solar array and SAR antenna are deployed. The SAR antenna Extendible Support Structure (ESS) is a key element in achieving reliable and accurate deployment necessary to mate the antenna sections and in maintaining the flatness of the SAR antenna under its varying thermal conditions.

Spar is the prime contractor for the space segment with responsibility specifically for the payload module, the SAR subsystem, and the integration, test and delivery of the spacecraft for launch. Table 2 summarizes the spacecraft characteristics.

TABLE 2 Spacecraft Characteristics

			Kg
Launch Mass	- Bus (including fuel)		1451
	- Payload		1475
	- Adaptor		55
	- Margin		171
		TOTAL	3152
Power	3 x 50 AH nickel cadium batteries		
Orbit	- altitude		792 km
	- inclination		98.5 deg
	- period		101.5 min
	- ascending node		1800 hr
Design Life			5 yrs

4.1 Payload Module

This module accommodates all payload elements, including the SAR and the X-band downlink antenna, the Earth sensor, and one of the two S-band telemetry antennas. Thermal control will be self-contained within the module, using active and passive techniques.

The on-board data handling system provides two X-Band downlink channels capable of operating at 100 Mbs-1 in the 8.215 to 8.400 GHz band. One of the two channels is used for realtime data while the second channel is used to downlink recorded data. A redundant tape recorder system is provided capable of storing 15 minutes of data at 85 Mbs-1.

Both channels (real time and recorded) are capable of simultaneous operation and both data streams include all image data and telemetry and ephemeris required to process the final image.

4.2 Bus Module

The bus will be purchased from a US company and selected for its proven reliability and heritage.

The hydrazine propulsion system is used to provide fine control of the spacecraft injection into its operational orbit and to maintain the altitude during its life. A 1C km altitude accuracy is necessary to achieve the required ground track repeat pattern.

The power system comprises a 2.5 kw solar array and three, 50 amp-hr batteries, which are sized to provide the required SAR operation throughout the 5 year life. Operation of the SAR is not required during eclipse. The eclipse experienced in the chosen orbit occurs for a short duration about summer solstice and then only over the south pole. During sunlit operation the payload will draw power from both the array and batteries and the batteries will be recharged during the interval that the radar is not operating.

The attitude control system uses reaction wheels, gyros, magnetorquers and sun and earth sensors to maintain the spacecraft pointing with an accuracy of better than ± 0.1° about all three axes.

An S-Band Telemetry Tracking and Command system (TT&C) is used to command and monitor the spacecraft "health" and allows for ground control of the SAR including in-orbit reprogramming of the pulse waveform, the pulse bandwidth and antenna beamshapes as well as the selection of other radar operation parameters.

5.0 GROUND SEGMENT

5.1 Data Acquisition and Processing

The SAR data may be downlinked in real time to an ERS-1 compatible ground station or stored in high speed tape recorders for subsequent downlink.

The data reception stations are equipped with 10m X-Band antennas which track the spacecraft above 5° elevation. The data is received at 100 MBPS, stored and subsequently retransmitted at a lower rate, on commercial communication channels to the Data Processing Centre (DPC). Here the data is processed to produce SAR images, which can then be transmitted to the end user. The objective is that the system will be designed such that no backlog of unprocessed data builds up. The time between data collection and reception by priority users will be of the order of 4 hours.

5.2 Mission Control System

The system is managed by a Mission Management Office (MMO), which has executive control of operations. The MMO receives requests for data from the various user groups, sets priorities for the operation of the spacecraft, and makes inputs to the detailed scheduling.

The detailed operation of the spacecraft is controlled by the Mission Control Centre (MCC) on a 24 hours a day, 7 days a week basis. The MCC is responsible for monitoring the status of the spacecraft, for scheduling the imaging sessions, as well as for implementing any necessary orbital and attitude corrections. The MCC also monitors the use of spacecraft resources such as high power amplifier, tape recorders, batteries and fuel. Detailed scheduling information is provided to the other elements in the network to ensure correct downlinking and processing of data. Information is fed back from the Data Processing Centre to the MCC so that the sensor operating parameters can be modified and image quality maintained.

Although a single TT&C station will suffice during the operational phase of the mission, additional ground stations will be required during and immediately after launch.

6.0 CONCLUSIONS

RADARSAT is the first Earth Observation system to be designed and developed in Canada. The many SAR operating modes make this system unique and pioneering in a global context both in terms of its versatility of operation and in its technology.

The RADARSAT system is being configured to provide a flexible response to a wide variety of user requirements. It will provide a regular, predictable fast turnaround data information system to the benefit of its users.

REFERENCES

1. Duchossois, G.; "Overview and status of the ERS-1 programme", Proc. Internat. Geoscience and Remote Sensing Symp., Zurich, 8-11 September 1986. ESA SP-254, pp 155-159 (1986).

2. Murai, S.; "Earth resources satellite (ERS-1) project in Japan", Proc. ISPRS Commission 1 Symp., Stuttgart, 1-5 September 1986. ESA SP-225, pp 487-488 (1986).

THE RADARSAT SYNTHETIC APERTURE RADAR SYSTEM

Anthony P. Luscombe

Spar Aerospace Limited,
21025, Trans-Canada Highway,
Ste-Anne-de-Bellevue,
Québec. H9X 3R2.
Canada.

ABSTRACT

The paper considers the Radarsat Synthetic Aperture Radar (SAR) system in terms both of its capabilities and its implementation. The system has been designed to be flexible in its operation, providing accessibility to areas over a wide range of incidence angles, and allowing imaging of any of a large number of swaths of different widths, and at a variety of spatial resolutions. This flexibility is achieved through a hardware design which incorporates elevation beamforming and digital pulse generation. The system also includes facilities for radiometric calibration of all image products. The paper describes the various standard modes of SAR operation, the approach to calibration, and the hardware components that make them possible. With these capabilities, the system should be able to provide the accurate quantitative radar scattering cross-section information required by a wide variety of users.

Keywords: Radarsat, Synthetic Aperture Radar (SAR).

1. INTRODUCTION

At the time that the feasibility studies for the Radarsat Synthetic Aperture Radar (SAR) were begun, the only civilian satellite SAR system that had flown was that on Seasat. The intention was to design a C-band system for Radarsat that could provide image quality equal to, or better than, that of Seasat's L-band SAR, but which would allow much greater choice in the swath position. Seasat had imaged a 100 km wide at an incidence angle of around 23° (Jordan, 1980). Radarsat was required to image swaths which were at least 100 km wide, but, additionally, to allow imaging of any specified point within a 500 km region from an incidence angle of 20° to at least 45°. A spatial resolution of about 25-30m in both ground range and azimuth had to be maintained for 4-look imaging anywhere within this region.

This set of requirements led to the introduction of two features in the system design:

a) The antenna had to have elevation beamforming capabilities so as to allow imaging of any of a set of defined swaths. These swaths occupied adjoining (slightly overlapping) sections of the 'accessibility region' (the full angular region that the SAR was required to be capable of imaging). Together, they covered the full width of that region.

b) The system had to be capable of generating a choice of pulses, each covering a different bandwidth. Depending on the incidence angle of the swath to be imaged, the appropriate pulse could then be selected to give the required (approximately 25m) ground range resolution.

With these two capabilities incorporated into the design, the system was capable of satisfying the resolution and coverage requirements with a set of seven beams, referred to below as "the Basic Beams". The additional operational flexibility also, however, gave the potential for a much wider variety of imaging modes. In particular, it allowed the SAR to image with various combinations of swath width and spatial resolution, and further sets of beams, referred to as 'the Wide Swath Beams' and 'the Fine Resolution Beams', were defined. The system is also able to operate experimentally in other modes besides those defined as standard: imaging at incidence angles greater than 50° is of particular interest.

Because the antenna design chosen for implementation of the beamforming capabilities inherently provided very rapid switching between beams, the system also had the potential for operating in a 'ScanSAR' mode. This mode, in which imaging is switched between two or more adjoining swaths, permits coverage of an even wider region in a single pass. ScanSAR operations for imaging 300 and 500 km wide regions were added to the set of standard Radarsat imaging modes.

In this paper, the various imaging modes of the SAR are described, and the hardware implementation that permits this operational flexibility is outlined. A different form of image is optimum for each application, but this system should allow any given user's requirements to be more closely satisfied than would be possible with a conventional fixed beam, single pulse design. For many of the potential applications, accurate radiometric calibration of the images is also necessary. The Radarsat calibration plans are outlined in the penultimate section.

2. IMPLEMENTATION OF THE RADARSAT SAR DESIGN

The operational flexibility of the Radarsat SAR is achieved principally through two components in the system: the pulse generation section of the low power subsystem, and the antenna, with its elevation beamforming network. This section contains brief outlines of each of these components, as well as the internal calibration subsystem used in the radiometric calibration procedures discussed in Section 4.

a) Digital pulse generator. The system must be capable of generating pulses with two different bandwidths (17.3 and 11.6 MHz) in order to provide reasonably uniform ground range resolution across the incidence angles covered by the set of Basic Beams. A third pulse covering a 30 MHz bandwidth, designed for fine resolution imaging, is also required. Digital technology has been chosen for the implementation of this component. This choice should enable a consistent form of pulse to be generated, and will allow experimental operation with pulses with other forms of modulation. The ground range resolutions obtained with each of the three current pulses are shown as functions of incidence angle in Figure 1.

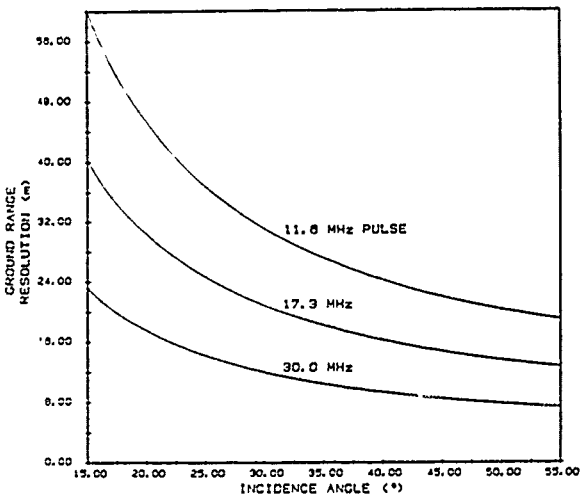


Figure 1: Ground Range Resolution

b) Antenna elevation beamforming network. The Radarsat SAR antenna is designed to provide the required elevation beamforming capabilities through the implementation of adjustable ferrite phase shifters at each of the elements across the antenna width (Zimcik et al, 1988). The coefficients for a number of standard beams will be stored in an on-board computer memory, but sets of coefficients for alternative beams can be transmitted from the ground for storage on-board. The hardware is inherently capable of rapid switching, enabling the system to operate in a ScanSAR mode (see Section 3.4) as well as a variety of conventional single swath imaging modes.

c) Calibration subsystem. A simple block diagram of the calibration subsystem is shown in Figure 2. The layout is designed for use in monitoring of the pulse characteristics as well as for tracking of internal gain variations in the system. During each imaging period, signals are tapped off before the antenna, attenuated and reinjected following the limiter and low noise amplifier. Calibration of these two components and of the receiver gain characteristic is performed at the start and end of each continuous sequence of pulse transmissions. Together, these provide the information used in the internal radiometric calibration updates.

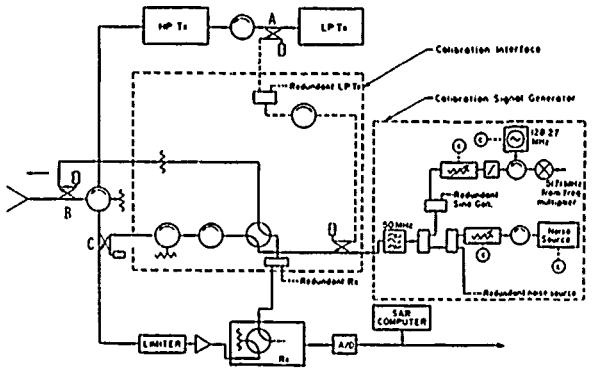


Figure 2: Radarsat Calibration Subsystem

3. THE RADARSAT IMAGING MODES.

Figure 3 is a schematic diagram of the coverage provided by each of the defined Radarsat beam modes. These modes are described in the following subsections.

3.1 The Basic Set of Beams

A set of seven beams have been defined to provide images which satisfy the original set of requirements. Each beam covers a swath of approximately 100 km and overlaps the next beam in the set by at least 35 km (except where other system constraints make this impossible). By reducing the swath width to near the specified minimum, the gain of the beam can be maximized and so optimum sensitivity to low scattering cross-section can be obtained. The large overlaps between swaths allow a better choice of beam to image any specific area.

For the purposes of normal imaging operations, each beam is associated with one of the available pulses: the 17.3 MHz pulse is used for the two beams nearest the subsatellite track, and the 11.6 MHz pulse for the remainder of the set. The ground range resolution, the data rate and the ground processing load are thereby kept reasonably uniform for all beams.

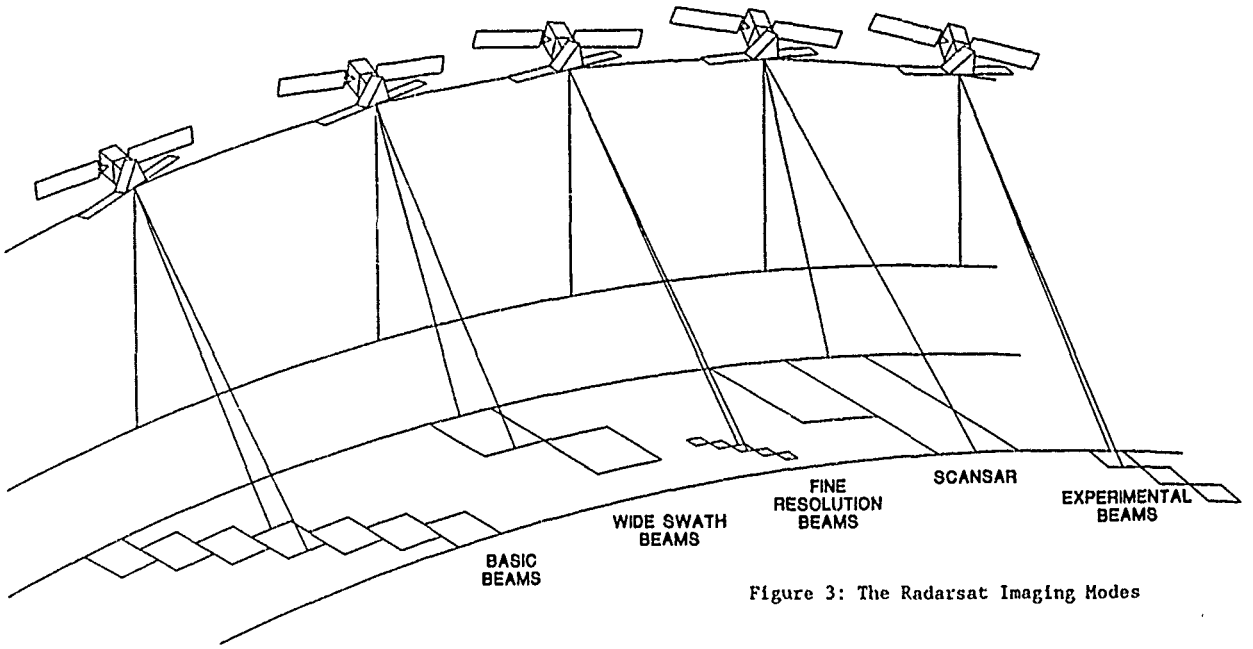


Figure 3: The Radarsat Imaging Modes

The form of image provided by the Basic Beams can be regarded as a compromise between wide coverage and fine resolution, optimized to give sensitivity to low scattering cross-section. In general, it is possible (subject to data rate limitations) to pair any of the beams with any of the available pulses to give alternative combinations of coverage and resolution. The following two sections describe the imaging modes that have been defined to provide the two extremes in the trade-off between these two performance parameters for conventional single swath imaging.

3.2 The Wide Swath Beams

For conventional imaging of a single swath (as opposed to imaging in a ScanSAR mode, as described in Section 3.4), the width of coverage is limited by the capacity of the data link to the ground. Two beams have been defined for swaths of maximum possible width, covering adjoining regions within the normal accessibility region of the SAR. Because a lower sampling rate can be used for operations with a narrower bandwidth pulse, both wide swath beams are defined for use with the 11.6 MHz pulse.

The swath widths are increased to around 150 km, compared with 100 km for the set of Basic Beams, but, because a lower bandwidth pulse is used, the range resolution is necessarily less fine (again, by a factor of approximately 1.5).

3.3 The Fine Resolution Beams

For the images produced with the Basic Beams, the spatial resolution is of about 25-30 m in each dimension for processing covering four looks. The same raw data could alternatively be processed to give a single-look image with an azimuth resolution of around 8 m. The fine resolution in range is obtained through pulse compression techniques. To improve this resolution, it is necessary to use a pulse with modulation covering an increased bandwidth.

When the Basic Beams are used to image any of the swaths at the far side of the accessibility region, the 11.6 MHz pulse would normally be employed. Some improvement in resolution (approximately in the ratio 3.2) could be obtained by using the wider (17.3 MHz) bandwidth pulse. In order to provide a significant enhancement in ground range resolution, a third pulse, with a bandwidth nominally of 30 MHz, was incorporated into the system. When this pulse is used for imaging of an area at 45° incidence angle, the resolution will be comparable to the single-look resolution in azimuth.

Since the wider bandwidth of signal must be sampled at a correspondingly higher rate, the data rate of the link to the ground limits the width of the swath that can be imaged. A set of Fine Resolution Beams have been defined for imaging at the far side of the accessibility region, where the incidence angle is highest and so the finest ground range resolution can be achieved. These beams are designed to cover the maximum swath widths (45-50 km) possible within the data rate limitations, but to provide higher gain and, hence, better sensitivity than the wider Basic Beams covering the same angles.

3.4 ScanSAR Imaging

For conventional single swath SAR imaging, the Wide Swath Beams represent the limit in swath width that can be achieved with the current data link. Even without the data rate limit, it would not be possible to image significantly wider swaths because of timing and range ambiguity constraints. If a modified form of operation known as ScanSAR is used, however, a much wider area can be covered in a single pass.

The principle of ScanSAR operation is to share the imaging time amongst a set of two or more distinct subswaths (Moore et al, 1981). Pulses are transmitted and returns received for a period in one subswath beam before operations are switched to another, and so on around the full set of subswaths. Each period must be sufficiently long to allow a synthetic aperture to be formed (i.e., to allow an image of a section of the subswath to be produced), but short enough that the successive periods in any one subswath cover adjoining or overlapping sections. (See Figure 4.)

Many combinations of subswaths, switching periods, and resolutions are possible, but two principal forms of ScanSAR operation have been defined for Radarsat:

a) Two-subswath ScanSAR, combining the coverage of the two Wide Swath Beams to give a total width of over 300 km.

b) Four-subswath ScanSAR, covering the full nominal 500 km accessibility region of the Basic Beams.

Because the raw data for any one subswath covers a period only a fraction of that from conventional imaging, there is a corresponding degradation either in the number of looks or in the azimuth resolution. For the two-subswath ScanSAR, an azimuth resolution of about 30 m could be obtained for one azimuth look and about 50 m for two looks. For the four-subswath ScanSAR, images with one and two azimuth looks can be expected to have azimuth resolutions of about 55 m and 100 m. (Luscombe, 1988)

Both proposed standard forms of ScanSAR are designed for use with the narrowest bandwidth pulse, but, because of the large change in incidence angle across the full swath, there is a wide variation in ground range resolution: from about 45 to 25 m for two-subswath operations and from about 45 to 20 m for four subswaths. The data can therefore be processed to produce images with a more consistent ground range resolution (nominally around 50 or 100 m), but with up to three times as many looks. For images with approximately 100 m resolution covering the full 500 km accessibility swath, for example, there would be around six looks in total.

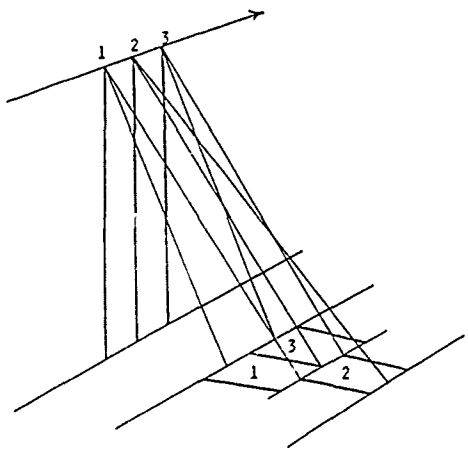


Figure 4: Two-Subswath ScanSAR Imaging

The beamforming capabilities designed into the system to provide the multiple standard beams can equally well be applied to the requirements of ScanSAR mode. Because the antenna design was inherently capable of switching rapidly between beams, it was only necessary to add additional memory to keep four sets of phase coefficients simultaneously accessible. Raw SAR data are generated in ScanSAR mode in the form of discrete blocks for each subswath, rather than as a continuous sequence of pulse returns, and so it is necessary to use modified processing algorithms to generate the images. (Wong, 1989).

3.5 High Incidence Angle Beams

The set of Basic Beams allow imaging of any region from 20° incidence angle to beyond 45°. For some applications, even higher incidence angles are regarded as optimum and the beamforming network enables imaging to be performed at these higher angles on an experimental basis. As the elevation angle of the beam increases, the timing constraints and range ambiguity effects become more restrictive and so coverage and performance estimates are tentative at this stage. Provisionally, however, an extra six beams have been defined for experimental imaging out to an incidence angle of nearly 60°.

4. RADIOMETRIC CALIBRATION

The radar system incorporates the internal calibration subsystem outlined in Section 2, which is designed to provide information for continuous updating of the radiometric measurements. This information represents only one aspect of the Radarsat radiometric calibration plan. The calibration procedures are also designed to make use of regular measurements from images of ground calibration targets, and of various functions and characteristics (beam patterns, temperature profiles, etc.) obtained pre-flight and updated periodically in-flight. All radar images produced with standard beams and imaging modes will be calibrated, and it is proposed to use an internationally-established calibration site to enable cross-calibration with other sensors.

5. SUMMARY

The optimum imaging characteristics (incidence angle, resolution, number of looks, swath width) depend very much upon the application. The Radarsat design should allow the system to be responsive to the requirements of a wide variety of users, and should enable more timely imaging of any specific area of interest than would be possible with a fixed beam system. The calibration procedures defined for the system are designed to provide as accurate quantitative radar scattering cross-section information as is possible within limitations of cost and technological development. This, in combination with the flexibility of operation, should make the Radarsat SAR truly "an economic tool for the nineties".

6. REFERENCES

1. Jordan, R.L., "The Seasat-A Synthetic-Aperture Radar System", IEEE J. Oceanic Eng., Vol. OE-5, pp 154-164, 1980.

2. Zimcik, D.G., Martins-Camelo, L. & Cowles, P.R., "Development of the Radarsat SAR Antenna", Canadian Aeronautics and Space Institute Journal, Vol. 34, No. 2, June 1988.

3. Moore, R.K., Claassen, J.P. & Lin, Y.H., "Scanning Spaceborne Synthetic Aperture Radar with Integrated Radiometer", IEEE Trans. Aerosp. Electron. Syst., Vol. AES-17, No.3, pp 410-21, 1981.

4. Luscombe, A.P., "Taking a Broader View: Radarsat Adds ScanSAR to Its Operations", Proc. IGARSS '88, Edinburgh, Scotland, Sept. 1988, pp 1027-32.

5. Wong, F., "An Algorithm To Access Scansar Data", submitted to Canadian Aeronautics and Space Institute Journal, 1989.

USER ACCESS AND PRODUCTION CONTROL OF CANADA'S OPERATIONAL SAR DATA PROCESSING FACILITY

Nigel Evans
MacDonald Dettwiler
13800 Commerce Parkway
Richmond, B.C., Canada V6V 2J3
Tel: (604)278-3411 - Telex: 04-355599 - Fax: (604)278-0531

Nigel Denyer
Canada Centre for Remote Sensing
2464 Sheffield Road
Ottawa, Ontario, Canada K1A 0Y7
Tel: (613)952-0202 - Telex: 0533777 - Fax: (613)954-2877

ABSTRACT

This paper presents a description of the Control and Management subsystem which provides user access and production control facilities at the Canadian Radar Data Centre Synthetic Aperture Radar Data Processing Facility (SARDPF). A complete description of subsystem architecture and functionality is presented.

The SARDPF is designed to receive and satisfy user requests from the user community for ERS-1 SAR Image Mode products.

The Control and Management subsystem provides the following functions within the SARDPF:

- an Order Desk facility,
- a Mission Management Office operator interface to support the identification and submission of data requests for new ERS-1 data acquisitions,
- a SARDPF operator interface to control all SARDPF operations,
- an ERS-1 raw and processed data archive catalogue,
- the control of an ERS-1 SAR processor subsystem, a GICS subsystem, a Transcription subsystem and a Data Path Switcher subsystem.
- control the distribution of products to the user community,

The paper concludes with a brief description of the planned upgrade of the Control and Management subsystem for the Radarsat phase.

KEY WORDS

ERS-1, Radarsat, SAR

1. INTRODUCTION

The SAR Data Processing Facility (SARDPF) is comprised of the following subsystems:

- Control and Management (CM) subsystem,
- Geocoded Image Correction System (GICS) subsystem,
- ERS-1 SAR Processor (SARP) subsystem,
- Product Transcription (TR) subsystem,
- Data Path Switcher (DPS) subsystem.

The CM subsystem provides all the required high-level system functions at the SARDPF. It is primarily designed to support the processing of SAR products derived from data acquired by the ERS-1 satellite in response to user requests submitted by the user community. In addition, the subsystem can support SAR product distribution over an Image Transfer Network (ITN) to Direct Users.

SAR products can be generated directly from raw data archive, the processed data archive or from newly acquired data which is transmitted from the Data Acquisition Facilities over a dedicated Data Transfer Network.

It is the intent of this paper to focus on the specific functions of the CM subsystem. The remaining subsystems are described in reference papers 1 and 2.

2. CM SUBSYSTEM ARCHITECTURE

2.1 Hardware Architecture

The CM subsystem architecture is designed to provide high reliability and availability. The host processor is a Digital Equipment Corporation MicroVAX II with 9 Megabytes of main memory (expandable to 17 Megabytes). The processor is supported by a number of peripheral devices and related control and data interfaces.

The processor is configured as follows:

- CPU board with console interface and 1 Megabyte of memory,
- 8 Megabyte memory board,
- spare slot for an additional memory board (to meet future expansion),
- disk controller supporting one 70-Megabyte disk drive plus a dual floppy disk drive,
- tape drive controller supporting one Fujitsu Computer-Compatible Tape (CCT) drive unit,
- disk controller supporting one Fujitsu 690 Megabyte disk drive unit.

An overview of the subsystem architecture is depicted in Figure 2-1 together with the interfaces to other SARDPF resources.

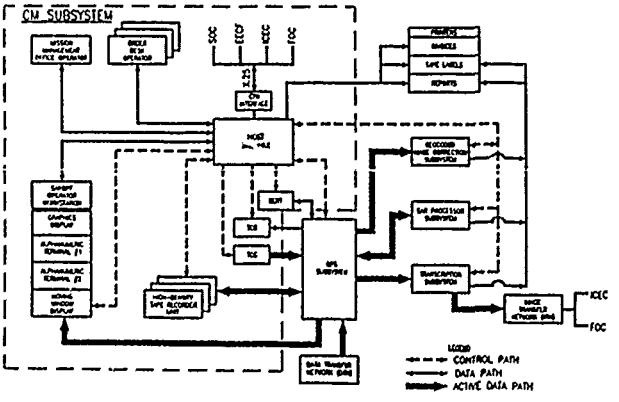


FIGURE 2-1 CM SUBSYSTEM ARCHITECTURE

The dual floppy disk drive is used to load vendor-supplied diagn packages. The drive is not used as part of normal system operation.

The 70-Megabyte disk drive is used to store all system application software tasks. In addition, all MicroVMS system software is resident on this disk.

CCT drive provides two functions:

- system backup operations to CCT directly from the MicroVM operating system,
- specific catalogue and product order backup operations as part of Management activities of the subsystem.

The Fujitsu 690-Megabyte disk drive is used to store information specific to the system application. Primarily the ERS-1 archive catalogue entries are stored on this disk.

2.1.1 Resource/Peripheral Control Interface

The High-Density Digital Tape Recorder (HDTR), Time Code Generator (TCG) and Time Code Reader (TCR) units interface to the host processor via the DEC UNIBUS. However, since the MicroVAX II is based on the DEC Q-BUS, a Q-BUS/UNIBUS bus interpreter is utilized.

2.1.1.1 HDTR Interface

The CM subsystem uses three identically configured HDTR units for recording and playback of raw and processed ERS-1 SAR data. All HDTR date input and output connections are made via the DPS subsystem under the control of the CM subsystem.

The HDTR units are Honeywell HD96 28-track units. The subsystem is able to support up to six of these units.

The HDTR control interface is manufactured by MacDonald Dettwiler and enables the selection of transport functions.

2.1.1.2 Time Code Reader/Generator Interface

The CM subsystem contains two identical Datum 9310 TCG/TCR units. One is configured as a TCG and the other is configured as a TCR. The TCG provides the IRIG-A signal that is recorded on the HDDT. The TCR extracts time code from HDDT. Both units are remotely controlled by the subsystem.

The TCR remote control enables selection of forward or reverse time code translation. The decoded time code is used by the subsystem for tape positioning.

2.1.1.3 Bit Error Rate Tester (BERT)

The BERT unit is controlled by the CM subsystem via an IEEE STD 488 interface. The unit is used for on-line maintenance tests of the HDTR units.

2.1.1.4 Moving Window Display (MWD)

An MWD unit is controlled by the CM subsystem via an RS232 serial interface. Up to two such units can be supported.

The MWD is used to display SAR imagery as it is being processed.

2.1.5 Control Transfer Network (CTN) Interface

The CM subsystem supports a CTN (X.25 protocol) communications link to external agencies. The interface supports the exchange of electronic mail messages between the SARDPF and the following agencies:

- Satellite Operations Centre (SOC),
- Earthnet ERS-1 Central Facility (EECF),
- Ice Centre Environment Canada (ICEC),
- Fisheries and Oceans Canada (FOC),

2.1.6 Operator Interface

- **Order Desk**

  Up to three Order Desk operator terminal interface are supported. It is assumed that the terminals are fully compatible with the DEC VT220/VT320 series of terminals.

- **Mission Management Office (MMO)**

  One MMO operator terminal interface is supported by the CM subsystem.

- **SARDPF Operator**

  The following terminal interfaces are supported by the CM subsystem and represent the SARDPF operator workstation:

  - colour graphics display,
  - two monochrome, alphanumeric terminals.

  The colour graphics display provides comprehensive status information in terms of a "mimic" representation of SARDPF resources and using colour to portray changes in state.

  One alphanumeric terminal display is dedicated solely to production control activities. The other is dedicated to management activities. This approach supports the future upgrade to separate processors.

2.1.1.7 Subsystem Interface

The CM subsystem controls the following SARDPF processing subsystems:

- GICS subsystem,
- TR subsystem,
- SARP subsystem,
- DPS subsystem.

The GICS, SARP and TR subsystems interface to the CM subsystem via an ethernet interface. The DPS subsystem in controlled via an RS232 serial interface.

The GICS subsystem is responsible for the generation of all the SARDPF geocoded products on CCT in response to work orders submitted by the subsystem.

The SARP subsystem is responsible for the generation of fine resolution georeferenced SAR data which is recorded on HDDT and represents the "processed data" archive. In addition, the SARP subsystem is responsible for the generation of special (SLC, SLD, MLD) and raw SAR products on CCT.

The TR subsystem is responsible for the generation of fine and coarse resolution georeferenced products on CCT; and over an Image Transfer Network (ITN) to the Direct Users. The products are derived directly from the ERS-1 "processed data" archive. In addition, the subsystem is responsible for all product quality control at the SARDPF with samples of product imagery output on a dedicated image recorder.

The DPS subsystem is responsible for performing all the necessary data path switching required to route image data between the GICS, SARP, TR subsystems and the HDTR units. Data path configuration commands are issued by the CM subsystem.

2.1.1.8 Printer Interface

The CM subsystem supports the following printers:

- invoice printer,
- tape label printer,
- reports printer.

The invoice printer is used solely for the generation of user invoices and is dedicated to the CM subsystem.

The tape label printer is shared by the CM, GICS, SARP and TR subsystems for the generation of HDDT and CCT tape labels.

The reports printer is shared by the CM, GICS, SARP and TR subsystems for the generation of SARDPF system performance and product reports.

2.2 Software Architecture

The CM subsystem is partitioned into two software functional components: Management and Control. This partitioning is a key design feature whereby the two components can be ported to independent host processors as part of a Radarsat upgrade program. The functions of the CM subsystem are described in terms of these components.

3. MANAGEMENT FUNCTIONS OF THE CM SUBSYSTEM

3.1 Order Desk Operator

The order desk operator is the primary interface between the user community and the SARDPF. The Order Desk operator is provided with the following operational features:

- to browse (query) the raw and processed ERS-1 data archive catalogue,
- to query SARDPF product prices,
- to enter, modify or query user details, ie. name, address, etc.,
- to enter, modify, query or cancel user requests for SARDPF products on behalf of users,
- to enter data requests for new ERS-1 data in terms of geographical areas.

3.2 Mission Management Office (MMO) Operator

The MMO operator represents the primary interface between the SARDPF and the Earthnet ERS-1 Control Facility (EECF). The operator is provided with the following operational features:

- to perform all Order Desk operator functions if necessary,
- to modify SARDPF product prices,
- to query, modify or cancel data requests submitted by the Order Desk operators,
- to submit data requests via the Control Transfer Network (CTN) to the EECF,
- to enter or query ERS-1 satellite instrument schedule parameters associated with data requests,
- to submit DAF assignment requests to the Satellite Operations Centre (SOC) via the CTN. Assignment requests are submitted in response to instrument schedules received from the EECF,
- to update the status of data requests and instrument schedules in response to electronic mail messages received from the EECF and SOC via the CTN,
- to submit postpass reports to the EECF,
- to monitor specific SARDPF resource status information.

3.3 SARDPF Operator Management Functions

The SARDPF operator is provided with a dedicated terminal to the Management component of CM subsystem. The operator is provided with the following operational features:

- to enter special product orders for the processing of SARDPF products which are derived from raw data on CCT. This data is not retained or referenced from the ERS-1 catalogue,
- to browse (query) the ERS-1 catalogue,
- to query or update default SAR processing parameters used in the product generation process,
- to ship completed product orders,
- to invoice shipped product orders,
- to enter ERS-1 raw data on HDDT for processing into the archive,
- to perform maintenance activities on the catalogue; ie. modification and deletion of entries,
- to perform purging operations on completed product orders.

3.4 Automatic Operations

The Management component software interfaces directly to the Control component. This interface is designed so as to enable one of the components to reside on another host computer without modifications to the interface. The interface can readily be configured to support DECnet/Ethernet communication between components.

The Management component performs the following operations on receipt of inputs from the Control component:

- update of the ERS-1 catalogue on the successful processing of newly acquired data,
- update of product order status as the order undergoes processing,
- update of SARDPF resource status.

4. CONTROL FUNCTIONS OF THE CM SUBSYSTEM

Control functions of the CM subsystem are summarized in terms of the operational features provided as part of the SARDPF operator workstation and the automatic operations provided by the Control component software.

4.1 SARDPF Operator Production Control Functions

4.1.1 Operator Equipment

The SARDPF operator production control workstation consists of the following:

- colour graphics display representing a "mimic" status of all SARDPF processing resources and external resources,
- a monochrome Moving Window Display (MWD) to monitor imagery as it is being processed,
- an alphanumeric, monochrome terminal to perform all SARDPF production control operations.

4.1.2 Operator Features

The SARDPF operator interface is designed so as to minimize data entry and thus workload.

4.1.2.1 Production Control Operations

The SARDPF operator is able to perform the following operations associated with product orders:

- cancel a selected product order,
- display the details of a product order,
- schedule a product order for immediate execution or for an assigned activation time,
- reschedule a product order,
- remove a product order from the processing queue,
- abort, suspend or resume product order execution,
- invoke and control the display of SAR imagery on a MWD for a designated product order,
- perform product and catalogue quality control operations,
- initiate windowing of products on an image recorder.

4.1.2.2 Retransmission Schedule Operations

For product orders which require data retransmission of raw ERS-1 data from the Prince Albert or Gatineau Data Acquisition Facilities, a retransmission schedule is automatically maintained. The SARDPF operator is able to perform the following operations:

- transmit the schedule to the SOC,
- print the schedule on the reports printer,
- display schedule details at the terminal.

4.1.2.3 ITN Product Distribution Schedule Operations

For product orders which require product delivery over an ITN, an ITN Product Distribution schedule is automatically maintained for each Direct User: ICEC and FOC. The SARDPF operator is able to perform the following operations:

- transmit the schedule to ICEC, FOC or both direct users,
- print the schedule on the reports printer,
- display schedule details at the terminal.

4.1.2.4 Auxiliary Data Entry

In order to support the processing of newly acquired ERS-1 data, the SARDPF operator is able to enter auxiliary data which is received from the EECF via the CTN. The following data can be entered by the operator:

- ERS-1 satellite ephemeris data,
- ERS-1 satellite time correlation parameters,
- ERS-1 SAR processing parameters.

4.1.2.5 Maintenance Operations

The SARDPF operator is able to invoke the following system operational tests:

- bit error rate testing of a HDTR unit,
- operational test exercising a specific processing string of the SARDPF. Ten tests are provided for selection.

4.2.1.6 Resource Monitoring and Update

The operator is able to update the "in service"/"out of service" status of resources at, or external to, the SARDPF.

4.1.2.7 System Performance Reports

The operator is able to invoke the print-out and purging of the following SARDPF system reports:

- System Event Log,
  outlining a summary of time-stamped system events over a specified time period,
- Performance Report,
  describing a summary of product orders processed by the SARDPF in terms of their completion status over a specified time period,
- SAR Parameter History,
  giving a comprehensive summary of critical SAR processing parameters over a specified acquisition time window.

4.2 Control Component Software Functions

The Control component software performs all the necessary resource allocation and control in response to the activation of a given product order. In addition, all processing requirements, in terms of work orders for the GICS, SARP and TR subsystems, are determined, distributed and monitored.

5. RADARSAT UPGRADE PHASE

For the Radarsat upgrade program where system throughput requirements will be higher, it is anticipated that the Control and Management components will reside on separate host processors. In this regard, the CM subsystem design has addressed the following design goals:

- a well-defined Control/Management interface which can support independent host processors,
- capability to support up to six HDTR units,
- capability to support up to six TCR units,
- capability to support up to six TCG units,
- capability to support a Radarsat SAR processor,
- capability to support up to two MWD units.

6. ACKNOWLEDGEMENTS

The authors would like to acknowledge the contribution to this paper of the other members of the MacDonald Dettwiler and CCRS project team responsible for the SARDPF.

7. REFERENCES

1. Kavanagh, P.F., Wong, F., Guindon, B., and Princz, G.J., "Canada's Fast Operational ERS-1 SAR Precision Processing and Geocoding System". See these conference proceedings, 1989.

2. Sack, M., Ward, J. and Princz, G.J., "Overview of Canada's ERS-1 SAR Data Processing Facility". See these conference proceedings, 1989.

TELEMETRY ANTENNA FOR RADARSAT

A. Kumar

AK Electromagnetique Inc.
492 Westminster Avenue
Dollard-des-Ormeaux
Quebec
Canada H9G 1E5

ABSTRACT

The paper describes a highly shaped beam telemetry antenna which can be used for the RADARSAT satellite payload. The antenna comprises a symmetric reflector of shaped profile and a circularly polarized feed horn. Three solutions of antenna are investigated, the first one that involves seventh-order polynomials, the second one that involves the geometrical optics principle, and the third one that involves modified geometrical optics principle. Computed radiation patterns are shown for 600 mm and 800 mm (dia.) shaped reflector antennas in free space.

INTRODUCTION

RADARSAT is a Canadian satellite which will map a region of the earth's surface using microwave synthetic aperture radar (SAR) techniques. This paper describes a shaped beam telemetry antenna suitable for re-transmitting the radar data back to an earth. The telemetry coverage area extends in all directions from nadir to the horizon as seen from the satellite. The X-band antenna beam is required to be circularly polarized, and to peak sharply on a 59.2 deg. off-axis annulus to provide compensation for range and atmospheric attenuation.

The antenna comprises a centre-fed shaped reflecor and a horn feed. This type of antenna has been designed and developed by Kumar [1] for the ERS-1 satellite. In the present paper, the author has shown that the axial ratio, bandwidth and gain of the antenna can be improved to meet the system requirements for the RADARSAT telemetry antenna.

FEED DESIGN

The feed design has been based on the approach developed by Kumar [1-3]. It comprises a crossed-dipole mounted in a feed horn. The use of a crossed-dipole with step cup ground plane provodes a simple and efficient method for achiving circular polarization.

Kumar [4-6] has shown that the dielectric lining or loading can improve the axial ratio and gain of a horn antenna. This technique has been reported by Oliver Clarricoats and Raghavan [7] in 1988. We have used a similar technique to improve the axial ratio and gain of the feed horn.

The feed horn consists of a piece of circular waveguide with steps. Given the dimensions of the horn, the boundary condition that the tangential E-field must vanish on the perfectly conducting surface, is used to set up an integral equation for the surface current density. The integral equation is then solved using the method of moments, and once the currents are found, the fields are readily calculated. This method gives more accurate results for small horns than the simpler aperture models, especially for the far-out and backlobe radiation. The agreement between experimental and theoretical radiation patterns is excellent [8-9].

We have designed a feed horn to increase the bandwidth of the antenna from 2% to 5%. The feed horn has been built and tested in the frquency range of 8.0 to 8.4 GHz. The return loss of the feed horn is better than -20 dB in the above frequency range as shown in Fig. 1. Also shown is the measured return loss obtained for the ERS-1 telemetry antenna feed. Figs. 2a and 2b show mounting arrangements of the feed horn with the reflector. Kumar [9] has described a detailed design and radiation patterns of the feed horn.

ANTENNA DESIGN AND RESULTS

The optimization of the different parameters defining the antenna reflector, to obtain the required shaped beam, has been carried out by means of software developed at AK Electromagnetique using the following approaches:
(1) seventh-order polynomials,
(2) geometrical optics, and
(3) modified geometrical optics.

Fig. 3 shows the shaped reflector, a stepped horn, a support rod and a coaxial cable. Profiles have been derived for a reflector of 600 mm diameter using the software. A computer-calculated radiation pattern using seventh-order polynomials is shown Fig. 4 [10]. Fig. 5 shows a computed radiation pattern (using geometrical optics) and an experimental one at 8.3 GHz [11]. Figs. 6a and 6b show computed radiation patterns of 600 mm and 800 mm reflector diameters, respectively. The shape of the reflector can be optimized to achieve desired radiation pattern for the RADARSAT antenna using modified geometrical optics [12].

CONCLUSIONS

Two prototype feed horns have been designed, built and tested in the frequency range of 8.0 to 8.4 GHz. Computer programs are developed to design shaped beam reflector at AK Electromagnetique. These programs can be used for the RADARSAT telemetry antenna.

REFERENCES

[1] Kumar, A.: "Highly shaped beam telemetry antenna for the ERS-1 satellite," Proc. IEEE Montech'86 Conference on Antennas and Communications, Montreal, Quebec, Canada, IEEE Cat No. TH0156-0, October 1986, pp. 46-49.

[2] Kumar, A.: "Theoretical analysis of a dipole fed long backfire antenna," ICAP 83, IEE Conf. Pub. 219, 1983, pp. 136-140.

[3] Kumar, A., and Hristov, H. D.: "Microwave cavity antennas," Artech House, USA, 1989.

[4] Kumar, A.: "Experimental study of a dielectric rod enclosed by a waveguide for use as a feed," Electronics Letters, Vol. 12, December 1976, pp. 666-668.

[5] Kumar, A.: "Dielectric-lined waveguide feed," IEEE Trans. Ant. and Propag., Vol. AP-27, March 1979, pp. 279-282.

[6] Kumar, A.: "Dual band dielectric loaded horns," Proc. IEEE ICEC, Montreal, 1984.

[7] Olver, A.D., Clarricoates, P.J.B., and Raghavan, K.: "Dielectric cone loaded horn antennas," IEE Proc., Pt. H, Vol. 135, 1988, pp. 158-162.

[8] Kumar, A.: "Telemetry antenna," Proc. IEEE Seventh Annual Benjamin Franklin Symposium," N.J., March 1989.

[9] Kumar, A.: "Telemetry antenna feed horns," Research Report AK-11-88, Quebec, 1988.

[10] Tucker, R.: Private communication.

[11] Do Boi, H. : Private communication.

[12] Kumar, A.: " Theoretical study on telemetry antennas," Research Report AK-12-88, AK Electromagnetique, Inc., D.D.O., Quebec, Canada, 1988.

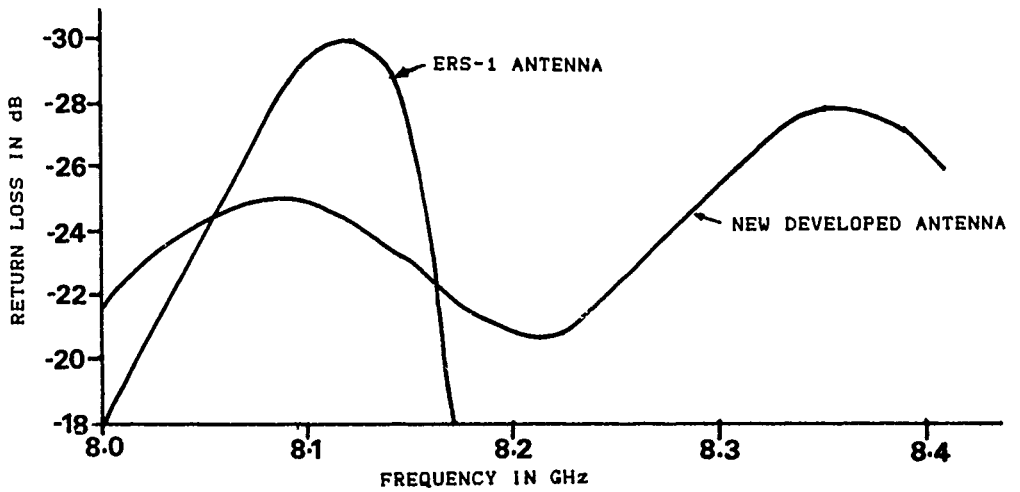


FIG. 1 RETURN LOSS VERSUS FREQUENCY FOR TELEMETRY ANTENNAS

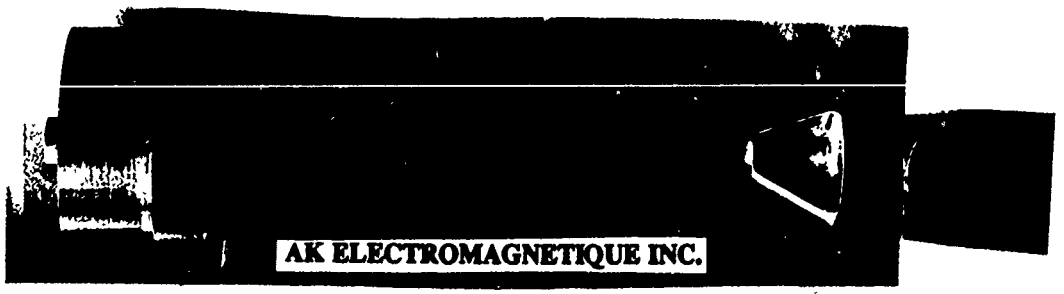


FIG. 2a NEW DEVELOPED FEED HORN WITH MOUNTING ARRANGEMENT



AK ELECTROMAGNETIQUE INC.

FIG. 2b NEW DEVELOPED FEED HORN WITH MOUNTING ARRANGEMENT '''

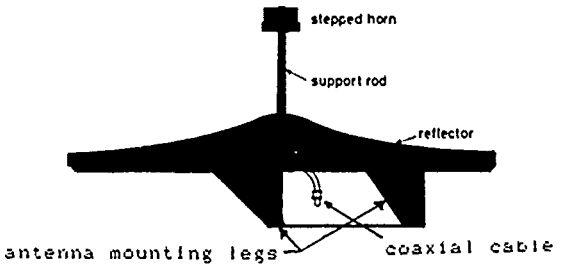


FIG. 3 ANTENNA ASSEMBLY

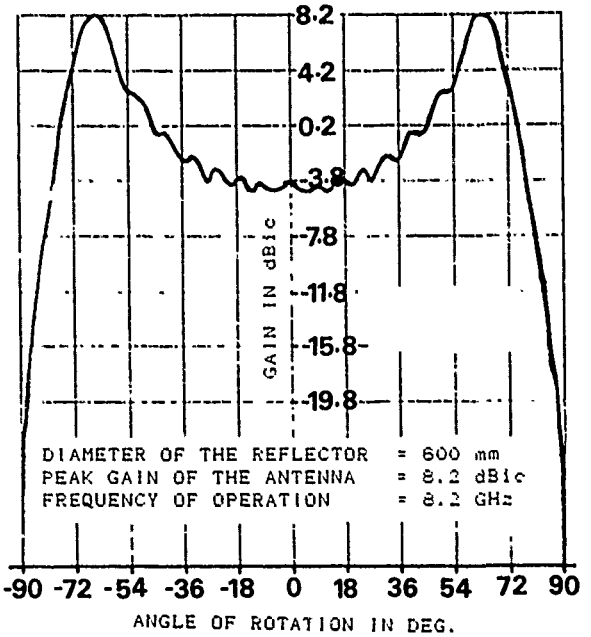


FIG. 4 A COMPUTER-CALCULATED RADIATION PATTERN USING SEVENTH-ORDER POLYNIMIALS

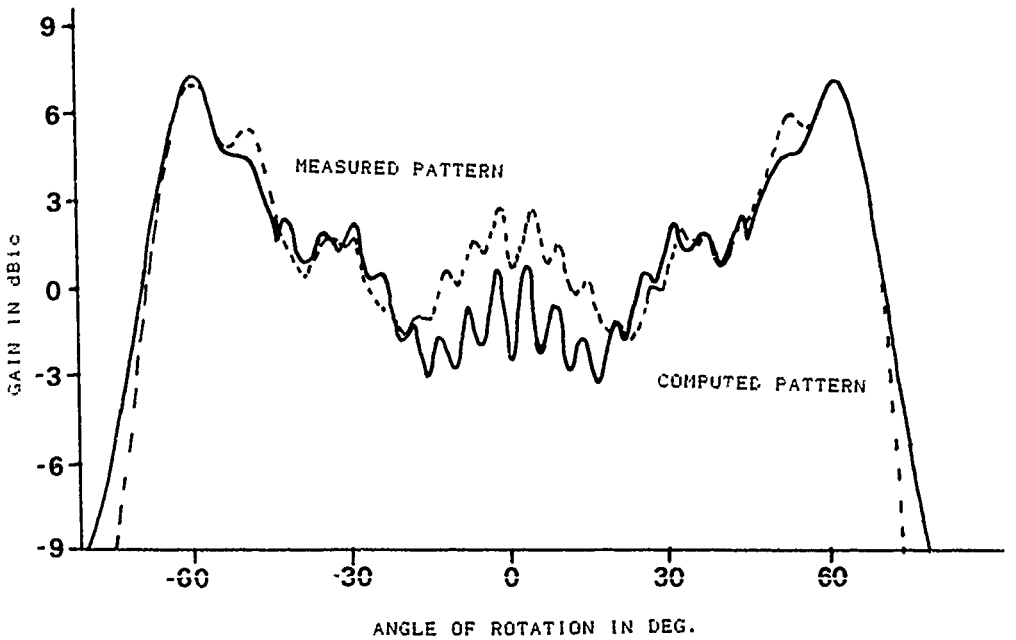


FIG. 5 COMPUTED (USING GEOMETRICAL OPTICS) AND MEASURED RADIATION PATTERNS OF A 600 mm DIAMETER SHAPED REFLECTOR ANTENNA AT 8.3 GHz.

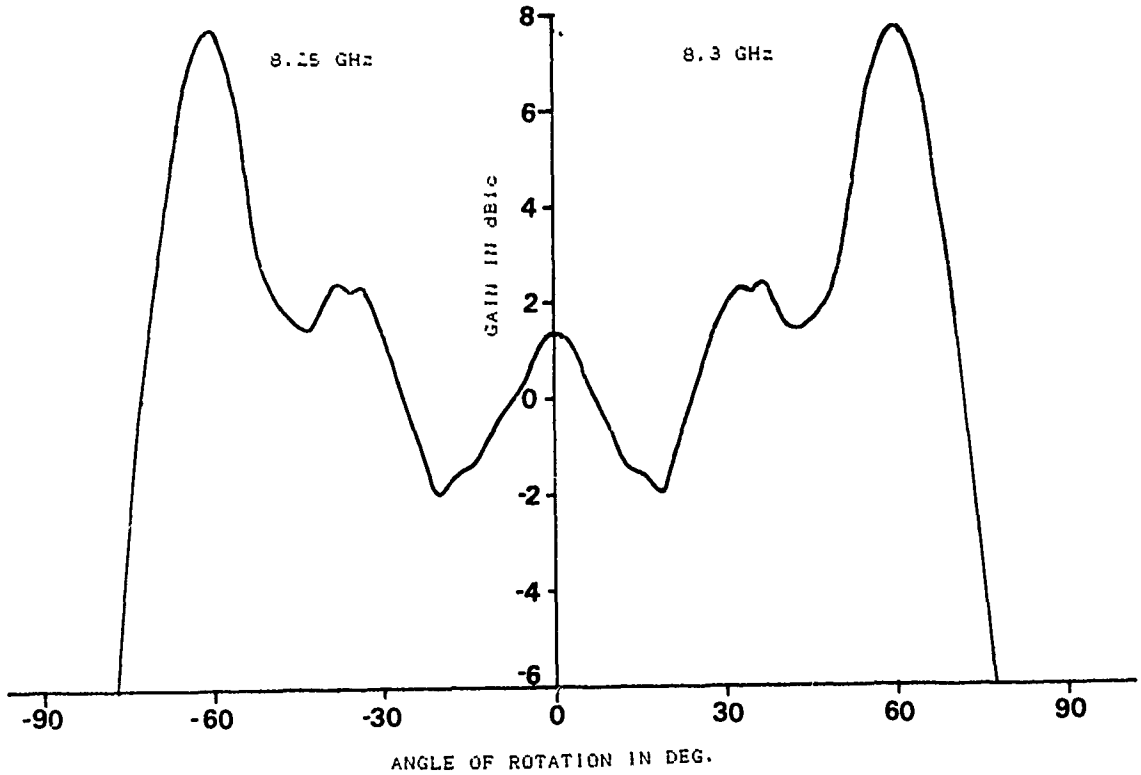


FIG. 6a COMPUTED RADIATION PATTERN OF A 600 mm DIAMETER SHAPED BEAM REFLECTOR ANTENNA USING MODIFIED GEOMETRICAL OPTICS AT 8.25 AND 8.3 GHz.

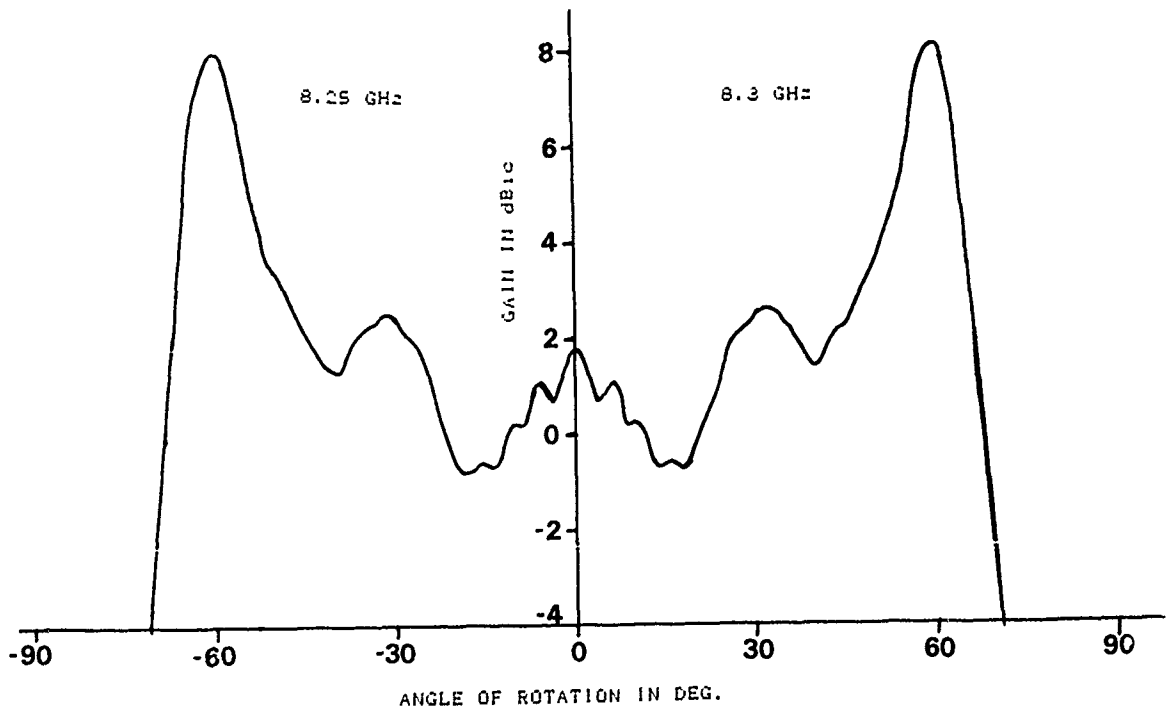


FIG. 6b COMPUTED RADIATION PATTERN OF A 800 mm DIAMETER SHAPED BEAM REFLECTOR ANTENNA USING MODIFIED GEOMETRICAL OPTICS AT 8.25 AND 8.3 GHz.

THE NEED FOR SAR CALIBRATION

A. Freeman

Jet Propulsion Laboratory
4800 Oak Grove Drive
Pasadena, CA 91109
USA

Tel: (818)354-1887
FAX: (818)354-3437
Telex: 67-5429

ABSTRACT

Microwave Remote Sensing with Synthetic Aperture Radar (SAR) sensors is a field now approaching maturity. In the next decade, several spaceborne SAR systems, including Radarsat, E-ERS-1, J-ERS-1, SIR-C/X-SAR and the EOS SAR, will be put into orbit. The streams of SAR images produced by these systems will pose a considerable challenge to the microwave remote sensing community in terms of understanding their information content, regarding the properties of the Earth's surface. There is an urgent need for the data from these sensors to be fully calibrated, so that adequate models for the relationship between radar backscatter and geophysical parameters can be developed.

In this paper the problem of routinely calibrating SAR data is discussed, with emphasis on some of the calibration requirements imposed by applications scientists. Within the framework of the Committee on Earth Observing Sensors (CEOS) working group on SAR Calibration these problems are being addressed by the engineers and scientists responsible for calibrating each national sensor.

KEYWORDS : SAR, Calibration

I. INTRODUCTION

Synthetic Aperture Radar (SAR) was first developed in the 1950's [1], for long-range mapping from airborne platforms. By the present date, the field has matured to the point where several space agencies throughout the world are poised to launch their own spaceborne SAR sensor into orbit ([2], [3], [4], [5] and [6]), and airborne SAR's for remote sensing abound. A new generation of SAR sensors is coming into existence, designed to be operationally stable, equipped with capabilities such as multi-frequency, multipolarization data-gathering, interferometric modes, wide-swath coverage, etc. The list of applications for SAR as a remote sensing tool is also rapidly expanding, in fields as diverse as oceanography, forestry, sea-ice monitoring, planetary exploration, geology, agriculture and terrain mapping.

To date, none of the SAR sensors which have produced commercially available data have been demonstrated to be successfully, reliably and repeatably calibrated. Numerous calibration experiments have been attempted, but these tend to yield results which are only applicable within a single SAR image frame. In the future, calibrated SAR imagery should be the norm, and easily related to the radar backscatter of the scene, e.g., via a look-up table relating image pixel number to σ, which can then be provided to the applications scientist. The new generation of multi-channel and interferometric SAR's mean that we must also concern ourselves with calibration of relative phase between (complex) SAR image pixels. The calibrated SAR image product should be repeatable, from day to day and from frame to frame; stable, within an image frame and between channels, and its accuracy known and understood.

Much has been achieved in the microwave remote sensing field without calibrated SAR images. The success of the SEASAT SAR [7] is a case in point. However, to advance to the quantitative use of SAR data, as opposed to the qualitative, will require calibrated images to become available in the near future. We have reached a stage where the applications scientist who wants to compare data from different sensors, extract geophysical parameters from backscatter measurements using models, carry out multi-temporal studies over large areas, build up a database of backscatter measurements for different types of terrain/incidence angle, etc. will not be able to proceed without calibrated SAR data products. Also, the full benefit of the new multichannel SAR's will not be felt unless the different channels can be properly compared with one another [8]. Another concern is that the SAR data-gathering and processing operations should be transparent to the user, so that the applications scientist does not have to spend his time checking whether the data he has received are of adequate quality for his purposes.

The Committee for Earth Observing Sensors (CEOS) has established a Working Group on SAR Calibration to address some of these concerns. The Group held its first full meeting in January, 1989 at the Jet Propulsion Laboratory in Pasadena, California, to tackle the problems of assessing SAR performance, establishing common international standards and definitions, co-ordinate multinational calibration campaigns, and educate the science community on the need for calibration and methodology.

In this paper, an overview of the need for SAR Calibration, from an applications standpoint, and the problems involved, in terms of practical calibration, is given.

II. SCIENCE REQUIREMENTS

A SAR system is simply a high-resolution active microwave sensor, capable of measuring the (complex) radar reflectivity of a surface. In its ideal realization, a SAR image should just represent:

$$\sigma (x, y, t, \theta, \phi, \lambda, E_T, E_R) \qquad (1)$$

which is the radar cross-section (amplitude and phase) of the scene as a function of spatial position (x,y), time (t), viewing geometry (θ,ϕ), radar wavelength (λ) and the polarizations of the transmitted and received electomagnetic waves (E_T, E_R). This measurement should be accurate and precise, repeatable, under the same conditions, by that or another sensor. To the majority of users of SAR data, the data collection and processing to produce RCS images should be irrelevant, except where they affect the quality of the data. For example, a geologist should not need to know whether or not an autofocus and/or clutterlock algorithm had been applied to his/her data during processing, unless these procedures caused deteriorations in the radiometric and geometric quality of his/her images. On the other hand, most users would probably want information on the level of system noise in their data, since this could directly affect their results. Unfortunately, quality control at this level has not been available to most users of SAR data in the past.

The interaction of the electromagnetic wave transmitted by a radar and the earth's surface, as a function of frequency and polarization, is fairly complex. Radar measurements of surface scatter are primarily sensitive to surface roughness and dielectric constant. The composition of the medium, in terms of particle size and orientation, is a dominant factor in volume scattering. In general, longer wavelengths will tend to penetrate diffuse surface boundaries such as leaf canopies and dry alluvium. The polarimetric signature of the backscatter is influenced by the orientation of the dominant scatters, and multiple bounces between scatters.

To study these interactions of the earth's surface (or near - surface) with microwave radiation quantitatively, using imaging radar, adequate calibration is usually necessary. The microwave remote sensing scientist's goal should be to adequately model the relationship between some geophysical parameter, X, e.g. soil moisture or biomass density, and the measured radar cross-section, σ or σ_0. The model need not be a direct relationship between X and σ, it may involve a correspondence between variations of the two with the variables given in (1), e.g. with space and time. The predictive nature of the model (forward problem) can usually be checked if the SAR data is calibrated and "ground truth" about the variations of the geophysical parameter are also available. Once a model has been demonstrated to adequately and reliably describe the relationship (or simply correlation) between X and σ or dX and dσ, it can then be used to routinely classify images in terms of X or dX. Further, if the relationship between X and σ can be shown to be unique, then the model may be inverted, so that SAR images can be turned directly into maps of geophysical parameters.

There are very few instances published in the literature of unambiguous relationships which have been developed between geophysical parameters and σ or σ_0. This situation must, in part, be attributed to the lack of routinely calibrated SAR data. In certain circumstances, the use of context, or the assumption of relative calibration within an image, have enabled the unambiguous estimation of X directly from SAR imagery. Examples include SAR images of the ocean, which have been used to map surface wind speeds and gravity wave-length and direction [9], and SAR images of ice floes, obtained on different days, which have been used to classify ice type (and therefore thickness) and to plot ice motion tracks [10], [11]. Under certain viewing conditions and assumptions, stereo or interferometric SAR data have been used to directly estimate surface height profiles [12], [13].

With calibrated SAR data the list of geophysical parameters which can be monitored should extend considerably. In Table 1 a list of such parameters is given, together with the measurement accuracy requirements for each, taken from [9] and [14]. These requirements impose their own in terms of absolute, long-term and short-term relative calibration of the SAR image products, which are also given in the Table. These requirements are quite stringent; to satisfy all these potential uses would require absolute calibration of $\pm$1dB, long-term relative calibration to +0.5dB and short-term relative calibration to better than 0.5dB. Calibration requirements for polarimetric SAR images are discussed in [15].

TABLE 1 SCIENCE REQUIREMENTS

GEOPHYSICAL PARAMETER	MEASUREMENT ACCURACY	Calibration Requirement		
		ABS	REL (SHORT)	REL (LONG)
ICE CLASSIFICATION[1] 3 CLASSES, WINTER	PROBABILITY CORRECT CLASSIFICATION > 80%	2.0	0.5	2.0
ICE MOTION[1]	PROBABILITY CORRECT MATCH > 95%	—	0.5	—
SNOWPACK WATER EQUIVALENCE	< 20 cm[1] ASSUMES DEPTH > 20 CM	1.0	1.0	1.0
WIND SPEED OVER OCEAN	< 20%	1.0	1.0	1.0
WAVE HEIGHT	SYSTEM TRANSFER FUNCTION	1.0	0.5	0.5
WAVE ENERGY	< 5%			
SOIL MOISTURE (5 SEPARATE LEVELS)	20% VARIATION IN EACH LEVEL ($\pm$0.3 gm/cm^3)	1 dB	<0.5 dB	1 dB
SURFACE ROUGHNESS (THROUGH COMPARISONS BETWEEN FREQUENCY BANDS)	—	1 dB	—	—
VEGETATION MAPPING/ MONITORING	BIOMASS DENSITY TO WITHIN 25%, LEAF AREA INDEX (LAI) $\pm$0.5 BETWEEN 0 AND 2	1 dB	<0.5 dB	0.5 dB
AGE OF LAVA FLOWS	ACCURATE CLASSIFICATION OF 3 AGE CLASSES	3 dB	1 dB	1 dB

[1] ULABY, EOS SAR PANEL REPORT
[2] R. KWOK AND J. CURLANDER, JPL

III. PROBLEMS INVOLVED IN PRACTICAL CALIBRATION

The problems involved in calibrating a SAR Sensor can be revealed by examining the radar equation appropriate to the representation of backscattered power, σ. In its most general form [16], this is:

$$P_I = \frac{P_t G_A^t (\theta_i) L_A G_A^r (\theta_i) \lambda^2 G_r \sigma G_p}{(4\pi)^3 R^4} + \bar{N} \qquad (2)$$

To radiometrically calibrate a SAR image (i.e. to obtain an accurate and precise estimate of σ from P_I), each of the terms in (2) must somehow be measured. The transmitted power, P_t, can often be measured using a power meter, in

which case the accuracy and precision of that instrument becomes a factor in the overall calibration accuracy and precision. The two antenna gain terms, $G_A^t(\theta_L)$ and $G_A^r(\theta_L)$, corresponding to the transmit and receive antennas, are functions of the look angle, θ_L, and are major sources of calibration uncertainty. The antenna patterns must be known very precisely. This can be achieved to a degree by pre-flight measurements but the patterns may change once mounted on the platform and subjected to the harsh environment of space or a high-flying aircraft. The patterns must, therefore, be re-measured in flight. Knowledge of the patterns themselves is not sufficient, the angle, θ_L, at which the real beams are pointing must also be known. The radar wavelength, λ, can be determined by preflight system tests. The receiver gain settings can also be tested before launch but should be verified in flight using a known signal, such as a sinewave or a chirp, injected into the receiver front-end. Propagation through the atmosphere may cause attenuation, L_A. Models of radar propagation usually predict small attenuation except at shorter wavelengths (X-Band) in the presence of rain. The range delay between radar and target, R, can be calibrated if the electronic delays through the radar system are known, and the system timing is calibrated pre-flight.

The processor gain, G_p, is another major source of error. To begin with, the terms in the processor gain which are invariant must be calibrated by simulation to determine the relationship between mean power in versus mean power out for the processor. Any range dependence in the processor gain should also be calibrated in this way [16]. The processor may perform many functions, besides the basic SAR correlation. An error in any of the algorithms involved, e.g. radiometric correction, doppler tracking, range cell migration correction, PRF ambiguity resolution, look extraction, slant-to ground range conversion, could lead to an error in the radiometric calibration of the final image product.

The system and background noise, N, in eq. (2) directly affects the estimate of σ from P_I, as both a bias and an uncertainty on each measurement. The bias could be removed by measuring the level of image noise in a processed sample of receive-only-noise (RON) data, i.e. with the transmitter turned off. The remaining uncertainty error can be reduced by non-coherent averaging, just as the variation due to speckle in radar measurements can be reduced.

The quantity we wish to estimate from eq. (2), σ, is often replaced by normalized radar cross-section, σ_0, via the relationship,

$$\sigma = \sigma_0 \rho_a \rho_g \qquad (3)$$

where ρ_a and ρ_g are the azimuth and ground-range resolution cell sizes. This seems straightforward, except that the actual image ρ_a and ρ_g may vary from that given in the specifications, due to processing error (impulse response broadening) and terrain fluctuations.

We have so far described a series of pre-flight and internal system measurements designed to enable us to estimate σ or σ_0 from P_I. Unfortunately, as stated in [17], and as can be gathered from the experiences of the international SAR community, this internal calibration procedure is very rarely sufficient.

In addition, external targets must be deployed and the resulting images calibrated, to 'close the loop' on the internal calibration. This external calibration procedure is also necessary for absolute determination of σ, verification of antenna patterns once in-flight, checking image quality, geometric calibration and calibration between channels of multi-frequency and multi-polarization SAR's.

External calibration, which is usually done by deploying target with known radar properties on the ground, brings its own problems. The first area of concern lies in uncertainties in the radar properties of the ground-based devices themselves, which should be measured or at least built to some standard specification. A second problem is the radar background contribution of the imaged scene, which now must be seen as a nuisance in measuring radar cross-section of the calibration targets, hence the term 'clutter'. A third problem with ground calibration is the relationship between single-point calibration measurements, and measurements averaged over large, homogeneous areas, which may not be straightforward.

So far, we have addressed the problem of radiometric calibration only. Since SAR is a coherent imaging sensor, we can make phase measurements which are related to the target - dependent phase in the following way:

$$\phi_I = \phi_t + \frac{4\pi R}{\lambda} + \phi_{target} + \phi_r + \phi_p + \phi_n \qquad (4)$$

where ϕ_I is the image phase measurement, ϕ_t and ϕ_r are the phase delays through the transmit and receive chains, $4\pi R/\lambda$ is the phase delay due to round-trip propagation of the radar pulse, ϕ_{target} the target-dependent phase, ϕ_p is any phase errors introduced during processing and ϕ_n is the total phase noise (random fluctuations) introduced by system noise. In principle, each of the terms in (4) can be measured by internal calibration. In practice, accurate phase calibration (which requires path length delay measurements to within a fraction of a wavelength) has only been possible using known ground targets.

IV. THE CEOS SAR CALIBRATION WORKING GROUP

At a meeting of the CEOS Calibration Working Group (including Optical/IR/Microwave Active and Passive sensors at ESTEC in July, 1988, it was decided there was a need for a separate working group to tackle the problem of SAR calibration. The new group had its first meeting at JPL in January, 1989, which was attended by over 40 engineers and scientists, representing the national space agencies of the USA, Canada, Europe and Japan.

The SAR Calibration Working Group's charter is as follows:

1. Establish a recommended set of SAR performance assessment techniques (i.e., quality control).
2. Establish a recommended set of standards for calibration techniques (internal and external), formats and definitions.
3. Coordinate multi-national calibration campaigns to develop expertise for intra-sensor and cross-sensor calibration.
4. Support SIR-C/X-SAR/ERS-1 Calibration activity as focal points for developing common calibration procedures.
5. Educate the science community (through support of papers, seminars, workshops) on capabilities/limitations of SAR sensors.

At the January meeting, representatives of each space agency outlined their own progress in SAR Calibration and planned activities and experiments for the future. The meeting also heard the results and status of a number of airborne calibration campaigns. Technical highlights included a presentation by L. Grey of CCRS on the extraction of radiometric measurements from point targets, using an integrated not a peak value, an overview of polarimetric calibration by R. Barnes of MIT/Lincoln Labs and a novel approach for polarimetric calibration using clutter statistics put forward by J. van Zyl of JPL.

The group also heard from the University of Michigan on multi-frequency, polarimetric scatterometer calibration and the U.S. National Institute of Standards and Technology on their microwave antenna calibration facility.

Besides the tutorial sessions a number of important areas of concern were identified for further, concentrated effort by the group, to be tackled by volunteer teams:

1. Image Quality: Norms and Techniques
2. Calibration Definitions
3. Radar Equation
4. Calibration Device Deployment
5. Error Models
6. Multi-sensor Campaigns
7. Processor Calibration
8. Polarimetric Techniques
9. Science Requirements
10. Internal System Calibration

These subject areas cover the whole range of problems involved in SAR calibration which have been described in this paper, from defining the scientific requirements to the implementation of a routine calibration procedure. It is the responsibility of each subject team (and it's leader) to establish a set of goals and draft a white paper on their chosen subject areas, to present at the next working group meeting, to be held in Italy, in September. The aim is to publish these white papers in the open literature, in journals and at conferences, and eventually a text on SAR Calibration.

ACKNOWLEDGEMENTS

The work described in this paper was carried out by the Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, under a contract with the National Aeronautics and Space Administration.

REFERENCES

1. Wiley, Carl A. Pulsed Doppler Radar Methods and Apparatus, U.S. Patent 3,196,436, Filed August 13, 1954, patented July 20, 1965.

2. Evans, D. and Elachi, C., Overview of the Shuttle Imaging Radar (SIR-C), in Proc. IGARSS '88. IEEE 88 CH2497-6, Vol. II, pp. 1015—1017.

3. Velten, E. H., X-SAR, A New Spaceborne SAR, Proc. IGARSS '88, IEEE 88 CH 2497-6 Vol II, pp. 1018—1020.

4. Luscombe, A.P., Taking a Broader View: Radarsat adds ScanSAR to it operations, Proc. IGARSS '88, IEEE '88 CH 2497-6, Vol. II, pp. 1027—32.

5. Louet, J., ERS-1 Operation Capabilities, Proc. IGARSS '88, IEEE 88 CH 2497-6, Vol. II, pp. 855-858.

6. Kashihara, H., et al, A Case Study of Space-borne SAR System Design for the Earth Resources Satellite (J-ERS-1), Proc. IGARSS '86, ESA SP-215, Vol. I, pp. 815—820.

7. Ford, J. P., et al, Seasat views North America, The Caribbean and Western Europe with Imaging Radar, JPL Publication No. 80-67, November 1980.

8. Freeman, A., Curlander, T. C. Dubois, P.D. and Klein, J. D., SIR-C Calibration Workshop Report, JPL Center for Radar Studies publication No. 88-003, November 1988.

9. EOS SAR Instrument Panel Report, NASA publication, 1988.

10. Lyden, J.D., Burns, B.A. and Maffett, A.L., Characterization of Sea-Ice Types Using Synthetic Aperture Radar, IEEE Trans. on Geoscience and Remote Sensing, Vol. GE-22, No. 5, pp. 431-439, Sept. 1984.

11. Curlander, J.C., Holt, B. and Hussey, K.J., Determination of Sea-Ice Motion using Digital SAR Imagery, IEEE Trans. on Oceanic Engineering, Vol. OE-10, No. 4, pp.358-367, October 1985.

12. Leberl, F., et al, SIR-B Stereo-Radargrammetry of Australia, International Journal of Remote Sensing, Vol. 9, No. 5, pp. 997—1012, May 1988.

13. Goldstein, R. M. and Gabriel, A. K., Crossed Orbit Interferometry: Theory and experimental results from SIR-B, International Journal of Remote Sensing, Vol. 9, No. 5, pp. 857—872, May 1988.

14. Kwok, R., private communication.

15. Dubois, P. D., et al, SIR-C Science Calibration Requirements, to appear in Proc. IGARSS '89, Vancouver BC, Canada, July 1989.

16. Freeman, A. and Curlander, J. C., Radiometric Correction and Calibration of SAR Images, submitted to Photogrammetric Engineering and Remote Sensing, March 1989.

17. Currie, N. C. (ed.) Techniques of Radar Reflectivity Measurement, publ. Artech House, 1984.

PRACTICAL SAR RADIOMETRIC CALIBRATION MEASUREMENTS AND EXPERIMENTS

Eric S. Kasischke

Radar Science Laboratory
Advanced Concepts Division
Environmental Research Institute of Michigan
Ann Arbor, Michigan USA

ABSTRACT

The philosophy behind and approaches to radiometric calibration of SAR imagery are discussed in this paper. The equations necessary for absolute and relative calibration of SAR imagery are presented and the critical terms identified. The effects of these critical terms on calibration are discussed, and approaches for the field experiments necessary for their measurement are presented.

KEYWORDS: SAR, radiometric calibration, calibration errors, calibration experiments

1.0 INTRODUCTION

The subject of radiometric calibration of digitally-processed data collected by synthetic aperture radars (SARs) has typically been limited to a select group of engineers and scientists directly involved with a specific sensor system. Because most SARs to date have been experimental in nature and not operational, or primarily used for mapping purposes where the radiometric precision or accuracy of the data were not an issue, potential users of SAR data have either accepted the often steep costs associated with calibration of their data or limitations in the data resulting from an incomplete calibration effort.

The experience in calibration of SAR systems has finally reached the point where engineers have developed procedures to minimize the errors of those system parameters which can influence radiometric calibration. Given the fact that during the 1990's a host of spaceborne SAR systems will be launched, now is an opportune time for the application scientist to develop an understanding of the issues concerning radiometric calibration of SAR imagery, especially those steps necessary to insure the SAR data being collected by future systems are useful for the intended applications.

The purpose of this paper is threefold: (1) to review the equations (and the required parameters for the equations) necessary to radiometrically calibrate SAR imagery; (2) to discuss how each of the calibration parameters can be obtained by the end user of the data; and (3) to discuss alternatives for the application scientist in conducting experiments and collecting measurements to obtain the calibration parameters.

2.0 RADIOMETRIC CALIBRATION OF SAR IMAGERY

In this paper we will address the problem of radiometric calibration of a single-frequency, single-polarization SAR system. We will not address cross-frequency or cross-polarization calibration issues. Using these constraints, three levels of radiometric calibration can be defined: **relative within-scene calibration**, which concernes with comparison of radar signatures from two different areas within the same SAR scene; **relative between-scene calibration**, which concerns comparison of radar signatures from the same or different areas in two distinct SAR scenes; and **absolute calibration**, which concerns how well the SAR-predicted radar cross section (σ) or radar scattering coefficient (σ°) values compare to the actual values contained in the scene.

In order to radiometrically calibrate SAR data, the power intensity recorded on the SAR image (P_i) is related to σ or σ° via the radar equation. The exact form of the radar equation depends on both the radar system utilized as well as the manner in which the data are digitally processed into imagery. In this paper, we utilize the equation developed by Kasischke and Fowler (1989) for an airborne SAR system. The relationship between σ° and P_i is expressed as

$$\sigma^\circ = \frac{(P_i - HP_n)\ 2(4\pi)^3\ R^3\ u\ \tau_p\ \sin\theta}{H\ (1+M)\ \lambda^3\ P_{av}\ g^2 G^2(\theta,\phi)\ r_s} \quad (1)$$

where H represents the overall system gain (from the receiver through digital processor), P_n is the additive noise power for the system, R is the slant range to the target, u is the platform velocity, τ_p is the transmitted pulsewidth, θ is the incidence angle to the target, M is the multiplicative noise ratio of the SAR, λ is the wavelength of the SAR, P_{av} is the average transmitted power, g is the peak antenna gain, $G(\theta,\phi)$ is the antenna gain pattern as a function of θ and ϕ (the azimuth angle to the target), and r_s is the slant range resolution of the SAR.

Other formulations of the radar equation for a SAR exist, and the major difference between these and Eq. (1) usually lies in the R^3 term, which may be an R^2 or R^4 term depending on how the data are processed.

In considering methods for radiometric calibration, the parameters defined in Eq. (1) can be categorized based on whether the parameters are:

1. Measured internally or externally;
2. Scene dependent;
3. Range variant; or
4. Azimuth variant.

2.1 INTERNAL VERSUS EXTERNAL PARAMETERS

Internal radar system parameters are those variables which are measured by the radar system itself or are derived directly from the image, and include parameters which are: (1) fixed (e.g., g, λ, τ_p, π); (2) derived from the image itself (i.e, P_i, P_n, r_s); (3) recorded or assumed radar system parameters (P_{av}, u, H, M); or (4) derived from a combination of system and image measurements (R, θ, ϕ).

External radar parameters are those variables measured with the aid of calibration targets deployed within the SAR scene, and are the major focus for measurement during calibration experiments. These parameters include measurement of the regions of system linearity and the antenna elevation pattern [$G(\theta,\phi)$].

2.2 SCENE-DEPENDENT CALIBRATION PARAMETERS

This category contains only one parameter, the measured image intensity (P_i). This parameter is the only one dependent on the scene itself, specifically the radar cross section of the targets within the resolution cell of the SAR.

2.3 RANGE-VARIANT CALIBRATION PARAMETERS

There are a group of parameters [R, θ, $G(\theta,\phi)$] which are range dependent, and typically are calculated separately for each pixel during the calibration process. These parameters are obviously dependent on the imaging geometry of the scene being collected by the SAR.

2.3 AZIMUTH-VARIANT CALIBRATION PARAMETERS

In a SAR image, the azimuthal dimension can be equated to the time over which certain calibration parameters can change. In an ideal SAR system, every calibration parameter would be temporally constant, i.e., they would never change over the operational lifetime of the SAR. Thus, these parameters would only have to be measured once. However, this situation us highly unlikely to occur. Some parameters are changed deliberately in order to optimize the operation of the SAR. These parameters include the overall system gain (H) and the average transmitted power P_{av}. Other parameters change because of changes in the operational environment. For instance, for future spaceborne SAR systems using large array antennas, the heating and cooling of the antenna panels by solar illumination will lead to a bending and warping of the antenna, thus changing $G(\theta,\phi)$.

The temporal stability of the SAR system parameters represents one of the major hurdles in radiometric calibration of the SAR. An awareness of the frequency over which these parameters change is a strong determinant in the design of calibration experiments.

In the beginning of this section, we discussed two types of relative calibration: within-scene relative calibration and between-scene relative calibration. The difference between these two types of relative calibration is basically a matter of how frequently the azimuth-variant calibration parameters change. In theory, a spaceborne SAR system can collect a continuous strip of imagery, which completely circles the earth. If the system is stable, then relative calibration between two parts in this continuous image would always be within-scene, even if it considered two different images of the same area collected at different times. However, since it is likely that some radar system parameters change between the different scenes, then in the context of our definitions, we would be faced with a between-scene calibration.

2.4 SUMMARY

In Table 1, we present a summary of the different categories describing the calibration parameters. The combination of the different categories which a parameter falls into, the difficulty in measurement of the parameter, and the influence of that parameter on the overall error bound for radiometric calibration all contribute to the strategy for measurement of that parameter.

3.0 FUNDAMENTAL CALIBRATION APPROACH

To actually calibrate a SAR system, we can define two variables, K and P', where

$$K = \frac{2(4\pi)^3 \, u \, \tau_p}{H \, (1 + M) \, \lambda^3 \, P_{av} \, g^2 \, r_s} \tag{2}$$

and

$$P' = \frac{(P_i - HP_n) \; R^3 \; \sin\theta}{G^2(\theta,\phi)} \tag{3}$$

Note that K consists of those terms which are primarily azimuth variant, while P' consists of those terms which are range variant or scene dependent. Using Eqs. (2) and (3), Eq. (1) can be redefined as

$$\sigma^\circ = K \, P' \tag{4a}$$

or as

$$\sigma^\circ \; (dB) = K \; (dB) + P'(dB) \tag{4b}$$

if we transform the radar equation parameters into dB.

In order to understand how calibration is achieved, consider the following example: Assume an area with a large number different targets or target areas whose radar scattering coefficients range between -10 and 40 dB m^2/m^2 is imaged by a SAR. Furthermore, assume the SAR system being used has only 30 dB worth of dynamic range, and that an attenuator in the receiver of the SAR defines the regions where the SAR is linear to $0 \leq \sigma^\circ \leq 30$ dB.

Figure 1 presents a plot of σ° plotted against P' for this example data set. This plot illustrates several important points about radiometric calibration of SAR data. First, only those regions of the SAR image where P_i is within the linear region of the system response curve can be calibrated. Thus, an important measurement for any calibration experiment is to define this system linearity window.

A linear regression model can be utilized to determine several important calibration parameters using those points in the example which are in the linear response window. The basic regression model is

$$Y_i = \alpha + \beta X_i + e_i \tag{5}$$

where $Y_i = \sigma^\circ$, $\alpha = K$, $X_i = P'$, $\beta = 1$ (since we have a linear detector), and e_i is the error term. The parameters α and e_i can be estimated using least squares regression techniques (see Kasischke and Fowler, 1989).

By using the K value determined from the calibrated reference targets in the examplee, the σ° of any other target in the scene can be estimated if the

radar system parameters in Eq. (3) are known or can be estimated and if $40 \leq P' \leq 70$. The additive noise level is usually determined via system engineering measurements and R and θ can easily be calculated from the radar imaging geometry. The most difficult parameter to measure in Eq. (3) is $G(\theta,\phi)$. If this parameter is known, calculation of σ° is a relatively straightforward process.

The scenario outlined above represents an integral part of the basic approach utilized to radiometrically calibrate a SAR system in an absolute sense. In summary, the following steps are used:

1. A set of targets whose σ or σ° values are known or are measured are imaged by a SAR system. The range in σ or σ° of these targets should be large enough to span the linear response envelope of the SAR.
2. A modified linear regression model (assuming $\beta = 1$; Kasischke and Fowler, 1989) can be used to estimate K for the SAR system using the SAR-derived P' values for the targets.
3. The standard error of the regression line (also estimated using the modified regression model) represents the relative or absolute calibration level of the SAR (see Kasischke, et al.,1988).

Next, consider the case where the objective is to calculate a σ° value from a different scene. The equation used in this case is

$$\sigma^\circ = K_1 P_1' / K_2 \qquad (6)$$

where the 1 subscript represents the area within the new scene to be calibrated and the 2 subscript represents the scene containing the calibrated targets. There are two basic approaches to estimate K_1: (1) targets whose radar cross section are known are deployed in the image, and K_1 is measured directly in the same manner as K_2; or (2) the radar system parameters in Eq. (2) are accurately measured, and the ratio of K_1/K_2 is calculated based on the relative changes in these parameters between the two passes.

For relative calibration, we have

$$\sigma_r^\circ = K_1 P_1' / K_2 P_2' \qquad (7)$$

where the 1 and 2 subscripts represent areas of the scene(s) to be calibrated relative to one another. For relative within scene calibration, where the SAR system is stable, $K_1 / K_2 = 1$. Otherwise, K_1 and K_2 are estimated in the same manner as for absolute calibration, i.e., either by measuring the parameters in Eq. (2) directly, or by using a reference target within the scene (not necessarily calibrated, but whose σ° is the same in both scenes).

4.0 EFFECTS OF CALIBRATION PARAMETER ERRORS

In the previous section, we discussed the basic approaches for radiometric calibration of SAR imagery. A key consideration in the design of calibration experiments is the effect a specific parameter has on the precision and/or accuracy of the SAR. The combination of these effects and the difficulty in measuring a specific parameter determine the strategy employed for measuring the parameter.

In summary, the following radar parameters must be measured or estimated in order to radiometrically calibrate an image: P_i, HP_n, R, θ, $G(\theta,\phi)$, and K. The image intensity, range and incidence angle to the target area can all be measured or estimated with very little error. Thus, the uncertainties in the calibration process stem mainly from the measurement errors associated with the other parameters.

In Figure 2, a plot of the change in the relative error bound for radiometric calibration of a SAR image based on the error in estimation of two key calibration parameters [K and $G(\theta,\phi)$] is presented. These bounds were calculated using the error model of Kasischke and Fowler (1989). Two cases for the antenna gain pattern are plotted, since for absolute calibration it enters the calibration equation only once, while for relative calibration it enters the equation twice. Clearly, the antenna pattern has the most significant effect on the relative error bound.

The effect of estimation of the additive system noise level (HP_n) on the relative error bound is somewhat complex. The error associated with the additive noise level is not so much dependent upon the error in measuring HP_n, but on the signal-to-noise-ratio (SNR) of the image (i.e., P_i / HP_n). Based on the model of Kasischke and Fowler, as $HP_n \rightarrow P_i$, the relative error bound increases significantly, as illustrated in Figure 3. From this plot, it can be seen that in order to minimize the calibration errors, the SNR should be greater than 6 dB.

5.0 MEASUREMENT OF CALIBRATION PARAMETERS

It is hoped that the discussions in the previous sections have outlined the basic approaches and problems associated with radiometric calibration of SAR data. Based upon the discussions in the previous sections, the following list represents the key measurements required for radiometric calibration of a SAR image:

1. The additive noise level for the SAR image;
2. The system linearity window;
3. The antenna gain pattern;
4. The relative values of all parameters in Eq (2), or estimates of K_1 and K_2; and
5. For absolute calibration, an external measurement of K_2 using calibrated reference targets.

With respect to these five areas, items 1 through 3 require an extensive set of calibration passes over target arrays which contain highly specialized targets and background clutter scenes. It is impractical for each user to collect these passes and analyze the resultant data independently. However, the parameters defined by these flights are typically provided by the operators of the SAR system. Thus, any person using a SAR for experimental purposes should request these measurements for each SAR scene they utilize.

With respect to measuring the radar parameters used in the estimation of K, two alternatives are available:(1) depend on the system's ability to measure each parameter used to estimate K; (2) or deploy calibrated targets so that K can directly be estimated from the scene of interest. If the first alternative is selected, then the user should demand data from the SAR system operators which demonstrates that the system is indeed calibrated to the extent that K can be accurately estimated using internal measurements alone.

If the second alternative is selected, then the following procedure is recommended:

1. Identify a target size which will be large enough to be in the middle of the system response envelope of the SAR;
2. Deploy the targets (approximately 6) near the area of interest in a low enough clutter background so that a target-to-clutter ratio of 20 dB is achieved. This is necessary to minimize the errors associated with measurement of these targets by the SAR.

Although most experimental airborne SAR systems in operation today are striving to develop techniques to internally measure the parameters necessary to estimate K, in the author's opinion none have yet to achieve this goal. Thus, it is in the best interest of those wishing to utilize these systems for experimental purposes to select the second option.

Unfortunately, there are significant costs associated with acquiring and deploying calibrated ground targets. Today, most operators of airborne SAR systems usually have a number of targets at their disposal, and it is well worth the additional costs associated with requesting their deployment.

6.0 SUMMARY

In this paper, the basic approach used for radiometric calibration of SAR imagery is outlined and the issues surrounding measurement of the parameters necessary for radiometric calibration have been discussed. Also, the critical parameters needed by the user for calibration of their data are identified.

Hopefully, in the near future the operators of SAR systems will progress to the point where they routinely provide calibrated data to the user. In addition, through the convening of international Working Groups, such as the Committee on Earth Observing Satellites' Working Group on SAR Calibration, it is hoped that an approach to radiometric calibration can be agreed upon so that cross-calibration of different SAR systems can be achieved.

TABLE 1
SUMMARY OF CALIBRATION PARAMETER CATEGORIES

Parameter	Internal or External	Scene Dependent	Range Variant	Azimuth Variant
P_i	Int	yes	yes	yes
HP_n	Int	no	yes	yes
R	Int	no	yes	no
u	Int	no	no	yes
τ_p	Int	no	no	no
θ	Int	no	yes	no
H	Int	no	no	yes
M	Int	no	yes	yes
λ	Int	no	no	no
P_{av}	Int	no	no	yes
g	Ext	no	yes	yes
$G(\theta,\phi)$	Ext	no	yes	yes

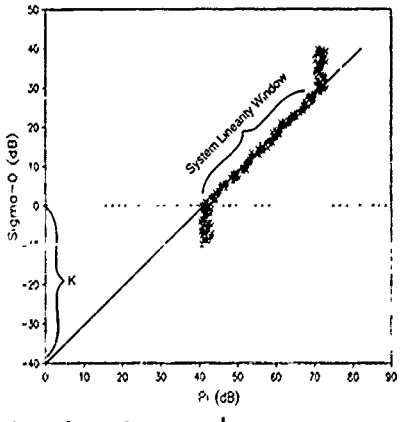


Figure 1. Plot of σ° vs P' for Theoretical Radar Targets

7.0 ACKNOWLEDGEMENTS

Many of the thoughts and ideas presented in this paper evolved as the result of discussions with colleagues at ERIM. I would like to acknowledge these individuals for their contributions to this paper. They include C. Wackerman, D. Sheen, R. Sullivan, D. Gineris, W. Dunlevy and N. Malinas.

8.0 REFERENCES

Kasischke, E.S. and G.W. Fowler, A Statistical Approach for Determining Radiometric Precisions and Accuracies in the Calibration of Synthetic Aperture Radar Imagery, IEEE Trans. Geosci. Remote Sens., in press, 1989.

Kasischke, E.S., D.R. Sheen, C.C. Wackerman and G.W. Fowler, Radiometric Calibration of Airborne SAR Data, Proc. 1988 Intern. Geosci. Remote Sens. Symp., Edinburough, Scotland, pp. 711-714, 1988.

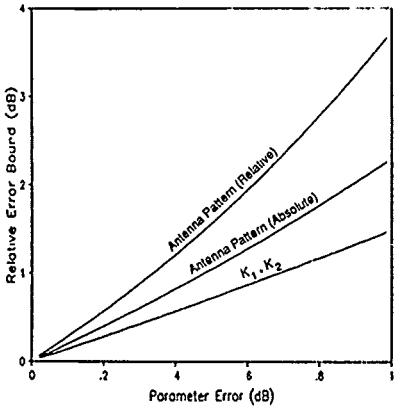


Figure 2. Relative Error Bounds for $G(\theta,\phi)$ and K

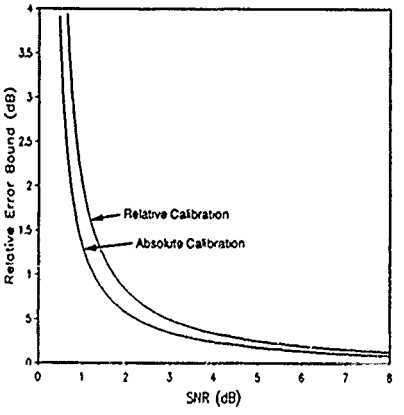


Figure 3. Relative Error Bounds for Different SNR Levels

Calibration for Airborne SAR*

R.K. Hawkins, T.I. Lukowski, A.L. Gray, C.E. Livingstone

Canada Centre for Remote Sensing

Abstract

Calibration may be considered to have many facets: radiometric, geometric (spatial), and interferometric (phase) comparisons; embracing static, temporal, and signal dependent measurement variances; and measurement tolerances dictated by radar, processor, and application or scattering space. This paper first deals generally with these issues but concentrates mainly on the radiometry of airborne SAR and specifically on relative and absolute radiometric calibration of the CCRS airborne SARs.

Airborne SAR has unique calibration difficulties due to its large swath/height ratio when compared with the satellite-borne case. This manifests itself in strongly varying terms in the radar equation with important implications for image interpretation and analysis. *Calibration* for airborne SAR may therefore lead to dissimilar data products from their satellite counterpart.

Recommendations for data acquisition and calibration using this facility will also be presented, along with CCRS plans for continuing calibration strategy, emphasizing the implications of calibration on data utility.

Keywords. SAR, calibration, radar, airborne, remote sensing

1 Introduction

Civilian use of airborne SAR has evolved from a limited research activity to an operational surveillance reality in Canada [1]. Internationally, there are now at least 8 airborne SARs for civilian remote sensing purposes in use or under development [2], [1], [3], [4], [5], [6], [7], [8]. This rapid progression has occurred through the recursive influences of sensor/processor performance potential and user requirements and has resulted in production of high quality imagery in multi-polarizations and multi-frequencies in near real time.

In parallel with this technological evolution, fostered by the combination of digital processing and control systems with more powerful and reliable hardware, has been a growth in user sophistication and application requirements. Digital image processing techniques, temporal and other comparative studies, and more recently, innovative data requirements, such as the polarimetric scattering matrix [9], [10], have pushed calibration into vogue as an important issue. Despite this enthusiasm, there will remain several useful applications of SAR for which calibration is unimportant (*e.g.* some land, ocean, and ice mapping requirements) and others for which only relative calibration is required (*e.g.* change detection).

Calibration may be considered to have many facets: radiometric, geometric (spatial), and interferometric (phase) comparisons; embracing static, temporal, and signal dependent measurement variances; and measurement tolerances dictated by radar, processor, and application or scattering space. Figure 1 shows this *calibration cube* which shows the interconnections of these aspects [11]. In this general sense, we can see that calibration cannot be simply considered in the narrowness of one scene or application but embraces a much broader scope. The perceived importance of SAR calibration can be seen in the recent formation of a CEOS working group dealing with the many issues which relate to it. Application requirements analysis by this group indicates that an absolute calibration accuracy between 0.5 and 1.0 dB is desireable even though it is well known [12] that due to fading measured data fluctuates much more than these figures.

Many attempts [13], [14], [15], [16], [17], [18] at airborne SAR calibration have been made with mixed success and sophistication. Current plans at CCRS include consideration of the radiometric and geometric planes of the calibration cube although new projects involving extensions of the current system to include interferometry and polarimetry will require phase calibration. Radiometric calibration seems to be the obvious first step.

*To be presented at IGARSS'89, Vancouver, July, 1989

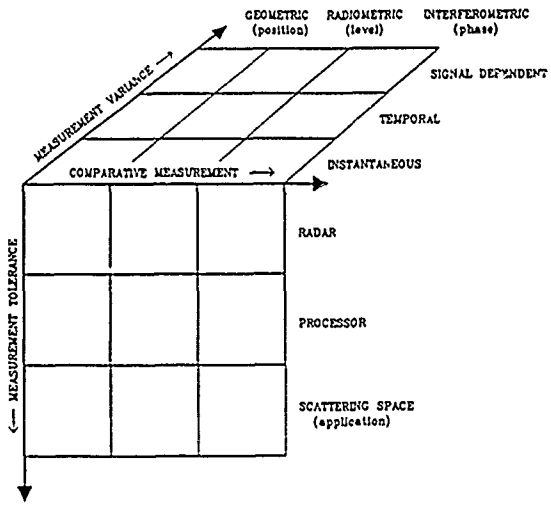


Figure 1: Calibration Cube.

This heuristic device is intended to show the interconnections and multi-dimensionality of the calibration problem.

2 Radiometric Calibration

Table 1 shows the main identified radiometric calibration issues for the radiometric plane of the cube, as it applies to the CCRS SAR.

Table 1: Radiometric Plane of Calibration Cube.

	INSTANTEOUS	TEMPORAL	SIGNAL DEPENDENT
RADAR	-STC -Ambiguities -Antenna pointing -Polarization sensitivity -Quantization -Band pass filters -System noise -Aliasing -Power -RF Interference -Gains	-System changes -Drifts in gain -Antenna pointing -Thermal stability -Motion Compensation -Antenna pattern -Aliasing -Power -RF Interference	-Dynamic range -Saturation -Linearity -Quantization
PROCESSOR	-Focus -Bandwidth -Computational efficiency -Multi-looking	-Processor change	-Dynamic range
SCATTERING SPACE	-Target pointing -Slope/area effect -Local θ_{inc}	-Signal to Clutter -Soil moisture	-Speckle

There are several available approaches to SAR calibration. One of these involves performing a systems analysis and determining all of the gains and losses in the system including radar and processor. This usually leads to some undetermined parameter which often contains the antenna gain or other uncertainties concerning the processor or radar. To circumvent these difficulties, an end-to-end approach to calibration is often adopted as has been shown by researchers at ERIM [18] for instance. End-to-end calibration may involve using distributed or point targets and both approaches have their own proponents and inherent difficulties.

The following relations have been used (*e.g.* [19]) in slightly different form for distributed and point target imaging by SAR:

$$S_r^d = \frac{P_t}{V}\left(\frac{\lambda}{4\pi}\right)^3 \frac{\rho_r \sigma^o}{2R^3 \sin\theta_{inc}} < G_{tr}(\vartheta) > e^{-\alpha R} \qquad (1)$$
$$\times g_{STC}$$
$$\times G_r$$
$$\times G_p$$
$$\times L$$
$$+\text{noise}$$

$$S_r^p = \frac{P_t}{V}\left(\frac{\lambda}{4\pi}\right)^3 \frac{\sigma_t}{2\rho_a R^3} < G_{tr}(\vartheta) > e^{-\alpha R} \qquad (2)$$
$$\times G_r$$
$$\times G_p$$
$$\times g_{STC}$$
$$\times L$$
$$+\text{noise}$$
$$+\text{clutter term}$$
$$+\text{interference term}$$

Here:

S_r^d	-received power from a distributed target in the image pixel,
S_r^p	-received power from the point target in the image pixel,
P_t	-average transmitted power,
V	-ground speed of the platform,
λ	-wavelength,
ρ_a	-azimuth resolution,
ρ_r	-slant range resolution,
σ^o	-backscattering coefficient,
σ_t	-radar cross section of the point target, assumed to be constant over ϕ_ξ,
R	-range to the pixel,
g_{STC}	-STC function [See (3).],
$< G_{tr}(\vartheta) >$	-average antenna gain over the *processed* beam width. $< G_{tr}(\vartheta) >= \frac{1}{\phi_\xi}\int_{\phi_\xi} G_{tr}(\vartheta,\varphi)d\phi$,
ϑ, φ	-antenna elevation and azimuth co-ordinates,
θ_{inc}	-incidence angle,
$G_{tr}(\vartheta,\varphi)$	-two-way antenna pattern including the radome,
α	-atmospheric attenuation coefficient,
G_r	-receiver gain function,
G_p	-processor efficiency,
L	-link budget loss factor which includes losses due to the transmission line, couplings *etc.*,
noise	-measured noise with the transmitter off [See (7).] and
ϕ_ξ	-range of azimuth antenna angle used in the *processed* bandwidth ξ.

In these relations, which explicitly involve the impulse function width, we have included terms for noise, atmospheric attenuation, system gains, clutter, and interference required for the discussion below. The STC function is given in section 2.1 with its implication for airborne SAR imagery interpretation.

Monitoring, where possible, changes in each of the terms in these equations is an essential systems requirement to understanding the calibration stability and producing anything other than piecemeal calibrated products as underlined by Attema [20].

2.1 STC Implications for Airborne SAR Interpretation

In the CCRS SAR [21], a time-dependent gain, known as STC, is applied before digitization to reduce the dynamic range of the signal and to produce a relatively 'flat' image product. This is a *model* relationship based on antenna gain, terrain reflectivity, and a zero relief earth. The form of the STC function is given in (3) and may be seen to cover the range-dependent terms in (1). Here, θ^{n} is the *nominal* boresight angle of the antenna.

$$\frac{1}{g_{STC}} = \frac{\kappa\, G_{el}(\theta - \theta^{n})\, \sigma^{o}(\theta)\, r^{-3}}{k^{2} \sin\theta_{inc}} \qquad (3)$$

The STC function for a model σ^{o} law for land is shown in Fig. 2 for the two high-resolution modes. The large and complex dynamic range (> 35 dB for *nadir* geometry) is related to the wide range of incidence angles imaged in an airborne SAR. Because the swath/height ratio is high, special consideration is required that is not present for the satellite-borne SAR[1]. These include

- dynamic range of output,
- variation of backscatter with incidence angle,
- variation of radar contrast with incidence angle.

Most researchers would therefore require not an absolutely calibrated (σ^{o}) image, but a relatively calibrated product which includes the STC function.

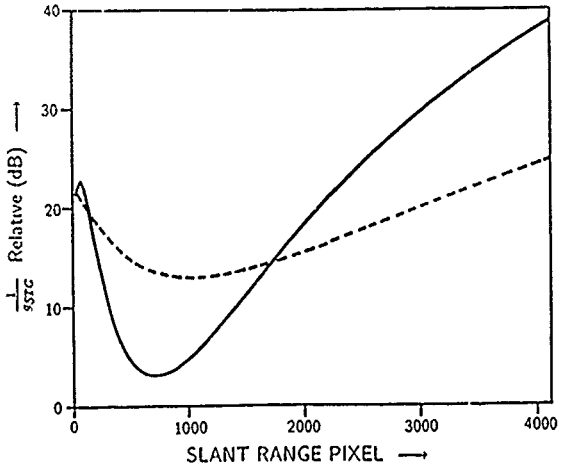


Figure 2. STC Function for CCRS SAR.

Two curves are shown. The dashed curve is for the *narrow swath* mode of the radar which covers an incidence angle range of $\sim 45° - \sim 76°$. The full curve is for the *nadir* mode which covers $0° - 7.1°$. In each case STC gain is plotted against the range pixel across the swath for a nominal flying height of 20000 ft and HH-polarization at C-band.

[1] As an example, the total swath of the ERS 1 SAR will subtend an angle of $\sim 7°$, whereas $\sim 75°$ are imaged in the *nadir* mode of the CCRS airborne SAR.

3 Calibration Strategy

Data from the CCRS SAR is processed either by an on-board, real-time, SAR processor (RTSP) or through ground based processing as explained in [21]. Because the compression gain for a point target is usually high, and the RTSP is an 8-bit processor with other restraints, work using the RTSP with point targets is not usually possible and the overall strategy is to work with the more flexible ground processor, establish calibration parameters there, and then using distributed targets, make the transfer to the real-time product where necessary.

3.1 Data Acquisition

During data acquisition, there is an automatic logging of many parameters on the MAID system [21] which are implicit parts of the radar's function. Flight parameters include: altitude, attitude, position, vertical acceleration, and horizontal velocities. Antenna parameters include: polarization, depression angle, azimuth angle. Radar parameters include: power, PRF, range delay, system gains, and clutter lock parameters.

To first order, fluctuations in the motion parameters are fully compensated by the motion compensation system. Higher order effects indicated in Table 1, such as antenna steering biases, clutterlock swings, and INS (inertial navigation system) errors, will necessarily require postflight analysis.

The recommended complete flight acquisition consists of:

(A) –recording of 'imagery' in signal and fully compressed forms,

(B) –recording of ancillary *MAID* parameters,

(C) –a noise record with the transmitter off but all other system parameters fixed,

(D) –a BITE (built in test equipment) oscillator recording used to monitor a reference signal through the system.

3.2 Absolute Calibration: An Area Approach

The use of point targets for specific scene calibration has been established in several investigations [14], [13], [22]. In many of these studies, the peak response of the point target has been used; however, Gray [23] has shown that this depends rather strongly on the processor focus and resolution whereas the integrated response from the point target is weakly dependent on these parameters. For this reason, we suggest it is more suitable to use an area approach as illustrated in Fig. 3, despite the apparent success obtained by other investigators [22].

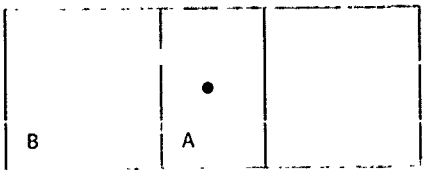


Figure 3: Area Approach to Calibration.

The point target, together with its sidelobes, are contained in area, A. Area, B, contains A and the excluded area B-A contains representative clutter and noise similar to that present in A.

In calculating the integrated image power, it is necessary to use the fact that image points are separated by ρ_r, in slant range and ρ_a, in azimuth. If we assume that energy is conserved and SAR processing or 'compression' is an energy redistribution process [24] (tempered by inefficiency associated with weighting), then it can be shown that

$$\sigma_i = kP_\Sigma \quad (1)$$

$$P_\Sigma = \left\{ \sum_{ij}^{B} P_{ij}^2 - \frac{N_B}{N_{B-1}} \sum_{ij}^{B-1} P_{ij}^2 \right\} \quad (5)$$

$$\sigma_{ij(i)}^{\circ} = \frac{k}{a} \left(P_{ij}^2 - < N_i^2 > \right) \quad (6)$$

$$< N_i^2 > = \overbrace{\frac{1}{M} \sum_{r=1}^{M} P_{ij}^2}^{transmitter\ off} \quad (7)$$

Here:

- k -calibration constant from image units to m^2.
- P_Σ -background subtracted image power associated with the point target.
- P_{ij} -pixel amplitude value for ith line and jth pixel.
- $P_{ij(i)}$ -pixel amplitude for the ith line and range pixel of target
- B -image space containing whole area
- $B-1$ -image space excluding the point target.
- N_{B-1} -number of pixels in $B-1$
- N_B -number of pixels in B.
- $\sigma_{ij(i)}^{\circ}$ -corresponding scattering coefficient of distributed target contained in the same range line having amplitude $P_{ij(i)}$.
- a -ground projected area of the pixel in m^2 $a = \frac{\text{pixel}_{\rho_r}\text{pixel}_{\rho_a}}{\sin\theta}$.
- N_j -noise profile for pixel j with transmitter off, for pixel j, and
- M -number of azimuth lines averaged for noise $M > 500$

By subtracting the background in this way, both the effect of average clutter and noise in the point target image are removed. It has the advantages of removing any system biases and uses direct measurements. It does *not* cover the interference term in (2) and a high signal-to-clutter ratio is still required [25].

3.3 Relative Calibration

In relative calibration, two or more images are to be compared using one as a reference. In the analysis below, a point target is assumed to be imaged at approximately the same range in each scene to be used as a reference for undetermined drifts in the system although a distributed target may also be used. When the images were obtained at the same flying altitude and geometry, the following relationship may be used to correct for the effects of system drifts[2], noise, power fluctuations, target cross section, STC law, and antenna pointing.

$$P_{ij}'^2 = \left(P_{ij}^2 - < N_j^2 > \right) \times \left(\frac{P_\Sigma}{\sigma_i^\circ} \right)_{Ref} / \left(\frac{P_\Sigma}{\sigma_i^\circ} \right) \times \frac{(g_{STC})_{Ref}}{g_{STC}} \times \frac{< G_{ij}(\vartheta_{Ref}' - \vartheta_j) >}{< G_{ij}(\vartheta' - \vartheta_j) >} \quad (8)$$

[2] This may also determined from cal generators as suggested by Larson [18] and in the CCRS system a BITE facility is also provided for this purpose

Here:

- P_{ij}' -corrected amplitude of the processed image for the data at line i and pixel j.
- P_{ij} -corresponding uncorrected amplitude.
- P_Σ -integrated image response given by target given by (5).
- σ_i° -radar cross section for the point target in the scene.
- ϑ' -*actual* antenna boresight setting.
- ϑ_j -incidence angle of the jth pixel, and
- Ref -indicates the function is evaluated for the reference image value.

This technique has been applied to multi-temporal data in Saskatchewan as described in [26].

4 Future Plans

A major obstacle to routine calibration in the airborne SAR is uncertainty in the antenna gain function [27]. The absolute calibration scheme indicated above, therefore implies that *only* imagery obtained in the same angular range in which the point target is located is 'calibrated'. Other lines in the image depend on an uncertain antenna gain function. It follows that this gain should be measured as closely as possible and estimates made of its dependence on pointing.

CCRS has underway two approaches to determine this pattern. One involves a 3-D mockup of the underside of the Convair-580 and antenna range measurements. The second is a rough target analysis [28]. With these measurements, it should be possible to extend measurements to the full range of the image although the tolerances may not be as small as 0.5 dB

System linearity and performance is being studied as part of the engineering assessment of the instrument. Indications to date are that these are more than adequate.

A host of second order calibration errors are also being investigated and analysis being carried out to quantify, reduce, where necessary, and compensate where possible. These relate principally to motion compensation and antenna steering systems. Extensions to the geometric and interferometric planes of the calibration cube are also under study.

References

[1] Bullock, B L. "Radar applications in remote sensing: an airborne remote sensing case history". In *Proceedings of the Twenty-First Symposium on Remote Sensing of Environment*, pages 619–627, Ann Arbor, Michigan, October 1987.

[2] Livingstone, C.E., A.L. Gray, R.K. Hawkins, and R.B. Olsen. "CCRS C/X airborne synthetic aperture radar. An R & D tool for the ERS-1 time frame". In *Proceedings of Radar'88*, Ann Arbor, Michigan, 1988 7p.

[3] Zhang, C.B. "A synthetic aperture radar with multichannel and multipolarization". In *IGARSS'88 Remote Sensing: Moving Toward the 21st Century*, pages 331–333, Edinburgh, Scotland, 1988.

[4] Held, D.N., W E. Brown, J.D. Klein, T. Sato, T. Miller, Q. Nguyen, and Y. Lou. "The NASA/JPL multifrequency, multipolarisation airborne SAR system". In *IGARSS'88 Remote Sensing: Moving Toward the 21st Century*, pages 345–348, Edinburgh, Scotland, 1988

[5] Horn, R. "E-SAR – The experimental airborne L/C-band SAR system of DFVLR". In *IGARSS'88 Remote Sensing: Moving Toward the 21st Century*, pages 1025–1026, Edinburgh, Scotland, 1988.

[6] Skou, N., S.N. Madsen, and E.L. Christensen. "The TUD C-band SAR". In *IGARSS'88 Remote Sensing. Moving Toward the 21st Century*, pages 1039–1040, Edinburgh, Scotland, 1988.

[7] Massonnet, D. "Preprocessing of the VARAN synthetic aperture airborne radar". In *IGARSS'88 Remote Sensing: Moving Toward the 21st Century*, pages 715–720, Edinburgh, Scotland, 1988.

[8] Kirby, M. "INTERA ice reconnaissance SAR" March 1989 Private communication.

[9] van Zyl, J.J. and H.A. Zebker "Radar polarization signatures of vegetated areas". In *IGARSS'87 Remote Sensing: Understanding the Earth as a System*, pages 835–837, Ann Arbor, 1987.

[10] van Zyl, J.J. "Unsupervised classification of scattering behavior using radar polarimetry data". *IEEE Trans. Geosci. Rem. Sensing*, 27(1):36–45, 1989.

[11] Raney, R.K. "Comments on SAR calibration". January 1988. (Private communication).

[12] Ulaby, F.T., T.F. Haddock, and R.T. Austin. "Fluctuation statistics of millimeter-wave scattering from distributed targets". *IEEE Trans. Geosci. Rem. Sensing*, 26(3):268–281, 1988.

[13] Freeman, A., C. Werner, and Y. Shen. "Calibration of multipolarization imaging radar". In *IGARSS'88 Remote Sensing. Moving Toward the 21st Century*, pages 335–339, 1988.

[14] Hartl, P. and S. Bhagavathula. "An attempt to calibrate airborne SAR imagery using active radar calibrators and ground-based scatterometers". In *IGARSS'87 Remote Sensing: Understanding the Earth as a System*, pages 501–508, 1987.

[15] Jackson, P L , R W Larson, A Dias, and D B Fraczak "Digital processing and absolute radar calibration of a stripmap synthetic aperture radar". In *IGARSS'84 Remote Sensing: From Research Toward Operational Use*, pages 611–614, August 1984.

[16] Hoogeboom, P. "On the use of distributed targets for the calibration of imagery from the European SAR-580 campaign". In *IGARSS'83 Remote Sensing: Extending Man's Horizons*, pages 4.1–4.6, 1983.

[17] Hoogeboom, P., P. Binnenkade, and L M.M. Veugen. "An algorithm for radiometric and geometric correction of digital SLAR data". *IEEE Trans. Geosci. Rem. Sensing*, GE-22(6):570–576, 1984.

[18] Larson, R.W., P.L. Jackson, and E.S. Kasischke. "A digital calibration method for synthetic aperture radar systems". *IEEE Trans. Geosci. Rem. Sensing*, 26(6):753–763, 1988.

[19] Kasischke, E.S. and G.W. Fowler. "A statistical approach for determining radiometric precisions and accuracies in calibration of synthetic aperture radar imagery". *Int. J. Remote Sensing*, –, 1988. 42p. (to be published).

[20] Attema, E. "Engineering calibration of the ERS-1 active instrumentation in orbit". In *IGARSS'88 Remote Sensing: Moving Toward the 21st Century*, pages 859–862, 1988.

[21] Livingstone, C.E., A.L. Gray, R.K. Hawkins, R.A. Deane J.G. Halbertsma, and R.B. Olsen. "The CCRS C-band radar – System description and test results". In *Proceedings of 11th Canadian Symposium on Remote Sensing*, pages 503–518, Waterloo, Ontario, 1987.

[22] Dobson, M.C., F.T. Ulaby, D.R. Brunfeldt, and D.N. Held. "External calibration of SIR-B imagery with area extended and point targets". *IEEE Trans. Geosci. Rem. Sensing*, GE-24(4):453–461, 1986.

[23] Gray, A.L , R.K. Hawkins, C.E. Livingstone, J. Campbell, E. Attema, and J.C. Morin. "Towards a calibration of the CCRS airborne SARs". In *IGARSS'88 Remote Sensing: Moving Toward the 21st Century*, pages 707–709, 1988.

[24] Raney, R.K. "SNR in SAR". In *IGARSS'85*, pages 994–999, 1985.

[25] Blacksmith JR., D.R., R.E. Hiatt, and R.B. Mack. "Introduction to radar cross-section measurements". *Proc. IEEE*, 53(8):901–920, 1965.

[26] Lukowski, T.I., Hawkins, R.K., B. Brisco, R. Brown, P. Daleman, and R. Ford. "The Saskatoon SAR calibration experiment". In *IGARSS'89 Remote Sensing: An Economic tool for the '90's*, 1989.

[27] Kasischke, E.S., D.R. Sheen, and G.F. Fowler. "Radiometric calibration of airborne SAR data". In *IGARSS'88 Remote Sensing: Moving Toward the 21st Century*, pages 711–714, 1988.

[28] Hawkins, R.K. "Determination of antenna elevation pattern for airborne SAR using the rough target approach". 1989. (in progress).

Acknowledgements

R.K. Hawkins would like to thank the *Science and Technology Agency*, STA, of Japan, and his colleagues at *The Communications Research Laboratory*, Koganei-shi, Tokyo, for their support during the tenure of his fellowship.

Approach to Derivation of SIR-C Science Requirements for Calibration

Pascale Dubois, Diane Evans, Anthony Freeman, Jakob van Zyl

Jet Propulsion Laboratory
California Institute of Technology

I. Abstract

Many of the experiments proposed for the forthcoming SIR-C mission require calibrated data, for example those which emphasize 1) deriving quantitative geophysical information (e.g. surface roughness and dielectric constant), 2) monitoring daily and seasonal changes in the earth's surface (e.g. soil moisture), 3) extending local case studies to regional and worldwide scales, and 4) using SIR-C data with other spaceborne sensors (e.g. ERS-1, JERS-1 and Radarsat). There are three different aspects to the SIR-C calibration problem, radiometric and geometric calibration which have been previously reported, and polarimetric calibration which is the topic discussed in this paper. We report here on a study undertaken to determine science requirements for polarimetric calibration. A model describing the effect of miscalibration is presented first, followed by the description of specific experiments and the resulting requirements levied on the system. We also discuss the effects of miscalibration on some commonly used measures which can be extracted from SAR data. We conclude that polarimetric calibration requirements are strongly application dependent and a compromise has to be made to maximize the domain of application of the data without prohibitive cost impact. A set of numbers summarizing SIR-C polarimetric calibration goals will conclude this paper.

Keywords: Calibration Requirement, SIR-C

I. Introduction

One of the most important problems confronting scientists is that of understanding the earth as a system. As a step in this direction, investigators analyzing Shuttle Imaging Radar (SIR-C) data will perform a series of geoscientific experiments leading to understanding the processes governing the biogeochemical, climatic and hydrologic cycles of the earth. These large-scale physical process studies will rely on our ability to: 1) extract biogeophysical parameters from SIR-C data; 2) intercompare multitemporal SIR-C data sets, and 3) analyze SIR-C data sets in conjunction with other data sets. Deriving quantitative geophysical information (e.g. surface roughness and dielectric constant), monitoring daily or seasonal changes in the Earth's surface or cover (e.g. soil moisture or plant biophysical parameters), and extending local case studies to regional or world-wide scales, using SIR-C data alone or in conjunction with other spaceborne radars (e.g. ERS-1, J-ERS-1, Radarsat, Eos SAR) all require calibrated data.

There are two different aspects to the SIR-C calibration problem, radiometric calibration, which allows scientists to compare backscatter coefficients, and polarimetric calibration which allows quantitative theoretical modeling. Polarimetry requires precise measurement of the phase and amplitude of a signal on four different data channels. Therefore the data have to be calibrated in both amplitude and phase. Radiometric calibration requirements have been outlined in the SIR-C Science Plan [1]. The study described in this paper is an attempt at deriving the science requirements on the polarimetric calibration for the SIR-C mission. A model representing the effects of the residual distortion of the polarimeter data after calibration is presented. This distortion model is then applied to specific examples of possible experiments. The approach used makes no pretense of analysing all the possible applications of the SIR-C data set. Further study is required to determine the feasibility of additional experiments. The applications we cover in this study are polarization synthesis and determination of surface roughness. The effects of system distortion on some measures commonly used in analysis of polarimetric SAR images are also discussed. The terminology and definitions have previously been defined [2], [3] and will not be repeated in this paper.

2. Model Description

The model is an attempt at representing the effects of the residual distortion remaining after the calibration process. The results of this study will therefore put requirements on the calibration process. Let S be the scattering matrix characterizing the backscatter properties of an area.

$$S = \begin{pmatrix} S_{hh} & S_{hv} \\ S_{vh} & S_{vv} \end{pmatrix}$$

where S_{hh}, S_{hv}, S_{vh} and S_{vv} are complex numbers.

The effects of the residual miscalibration of the system after the calibration process are modeled as follows. The scattering matrix is multiplied by two matrices representing the distorting effect of the receiving path and the emitting path (the ideal residual miscalibration matrices E_R and E_T are the identity matrix). A noise matrix is added to the system.

$$S_D = E_R \; S \; E_T \; + \; N \qquad (2.1)$$

with:

$$E_R = \begin{pmatrix} 1 & \epsilon_4 \\ \epsilon_3 & a_2 \end{pmatrix}, E_T = \begin{pmatrix} 1 & \epsilon_1 \\ \epsilon_2 & a_1 \end{pmatrix} \text{ and } N = \begin{pmatrix} N_{hh} & N_{hv} \\ N_{vh} & N_{vv} \end{pmatrix}$$

where E_R is the residual miscalibration of the receiver system, E_T the residual miscalibration of the transmitter system and N the noise matrix. a_1 and a_2 represent the mismatch between the like-pol channels. They are complex numbers to account for a phase difference (different path length) and an amplitude difference (difference in gain). The ϵ's are also complex numbers. They represent the crosstalk between channels and are a measure of the channel isolation. The N's in the noise matrix are independent random variables with a Gaussian distribution, zero mean and a variable standard deviation σ. The model was further simplified by setting $a_1 = a_2 = a$ and $\epsilon_i = \epsilon$ for all i.

3. Polarization Synthesis

When the scattering matrix is known, we can synthesize any polarization setup of the antennas [3]:

$$V = h_r^T \; S \; h_t \text{ and } P = VV^* \qquad (3.1)$$

where V is the field, P its power, h_r and h_t the polarization vectors of the receiving and transmitting antennas, S the scattering matrix. Polarization synthesis is a procedure used in many applications such as optimal polarization [4] and polarization filtering [5]. Given a scattering matrix S, we compute the distorted Stokes matrix S_d. We then compare the powers (P and P_d) obtained from a given pair of polaization vectors (the formation of P and P_d is illustrated in Fig. 3-1). In order to quantitatively estimate the impact of the distortion on the synthesis process, we define a normalized RMS error δ as follows.

$$\delta = \sqrt{\frac{\int_D (P - P_d)^2}{\int_D P^2}} \qquad (3.2)$$

where the domain D over which the integration is performed is the complete set of possible polarization setups of the system, i.e., the orientation angles of the receiving and transmitting antennas vary independently between 0° and 180° as the ellipticity angles vary between −45° and 45°. δ is a normalized RMS error in synthesis.

We investigated several target types: Ocean, smooth lava, rough lava, a forested area and an urban area. Table 3.1 summarizes the distortion resulting from 6 cases of miscalibration for different types of target. The ocean and the lava are the most sensitive targets. The S_{vv} term in the case of the ocean is much larger than any other term of the scattering matrix; a small channel imbalance or crosstalk distortion has therefore a large impact on the synthesis. A similar but not as pronounced effect occurs for smooth lava. Note the RMS error is very sensitive to crosstalk error. For example, a change in crosstalk from $-30dB$ to $-25dB$ results in a change in RMS error of $2dB$. The RMS error is not as sensitive to the other parameters.

Next, we vary the standard deviation of the noise matrix which is related to the Signal to Noise Ratio (SNR) (Fig. 3-2) . To keep the contribution of noise to the total error δ below -12.5 dB, we need an SNR of better than $20dB$ for all targets investigated.

4. RMS Height determination

Many geoscientific investigations require information about surface roughness. For example, The RMS height of the Cima lava flow in the Mojave desert of California decreases at first due to weathering and deposition of wind-blown material, then stream development reverses the phenomenon (see Fig. 4-1, which shows RMS heights of lava flows plotted against age) [6]. Therefore knowledge of the RMS height is an important factor in dating flows in a relative way.

Quantitative interpretation of surface roughness relies on the Stokes matrix representation. It is a more complete way to describe the scattering properties of a surface than the scattering matrix as it can include some depolarization effects like multiple scattering.

In this example, we compare the RMS heights obtained from a Stokes matrix and its distorted counterpart. Given an RMS height, a Stokes matrix M is first obtained using a second-order Bragg mode developed at JPL. A distorted Stokes matrix M_d is computed and then inverted to give a distorted estimate of RMS height (the algorithm used is shown in Fig. 4-2). Four RMS heights were studied: 0.1cm, 2cm, 4cm and 6cm. Figure 4-3 plots the variation of the estimated RMS height as a function of amplitude of a and ϵ. The phase of a has no effect in this case. When the amplitude of a is $0dB$ or the amplitude of ϵ is $-30dB$, the error in the estimate of RMS height is negligable. If an estimation accuracy of 2cm for the RMS height is desired, the requirements on the system are a maximum amplitude of $0.3dB$ for parameter a and a maximum amplitude of $-30dB$ for ϵ. The operation that inverts the Stokes matrix to give the RMS height relies only on the normalized Stokes matrix. The absolute backscatter coefficient (the 1,1 element of the Stokes matrix) is not used to determine the RMS height.

5. Science Calibration Goals for SIR-C

Both examples developed in this paper show the importance of ϵ, the crosstalk term and we conclude that the residual crosstalk ϵ should not exceed $-30dB$. While the derivation of surface roughness may be less stringent, polarization synthesis requirements on a are : Amplitude of a less than $\pm 0.2dB$ ($0.4dB$ two-way), Phase of a less than 5° (10° two-way). Recommended calibration goals are summarizes in Table 5.1.

The next section analyzed the impact of these calibration requirements on some parameters extensively studied in the polarimetric literature [7].

6. Miscalibration and noise distortion on some commonly used parameters

Up to now,only a few polarimetric parameters such as the relative amplitude of S_{hh}/S_{vv}, the phase difference between S_{hh} and S_{vv}, have generated a widespread interest in the scientific community . In this section, the expected distortions due to miscalibration and due to noise are evaluated separately for these parameters.

In the Jet Propulsion Laboratory DC-8 airborne SAR system, which is in many respects a test-bed for SIR-C, the output products are converted to a Stokes matrix format [8]. In this process, the scattering matrix is symmetrized. After symmetrization, neglecting higher order terms, (assuming $S_{hv} < S_{hh}$ or $< S_{vv}$ and ϵ small) (2.1) can be written as follows:

$$S^d_{hh} \simeq S_{hh} + N_{hh}$$

$$S^d_{hv} = S^d_{vh} \simeq \frac{a}{2}(S_{vh} + S_{hv}) + \epsilon(S_{hh} + aS_{vv}) + \frac{1}{2}(N_{hv} + N_{vh})$$

$$S^d_{vv} = a^2 S_{vv} + N_{vv} \qquad (6.1)$$

where superscript ^{d} indicates distortion. Note the dependence of cross-pol measurements on the system isolation, ϵ, and the relative levels of the like- and cross-pol backscatters. Because of this target dependence, it is difficult to establish a singular requirement for the cross-pol measurements.

6.1 Distortion due to Miscalibration:

The distortion due to residual miscalibration is analyzed here. The N matrix in (2.1) is set to zero as the impact of noise will be studied in the next section. From (6.1), we have, for a comparison of the like-pol measurements:

$$\frac{S^d_{vv} \cdot S^{d*}_{vv}}{S^d_{hh} \cdot S^{d*}_{hh}} \simeq \frac{S_{vv} \cdot S^*_{vv}}{S_{hh} \cdot S^*_{hh}} \; |a^4| \qquad (6.2)$$

The relative amplitude of S_{vv}/S_{hh} is multiplied by the channel imbalance term squared. Therefore, a channel imbalance of $\pm 0.2dB$ will result in a relative error of $\pm 0.4dB$ which is an acceptable error.

The phase difference error between S_{hh} and S_{vv} due to miscalibration can also be computed.

$$Arg(S^d_{hh}\ S^{d*}_{vv}) \simeq Arg(S_{hh}\ S^*_{vv}) + 2\ Arg(a) \qquad (6.3)$$

The bias between the phase differences introduced by the miscalibration is 10° for the SIR-C experiment. Relying on our experience in polarimetry, this distortion is acceptable. Phase comparisons between like- and cross-pol measurements are again complicated by their dependence on the target properties, as well as system distortion.

Because the distortion model is deterministic and in scattering matrix format, neither the channel imbalance nor the crosstalk will create a depolarization effect. (i.e. the pedestal height of the polarization signatures will not changed).

6.2 Distortion due to noise:

The noise is now assumed to be the only distortion factor; the E_R and E_T matrices in (2.1) are replaced by identity matrices.

The elements of the noise matrix N are assumed to be independent random Gaussian variables with zero mean but the same standard deviation, σ. It follows that the Stokes matrix corresponding to the noise matrix is of the form:

$$< M_N > = \begin{pmatrix} \sigma^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad (6.4)$$

where $< >$ denotes spatial averaging. This is equivalent to adding the constant σ^2 to the total signal power. While all other terms in the noise Stokes matrix have zero mean, they still introduce distortion in the target Stokes matrix.

We also have:

$$< S^d_{hh} S^{d*}_{hh} > \; = \; < S_{hh} S^*_{hh} > + \; \sigma^2 \qquad (6.5)$$

$$< S^d_{vv} S^{d*}_{vv} > \; = \; < S_{vv} S^*_{vv} > + \; \sigma^2 \qquad (6.6)$$

The average value of the return power is increased by σ^2 for $S_{hh}S^*_{hh}$ and $S_{vv}S^*_{vv}$. It can also be shown that noise has a similar effect on the average power of the signal synthesized for any polarization of the receive and transmit antennas.

For the phase comparison of the like-pol terms, we have:

$$Arg(< S^d_{hh} S^{d*}_{vv} >) =$$

$$Arg(< S_{hh} S^*_{vv} > + < N_{hh} N^*_{vv} > + < S_{hh} N^*_{vv} > + < N_{hh} N^*_{vv} >)$$

$$= Arg(< S_{hh} S^*_{vv} >) \qquad (6.7)$$

assuming that the noise terms are uncorrelated with each other and with the S_{hh} and S_{vv} terms. The argument of the spatially averaged cross-product $S_{hh}S^*_{vv}$ is independent of the noise level but the RMS variation

$$\delta = \sqrt{\frac{\int_D (P - P_d)^2}{\int_D P^2}} \qquad (3.2)$$

where the domain D over which the integration is performed is the complete set of possible polarization setups of the system, i.e., the orientation angles of the receiving and transmitting antennas vary independently between 0° and 180° as the ellipticity angles vary between −45° and 45°. δ is a normalized RMS error in synthesis.

We investigated several target types Ocean, smooth lava, rough lava, a forested area and an urban area. Table 3.1 summarizes the distortion resulting from 6 cases of miscalibration for different types of target The ocean and the lava are the most sensitive targets The S_{vv} term in the case of the ocean is much larger than any other term of the scattering matrix; a small channel imbalance or crosstalk distortion has therefore a large impact on the synthesis. A similar but not as pronounced effect occurs for smooth lava. Note the RMS error is very sensitive to crosstalk error. For example, a change in crosstalk from $-30dB$ to $-25dB$ results in a change in RMS error of $2dB$. The RMS error is not as sensitive to the other parameters.

Next, we vary the standard deviation of the noise matrix which is related to the Signal to Noise Ratio (SNR) (Fig. 3-2) . To keep the contribution of noise to the total error δ below -12.5 dB, we need an SNR of better than $20dB$ for all targets investigated.

4. RMS Height determination

Many geoscientific investigations require information about surface roughness. For example, The RMS height of the Cima lava flow in the Mojave desert of California decreases at first due to weathering and deposition of wind-blown material, then stream development reverses the phenomenon (see Fig. 4-1, which shows RMS heights of lava flows plotted against age) [6]. Therefore knowledge of the RMS height is an important factor in dating flows in a relative way

Quantitative interpretation of surface roughness relies on the Stokes matrix representation. It is a more complete way to describe the scattering properties of a surface than the scattering matrix as it can include some depolarization effects like multiple scattering.

In this example, we compare the RMS heights obtained from a Stokes matrix and its distorted counterpart. Given an RMS height, a Stokes matrix M is first obtained using a second-order Bragg mode developed at JPL. A distorted Stokes matrix M_d is computed and then inverted to give a distorted estimate of RMS height (the algorithm used is shown in Fig. 4-2) Four RMS heights were studied. 0 1cm, 2cm, 4cm and 6cm Figure 4-3 plots the variation of the estimated RMS height as a function of amplitude of a and ϵ. The phase of a has no effect in this case When the amplitude of a is $0dB$ or the amplitude of ϵ is $-30dB$, the error in the estimate of RMS height is negligable. If an estimation accuracy of 2cm for the RMS height is desired, the requirements on the system are a maximum amplitude of $0.3dB$ for parameter a and a maximum amplitude of $-30dB$ for ϵ. The operation that inverts the Stokes matrix to give the RMS height relies only on the normalized Stokes matrix. The absolute backscatter coefficient (the 1,1 element of the Stokes matrix) is not used to determine the RMS height.

5. Science Calibration Goals for SIR-C

Both examples developed in this paper show the importance of ϵ, the crosstalk term and we conclude that the residual crosstalk ϵ should not exceed $-30dB$. While the derivation of surface roughness may be less stringent, polarization synthesis requirements on a are : Amplitude of a less than $\pm 0.2dB$ ($0.4dB$ two-way), Phase of a less than 5° (10° two-way). Recommended calibration goals are summarizes in Table 5 1

The next section analyzed the impact of these calibration requirements on some parameters extensively studied in the polarimetric literature [7].

6. Miscalibration and noise distortion on some commonly used parameters

Up to now,only a few polarimetric parameters such as the relative amplitude of S_{hh}/S_{vv}, the phase difference between S_{hh} and S_{vv}, have generated a widespread interest in the scientific community . In this section, the expected distortions due to miscalibration and due to noise are evaluated separately for these parameters.

In the Jet Propulsion Laboratory DC-8 airborne SAR system, which is in many respects a test-bed for SIR-C, the output products are converted to a Stokes matrix format [8]. In this process, the scattering matrix is symmetrized. After symmetrization, neglecting higher order terms, (assuming $S_{hv} < S_{hh}$ or $< S_{vv}$ and ϵ small) (2.1) can be written as follows:

$$S^d_{hh} \simeq S_{hh} + N_{hh}$$

$$S^d_{hv} = S^d_{vh} \simeq \frac{a}{2}(S_{vh} + S_{hv}) + \epsilon(S_{hh} + aS_{vv}) + \frac{1}{2}(N_{hv} + N_{vh})$$

$$S^d_{vv} = a^2 S_{vv} + N_{vv} \qquad (6.1)$$

where superscript ^{d} indicates distortion. Note the dependence of cross-pol measurements on the system isolation, ϵ, and the relative levels of the like- and cross-pol backscatters. Because of this target dependence, it is difficult to establish a singular requirement for the cross-pol measurements

6.1 Distortion due to Miscalibration:

The distortion due to residual miscalibration is analyzed here The N matrix in (2.1) is set to zero as the impact of noise will be studied in the next section. From (6.1), we have, for a comparison of the like-pol measurements:

$$\frac{S^d_{vv} \cdot S^{d*}_{vv}}{S^d_{hh} \cdot S^{d*}_{hh}} \simeq \frac{S_{vv} \cdot S^*_{vv}}{S_{hh} \cdot S^*_{hh}} |a^4| \qquad (6.2)$$

The relative amplitude of S_{vv}/S_{hh} is multiplied by the channel imbalance term squared. Therefore, a channel imbalance of $\pm 0.2dB$ will result in a relative error of $\pm 0.4dB$ which is an acceptable error.

The phase difference error between S_{hh} and S_{vv} due to miscalibration can also be computed.

$$Arg(S^d_{hh}\ S^{d*}_{vv}) \simeq Arg(S_{hh}\ S^*_{vv}) + 2\ Arg(a) \qquad (6.3)$$

The bias between the phase differences introduced by the miscalibration is 10° for the SIR-C experiment. Relying on our experience in polarimetry, this distortion is acceptable. Phase comparisons between like- and cross-pol measurements are again complicated by their dependence on the target properties, as well as system distortion.

Because the distortion model is deterministic and in scattering matrix format, neither the channel imbalance nor the crosstalk will create a depolarization effect. (i.e. the pedestal height of the polarization signatures will not changed).

6.2 Distortion due to noise:

The noise is now assumed to be the only distortion factor. the E_R and E_T matrices in (2.1) are replaced by identity matrices.

The elements of the noise matrix N are assumed to be independent random Gaussian variables with zero mean but the same standard deviation, σ. It follows that the Stokes matrix corresponding to the noise matrix is of the form:

$$< M_N > = \begin{pmatrix} \sigma^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad (6.4)$$

where $< >$ denotes spatial averaging. This is equivalent to adding the constant σ^2 to the total signal power. While all other terms in the noise Stokes matrix have zero mean, they still introduce distortion in the target Stokes matrix.

We also have:

$$< S^d_{hh} S^{d*}_{hh} > \ = \ < S_{hh} S^*_{hh} > + \ \sigma^2 \qquad (6.5)$$

$$< S^d_{vv} S^{d*}_{vv} > \ = \ < S_{vv} S^*_{vv} > + \ \sigma^2 \qquad (6.6)$$

The average value of the return power is increased by σ^2 for $S_{hh}S^*_{hh}$ and $S_{vv}S^*_{vv}$. It can also be shown that noise has a similar effect on the average power of the signal synthesized for any polarization of the receive and transmit antennas.

For the phase comparison of the like-pol terms, we have:

$$Arg(< S^d_{hh} S^{d*}_{vv} >) =$$

$$Arg(< S_{hh} S^*_{vv} > + < N_{hh} N^*_{vv} > + < S_{hh} N^*_{vv} > + < N_{hh} N^*_{vv} >)$$

$$= Arg(< S_{hh} S^*_{vv} >) \qquad (6.7)$$

assuming that the noise terms are uncorrelated with each other and with the S_{hh} and S_{vv} terms. The argument of the spatially averaged cross-product $S_{hh}S^*_{vv}$ is independent of the noise level but the RMS variation

will clearly depend on the noise level in the data channels.

7. Conclusion

It should be noted that the recommended calibration requirements derived here rely on specific experiments Polarimetry is still at an investigation stage and its applications are expected to increase in the future as the scientific community develops familiarity with polarimetric data sets. It follows that the calibration requirements imposed on the SIR-C system may be over constraining for some experiments while not strict enough for some others Therefore, a complete analysis of the required calibration accuracy for every planned experiment should be compared to the calibration requirements imposed on SIR-C to see how adequate the SIR-C data will be to this experiment.

The model used here is a first step to simulate the effect of miscalibration The deterministic approach and the assumption that the distortion remains constant throughout one synthetic aperture are the main weaknesses of the model

Finally, polarimetry is not a standalone tool. For example, in the case of RMH height determination, the radiometric information can be very useful as the pixel brightness depends on the roughness of the corresponding area. If this information is used together with the polarimetric information, the resulting requirements on polarimetric calibration accuracy may become less strict In general, an efficient experiment will rely on all the available information, taking advantage of the multidimensionality of the data set multipolarimetry, multifrequency and radiometry.

Acknowledgment

The research described in this abstract was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. We would like to thank our colleagues at JPL for their constructive comments which helped improving the quality of this paper.

References

[1] Shuttle Imaging Radar (SIR-C) Science plan, JPL Publication 86-29

[2] van Zyl, Zebker, Elachi. Imaging radar polarization signatures: theory and observation, Radio Science '87, 22, no. 4, pp. 529-543.

[3] Zebker, van Zyl, Held. Imaging radar polarimetry from wave synthesis, Jour. Geophys. Res., January '87, pp. 683-701.

[4] Ionnadis, Hammers. Optimum antenna polarizations for target discrimination in clutter, IEEE Trans. Antennas and Propagation, Vol. AP-27, No. 3, May 1979.

[5] Dubois, van Zyl, Polarization filtering of SAR data, Proc. of IGARSS '89 Symposium, 89

[6] Farr :Personal communication.

[7] Ulaby, Held. Relating polarization phase difference of SAR signals to scene properties, IEEE trans. GRS, Vol GE-25, pp. 83-87, 1987.

[8] Dubois, P. C. and L. Norikane, Data Volume Reduction for Imaging SAR Polarimetry, Proc. of IGARSS '87 Symposium, 87

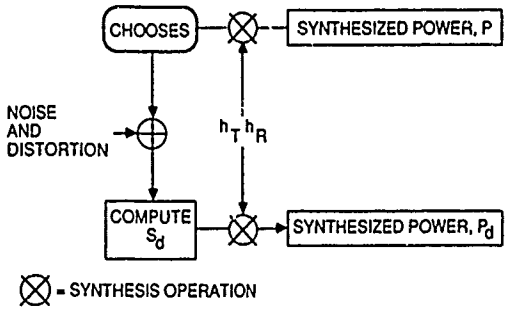


Figure 3.1: Synthesis Test Procedure

DISTORTION LEVEL	a = 0.2 dB 5° ε = -30 dB	a = 0.3 dB 5° ε = -30 dB	a = 0.2 dB 5° ε = -25 dB	a = 0.3 dB 5° ε = -25 dB	a = 0.2 dB 4° ε = -30 dB	a = 0.2 dB 6° ε = -30 dB
OCEAN	-9.87	-9.80	-7.91	-7.89	-10.08	-9.65
LAVA	-9.59	-9.39	-7.70	-7.62	-9.79	-9.36
FOREST	-11.34	-11.02	-9.23	-9.09	-11.37	-11.33
URBAN	-12.40	-12.11	-11.03	-11.03	-12.90	-12.00

Table 3.1: Normalized RMS error on synthesis (δ) for different target types for some miscalibration parameters

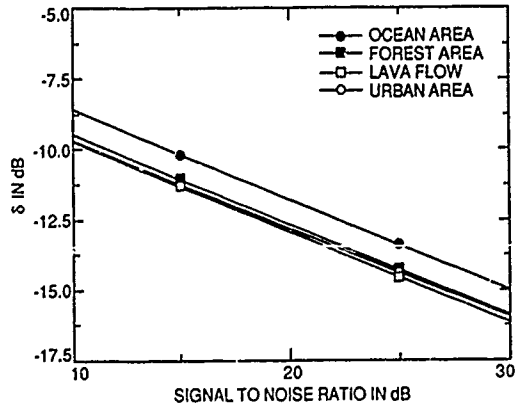


Figure 3.2: Normalized RMS error on synthesis (δ) for different target types as a function of SNR.

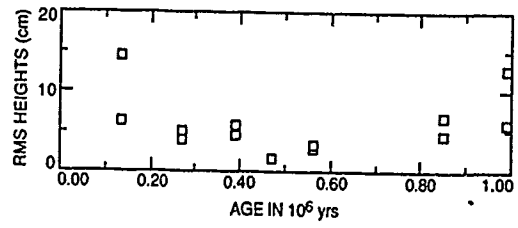


Figure 4.1: Variation of RMS Height as a Function of Age of Lava.

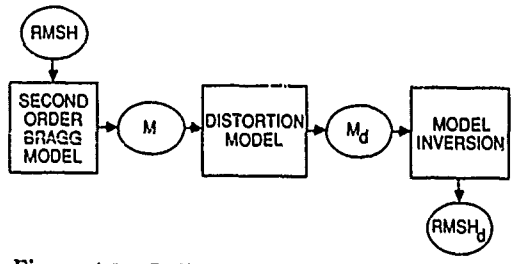


Figure 4.2: RMS height Test Procedure

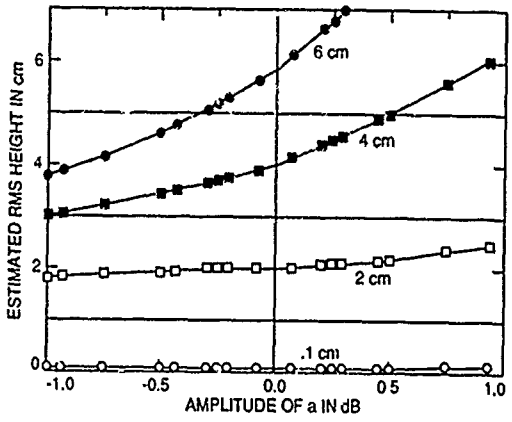


Figure 4.3: Estimated RMS height versus Amplitude of channel imbalance.

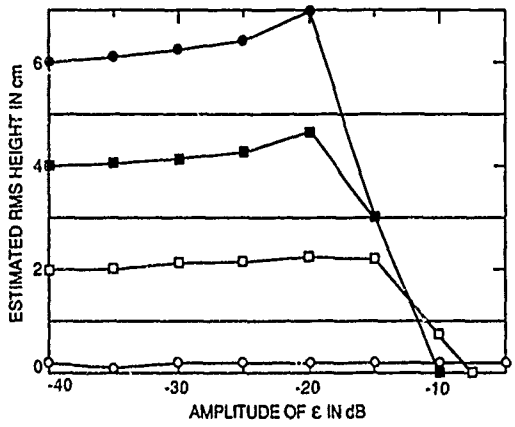


Figure 4.4: Estimated RMS height versus Amplitude of crosstalk (ϵ).

Polarimetric Amplitude Imbalance (2-way) $\pm 0.4dB$
Polarimetric Relative Phase Calibration (2-way) $\pm 10^{\circ}$
Polarimetric Cross-talk error $< -30dB$
Signal-to-Noise Ratio $> 20dB$

Table 5.1: SIR-C Calibration Goals

SIR-C CALIBRATION PROCEDURES

A. FREEMAN AND J. C. CURLANDER
JET PROPULSION LABORATORY
4800 OAK GROVE DRIVE
PASADENA, CA 91109
TELEPHONE: (818)354-1887
FAX: (818)354-3437
TELEX: 67-5429

The Shuttle Imaging Radar-C (SIR-C) mission is part of an on-going spaceborne SAR program by NASA/JPL, which will culminate in the EOS polar platform SAR towards the end of the century. SIR-C is unique in that it will be the first multi-frequency, multipolarization SAR to operate from space. The sensor will collect fully coherent, polarimetric backscatter information (HH, HV, VH and VV polarization) at two frequencies (L- and C-Band) about the earths surface and will also collect backscatter data in one X-Band channel.

This paper will describe project-recommended approaches for calibrating SIR-C images, using calibration devices such as corner reflectors transponders and receivers deployed on the ground. The recommendations will include the correct procedures to be adopted by SIR-C Principal Investigations to locally calibrate their image data, near to their site of scientific interest. The paper will also describe the project calibration plans for SIR-C, which will include both external calibration at major sites around the SIR-C orbit track, and calibration routines internal to the sensor itself. Finally, the expected calibration accuracies for SIR-C images will be presented.

The research described in this paper was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

PROGRESS IN CALIBRATION: SAR-580 TO ERS-1

A M Smith D G Corr I Ward
SD-SCICON
Sanderson House
49 Berners Street
London W1P 4AQ
UK

Tel 01 580 5599
Fax 01 580 7716
Telex 24293 SCICON G

J-P Guignard
ESTEC
Noordwijk
The Netherlands

M Doherty
ESRIN
Frascati
Italy

Analysis of the SAR-580 Calibration Campaign data collected in 1981, brought in to attention a number of factors which appear to have influenced both the design of the ERS-1 SAR and its associated image processing techniques.

The principal problem with the SAR-580 data was that due to an instability in the transmitter system the transmitted waveforms were very different from their nominal specification. Additionally, the waveforms varied from pass to pass. Fortunately replica pulses were recorded as part of an on board echo simulator. This replica data was extracted and used as the replicas for range compression.

In the analysis of the images of point targets a non linear behaviour was observed for the largest targets. This was traced to saturation of the ADC by the echo signals. Further theoretical work predicted that non linear behavious would particularly occur with the ERS wave mode data, due to the limited dynamic range.

As a result of the analysis of the many target images a detailed comparison of peak and integration methods of energy measurement was brought about.

Recently attention has been given by us on how to correctly use the replica data transmitted by ERS-1, and how to correct for ADC non linearities.

After a brief review of the findings from our previous work, this paper describes alternative approaches to range compression and ADC correction which are intended to provide image products with very high accuracy.

RESULTS OF THE 1988 NASA/JPL AIRBORNE SAR CALIBRATION CAMPAIGN

A. Freeman, C. Werner and J.D. Klein

Jet Propulsion Laboratory
4800 Oak Grove Drive
Pasadena, CA 91109
USA

ABSTRACT

During Spring of 1988, the NASA/JPL multi-frequency, multipolarization SAR flew in a series of calibration experiments over the Goldstone dry lake bed. At Goldstone, an array of calibration devices was deployed, including dihedral and trihedral corner reflectors, polarimetric active radar calibrators (PARC's), passive receivers and CW tone generators. The aim of the campaign was to calibrate both amplitude and phase of the resulting SAR images, over long and short time scales.

In this paper, the results of the analysis of the calibration data collected in Spring 1988 are presented. Trihedral corner reflector signatures and certain image background measures are used to externally calibrate relative amplitude and phase between polarizations at a given frequency, and to calibrate across frequencies. Assessments are made of the calibration accuracy as a function of image frame position for each frequency, and the stability of the radar calibration over long and short timescales.

Finally, the problem of obtaining calibrated images from the standard DC-8 SAR products is discussed.

KEYWORDS : SAR, Calibration

I. INTRODUCTION

There is a growing emphasis on the need for calibrated SAR imagery within the microwave remote sensing community. Calibrated images would significantly advance the use and testing of models to study backscatter phenomena, and in relating the image pixel values to geophysical parameters. With the launch of several spaceborne SAR's, such as E-ERS-1, J-ERS-1 and Radarsat, ahead in the 1990's, and with the advent of a new generation of multi-frequency, multi-polarization SAR's, such as SIR-C and the Eos SAR instrument, the need becomes more urgent. The techniques for routine accurate and precise calibration of SAR image products should be available well before the first stream of spaceborne SAR imagery starts to flow. The best approach to develop these new techniques and start the learning process is to use aircraft SAR systems as test beds.

The NASA/JPL DC-8 SAR took part in a series of calibration experiments at Goldstone Dry Lake bed in Spring 1988. An additional calibration experiment was carried out at Fairbanks, Alaska in March of that year. The aims and execution of the DC-8 SAR calibration campaign have been described in (Freeman,1988). The DC-8 SAR, which is capable of simultaneously collecting multi-polarization and multi-frequency data, has been described in (Held,1988).

In summer, 1989 the DC-8 SAR will take part in a series of campaigns, in the USA and Europe. Calibration experiments are planned for sites at Goldstone, near Mt. Shasta, CA and at several European sites, including DLR in W.Germany, Bari in Italy and Flevopolder in The Netherlands.

In this paper, the results of a series of experiments aimed at calibrating the multi-frequency, multi-polarization DC-8 SAR images over an entire season are summarized. Image quality, radiometric calibration, polarimetric calibration (amplitude and phase), cross-frequency calibration and antenna pattern measurements are included in the analysis.

II. IMAGE QUALITY

Nominal values for the image quality parameters of the NASA/JPL SAR are given in Table 1. These values were predicted from systems tests and analysis prior to the operational stage of the system. In Table 2, typical results for the quality of the measured range impulse response functions are presented. The responses were obtained from the Goldstone trihedral corner reflectors data . Over the data gathering season, very little variability of the range impulse response was apparent.

The differences between the nominal and measured range impulses are probably due to the introduction of a weighting function in the processor, which was not taken into account in the pre-flight systems analysis and whose effects have not been estimated using simulated data. Since the measured quantities do not vary, the discrepancies between the two sets of values are not critical.

In Tables 3 and 4 the best and worst case examples of the azimuth impulse response functions, as measured from Goldstone Trihedral data, are presented. In the best-case results, the system performance is fairly close to the nominal, except for a slight broadening and drop in sidelobe level, due to the weighting imposed by the azimuth prefilter in the SAR processor.

In the worst-case, the L-Band response is slightly broader than the best-case, mainly due to a degradation in resolution for a few of the trihedrals in the array, centered around boresight (50° Incidence). This was a rare

occurrence, not repeated in the rest of the Goldstone data. For C-Band, the impulse response has a tendency to divide into two narrow peaks Hence the worst-case ISLR of +8.4 dB. In this extreme situation, the ISLR calculation used in this paper breaks down, so that the measure becomes of little value. For the P-Band worst-case, the impulse response is degraded so badly that the resolution is broadened by a factor of 4. In this situation, both the ISLR and PSLR measures used in this paper break down. The positive ISLR value does not necessarily mean that there is more energy in the trihedral responses concentrated outside the main lobe, than within it, just that the estimation of where those limits on the main lobe actually lie has failed The most probable sources of these errors are mis-estimation of the drift angle for the C-Band data and nonlinear frequency modulation caused by aircraft motion at P-Band (which has the longer aperture synthesis time in azimuth). For both frequencies, an error in the drift angle can lead to an error in the range cell migration correction in the processor.

III. RADIOMETRIC CALIBRATION

Tables 5 and 6 are the summary tables of the radiometric calibration performance of the NASA/JPL DC-8 SAR during Spring 1988, as estimated from the trihedral corner reflector data.

In Table 5 we present the estimates of short- and long-term calibration precisions. The short-term values are the standard deviations taken across the swath of the trihedral HH and VV polarization measurements for a given image. The results presented in Table 5 are the worst case, taken over the data-gathering season. These fluctuations could be due to a variety of causes: imperfections in the corner reflector construction; ripples in the radar elevation antenna pattern; or variations due to the background clutter and system noise. The long-term values [x ($\pm$y)] in Table 5 are the range of variation (maximum x+y, minimum x-y) of the mean value (for each image data set) over the data-gathering season. The short-term (1σ) variations for each image are not included in the calculation of the range of variation.

Assuming that the trihedral corner reflector cross-sections are, on average, as given in (Ruck,1970) we can easily estimate the appropriate correction factors to obtain RCS or σ given in Table 6. The uncertainty estimates given in the table are a root-sum-square of the worst-case short-term relative variation and the long-term variation estimates So, in the case of the LHH value, after April 4th,

$$\text{Correction factor, } C_F = 72.2 - 29.1 \pm (1.9^2 + 1.4^2)^{1/2} = 43.1\ (\pm 2.4)\ \text{dB} \qquad (1)$$

The C_F values for before April 4th, which include correction for a change in one of the system operation parameters (the caltone level), are given in the first column of Table 6.

To apply these correction factors to a given image, in order to estimate the radar cross-section, σ, the following formula should be used:

$$\hat{\sigma} = 10 \log_{10}\left[DN \sin\theta_i \cdot \frac{R_0^2}{(8 \times 10^3)^2}\right] - C_F \qquad (2)$$

where DN is the power representation of the image pixel value ($= I^2 + Q^2$ for the high-resolution complex image product) and θ_i is the incidence angle.

To estimate normalized radar cross-section, σ_0, the following formula should be used:

$$\hat{\sigma}_0 = 10 \log_{10}\left[DN \cdot \frac{R_0^2}{(8 \times 10^3)^2} \cdot \frac{1}{\rho_a \rho_r}\right] - C_F \qquad (3)$$

Where the product $\rho_a \rho_r$ is the image resolution cell area, nominally 4.0 x 11.0 m (or 16.4 dBm^2).

The correction formulae given above should be used with caution: they strictly only apply to the calibration data analyzed in this paper. Extensions to further data sets may not be valid.

From Tables 5 and 6 it is immediately clear that only the L-band radar performed well in a radiometric calibration sense, when the results are compared with the goals in (Freeman,1988). The ~4.0 dB short-term variability in the C-Band signatures is, we believe, due to imperfection in the construction of some of the corner reflectors. Some of the older reflectors in the Goldstone array gave consistently bad results at C-band from day-to-day and pass to pass, thus increasing the short-term variation across the swath. The long-term variability in the C-Band data is more serious, since it indicates a severe problem in the C-Band radar stability. This problem may lie in the level of transmitted power, the receivers or the input caltone. At P-Band, the problems of calibration are compounded by two factors: firstly, the trihedral responses, particularly in the HH images, behave strangely, and secondly, the low signal-to-clutter background level. The former effect, which is marked by a rapid (6-10 dB) fall-off in the HH response a few degrees array from the boresight angle, is probably due to the low size/wavelength ratio of the trihedrals at P-Band, which may allow edge effects to dominate in the trihedral signature. The latter effect, characterized by an SCR of typically ~25 dB, could be a significant additional error source.

The obvious conclusion to be drawn from Tables 5 and 6 is that only in the case of L-Band is there any hope of radiometrically calibrating the DC-8 SAR images without the assistance of a calibration device such as a trihedral corner reflector within the scene. Even then, the corner reflector must be of good quality for C-Band calibration and imaged at very close to boresight for the P-Band.

It had been hoped, before the calibration campaign began, that the external calibration results would be compared with internal calibration measurements. In the internal calibration pre-flight measurements of system parameters such as antenna gain are included with in-flight measurements of transmitter power and receiver gain, for example, in the appropriate form of the radar equation to estimate the DN value corresponding to a target of given RCS. Unfortunately, a vital part of the radar chain, the SAR processor, remains uncalibrated, so this part of the experiment has not yet been completed.

In the above calculations and discussion the effects of system noise on the radiometric calbration have been ignored. There are two reasons for this: firstly, since the SAR processor is uncalibrated, we have no direct measurements of the actual image noise level; secondly, the noise-equivalent sigma-zeros estimated from preflight systems analysis were quite low, $\lesssim$ -35 dB, -30 dB and -35 dB for L-, C- and P-Band respectively.

IV. POLARIMETRIC CALIBRATION RESULTS

In this section, the results for polarimetric calibration of the DC-8 SAR images are presented, frequency by frequency. Because the data collected allows us to look at the variation in the polarimetric parameters on both long- and short-term bases, the results are divided in this fashion.

Table 7 contains a summary of the polarimetric calibration results for L-Band. The short-term variations given are the worst-case, taken over the whole season, of the standard deviations across the swath of the amplitude ratios and phase differences, as extracted from the trihedrals and image background. The long-term variations are just the range of variation of the mean value for each ratio, taken over the data-gathering season.

From Table 7, we can draw the following conclusions about the polarimetric calibration performance of the L-Band radar:

i) The backscatter ratios, L HH/VV, L HV/VH and L HV/HH are fairly stable within an image and between data sets

ii) The L-Band cross-pol leakage is high (~ -15.8 dB) and varies significantly with incidence.

iii) The L-Band cross-pol leakage has a significant effect on the relationship between the like- and cross-pol phase.

iv) Trihedral corner reflectors and image background phase calibration approaches (Sheen,1989) give similar results at L-Band. However, the trihedral signatures are sometimes subject to wild phase fluctuations, with unknown cause

v) The L HH/VV phase difference varies greatly from day to day. The L HV/VH phase difference, on the other hand, does not

vi) The above results indicate that the L-Band DC-8 SAR images are, in general, calibratable in a polarimetric sense.

The polarimetric radar characteristics revealed by the results given in Table 8 for C-Band can be summarized as follows:

i) An imbalance between the HH/VV amplitude ratios which varied significantly across the swath (~3dB over 10 degrees of incidence angle) was apparent in most of the C-band data This short-term variation was attributed to the C-band antenna pattern and a radiometric correction curve (derived by assuming reversal of the physical antenna) was applied to the C-=band results. After correction, the HH/VV and HV/VH amplitude ratios are also offset from 0 dB, but do not vary greatly in the short-term. In the long-term, however, they show significant variation from day-to-day and pass-to-pass These results indicate that the C-Band radar system was not radiometrically stable and may vary over comparatively short time scales (~20 mins).

ii) The C-Band cross-pol isolation performance, ~ -28 dB, is very good.

iii) The HH/VV and HV/VH phase differences are stable within image frames and over the May and June data set, but can vary widely if the February and March data are included.

iv) The HV/HH phase difference is generally distributed uniformly between ±π.

The polarimetric calibration measurements for P-Band are given in Table 9 The main conclusions to be drawn from this table are as follows:

i) Interference in the P-Band data causes the terms involving cross pol measurements to fluctuate widely.

ii) The P-Band cross-pol isolation performance, as measured from the trihedrals, is ~ -22.0 dB, which is acceptable. There is some evidence to suggest that the performance may in fact be better than this, but the presence of interference precludes a more accurate measurement.

iii) The P-Band HH/VV amplitude ratios extracted from trihedrals are badly affected by the variations of the HH trihedral response noted in Section VI. However, their median value is fairly close to 0.0 dB and the HH/VV image background ratio indicates a degree of stability in this measurement.

iv) The phase differences between HH/VV and HV/VH, as measured from the image background, are relatively stable over the data-gathering season.

V. CROSS-FREQUENCY CALIBRATION RESULTS

In Table 10 results of the cross-frequency comparison of the individual trihedral DN values are summarized The measurements were obtained by calculating statistics on the frequency ratio for each trihedral signature over the data sets. The short-term results are the worst-case of the standard deviations taken across each image swath. Also given in the table are the expected short-term variations, obtained by a root-sum-square of the short-term variations for each frequency, from Tables 6-9. The long-term results in Table 10 are the mean values for each ratio taken over all the data sets, together with the range of variation about that mean. In the fourth column, the expected values for the mean long-term results are given. These expectations were derived using equation (6) and post-April 4th values for the radar parameters Results for comparisons of HH and VV values between frequencies are given, as well as comparison of the ratios of HH/VV.

The close correspondence between the measured and 'expected' short-term results in Table 10 indicates that the short-term variations in the trihedral signatures are uncorrelated Thus, the ratio of LHH/CHH, say, tends to reflect the uncertainties in the individual frequences across the swath. Only in the case of the L/C HH/VV ratio can the short-term performance be considered satisfactory.

In the long-term results, the range of variations about the mean again probably reflect the uncertainties in the individual frequencies. The mean values in the case of L/C-Band are fairly close between measured and expected, with a 2.5 dB offset between the HH values, 2.6 dB between VV values and just 0.6 dB for the HH/VV ratio. These offsets can probably be explained in terms of the previously noted variability of the C-Band radar system. The offsets between L and P-Band measurements, except in the case of the HH/VV ratio, are more severe, with unknown cause.

VI. SUMMARY AND DISCUSSION

Image quality and calibration results were derived for trihedral and image background measurements. The results were divided into four sections: image quality, radiometric, polarimetric and cross-frequency calibration

The vast majority of the standard images produced by the DC-8 SAR system are uncalibrated. To obtain polarimetrically calibrated data from standard DC-8 SAR images, the approaches put forward in (Klein, 1989) and (van Zyl,1989) are recommended, if the imaged scene contains a trihedral corner reflector. For radiometric calibration the correction factors given in this paper should be used with caution. Results indicate that the correction factors should give better results for the more stable L-band system. In general, it is recommended that trihedral corner reflectors or other known calibration devices be used for radiometric calibration, if available.

The best form of device for radiometric calibration would be a target with a polarization signature of the form,

$$S = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

or similar, since it allows radiometric calibration of both like- and cross-pol channels. If the cross-talk can be ignored, then such a device will also allow full polarimetric calibration within a frequency band at one point in the image.

When using trihedral or other calibration device signatures the radiometric corrections put forward in Section III should first be applied to obtain range-invariant DN values. Phase measurements from trihedrals are occasionally subject to fluctuations, so the use of more than one trihedral signature is recommended.

Care must also be taken in deriving calibration parameters from image background properties. Regions of low SNR, containing significant interference or including the nadir return should be avoided. For estimating the cross-talk and HH/VV phase balance, only regions where the HV/HH and HH/VV returns are uncorrelated should be used. These assumptions should hold in areas where the HV backscatter is non-trivial, and the scatters are azimuthally symmetric and predominantly single-bounce.

ACKNOWLEDGEMENTS

The research carried out in this paper was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. The authors would like to thank K. Wang, D. Tomlinson and L. Nguyen for their assistance in analyzing the data and the DC-8 aircraft SAR team for their efforts in building the radar, collecting and correlating the data.

REFERENCES

1. Freeman, A., et al, SIR-C Calibration Workshop Report, JPL, Center for Radar Studies Technical Report 88-003, November 1988.

2. Held, D.N., et al, The NASA/JPL multifrequency, multi-polarization Airborne SAR System, Proc. IGARSS '88, Edinburgh, Scotland, Sept. 1988, pp. 345—350.

3. Ruck, G.T., et al., Radar Cross-Section Handbook, Vol. II, p. 588, publ. Plenum, NY, 1970.

4. Sheen, D. Freeman, A. and Kasischke, E., Phase Calibration of Polarimetric Radar Images, submitted to IEEE GRS Trans., 1989.

5. Klein, J.D., Calibration of Quad-Polarization SAR data using Backscatter Statistics, Proc. IGARSS '89, Vancouver, BC, 1989.

6. van Zyl, J.J., A technique to calibrate polarimetric radar images using only image parameters and trihedral corner reflectors, Proc. IGARSS '89, Vancouver, BC, 1989.

7. Zebker, H.A. and Lou, Y-L., Phase Calibration of Imaging Radar Polarimeter Stokes matrices, Proc. IGARSS '89, Vancouver, BC, 1989.

Table 1 Nominal Image Quality Parameters for JPL A/C SAR High-Resolution Images

	Res (m)	ISLR(dB)	PSLR(dB)	Channel offsets(m)
Range	8.0	-13.0	-16.0	<0.8
Azimuth	3.0	-8.0	-13.0	<0.4

Table 2 Typical Measured Range Impulse Response Parameters for Goldstone Data

Band	Res (m)	ISLR (dB)	PSLR(dB)	Channel offsets (m)
L	11.1 (±0.8)	-9.6 (±0.7)	-21.5 (±1.0)	≤0.7
C	10.7 (±0.8)	-11.6 (±3.2)	-20.9 (±5.1)	≤1.4
P	11.2 (±0.7)	-16.7 (±6.6)	-22.2 (±1.6)	≤1.4

Table 3 Best-Case Azimuth Impulse Responses for Trihedral Data

Band	Res (m)	ISLR(dB)	PSLR(dB)	Channel Offsets(m)
L	4.0 (±0.5)	-9.4 (±6.9)	-22.7 (±2.6)	0.0
C	4.5 (±0.9)	-8.2 (±6.4)	-19.3 (±3.8)	0.0
P	3.9 (±0.6)	-7.2 (±6.7)	-23.2 (±2.8)	0.0

Table 4.Worst-Case Azimuth Impulse Responses for Trihedral Data

Band	Res. (m)	ISLR(DB)	PSLR(dB)	Channel Offsets(m)
L	5.3 (±2.1)	-8.6 (±6.1)	-22.7 (±2.6)	1.8
C	4.1 (±0.5)	+8.4 (±11.1)	-19.3 (±3.4)	1.5
P	16.6 (±2.7)	+8.6 (±11.9)	-	3.3

Table 5 Short-term and Long-term Relative Calibration Precisions in dB Estimated from Trihedrals

Freq.	Pol	Short-Term	Long-Term	RSS
L	HH	±1.4	±1.9	±2.4
	VV	±1.4	±2.0	±2.4
C	HH	±3.8	±6.6	±7.6
	VV	±4.0	±8.4	±9.3
P	HH	±3.5	±4.0	±6.7
	VV	±2.2	±3.3	±4.0

Table 6. Correction Factors, C_F, in dB for Absolute Calibration (from Trihedrals)

Freq.	Pol	Before* 4/4/88	After* 4/4/88
L	HH	52.6 (±2.4)	43.1 (±2.4)
	VV	52.7 (±2.4)	43.3 (±2.4)
C	HH	-17.2 (±7.6)	-14.8 (±7.6)
	VV	-20.6 (±9.3)	-18.9 (±9.3)
P	HH	64.1 (±6.7)	55.4 (±6.7)
	VV	65.8 (±4.0)	56.6 (±4.0)

*Mean value only changes. Variability taken over whole data set for each freq./pol.

Table 7. Polarimetric Calibration Results for L-Band

		Trihedrals		Image Background	
		Short-term	Long-term	Short-term	Long-term
L HH/VV	Ampl	±0.9 dB	-0.3 (±0.3) dB	±0.9 dB	-0.3 (±0.3) dB
	Phase	±31.1°	96.1 (±76.2)°	±13.1°	90.2 (±68.4)°
L HV/VH	Ampl	±2.3 dB	-1.8 (±1.4) dB	±1.8 dB	-0.8 (±1.0) dB
	Phase	±98.9°	64.8 (±18.5)°	±12.2°	65.9 (±15.2)°
L HV/HH	Ampl	±2.6 dB	-15.8 (±2.9) dB	±2.0 dB	-8.4 (±0.5) dB
	Phase	-	-	±30.9°	47.5 (±35.8)°

Table 8. Polarimetric Calibration Results for C-Band

		Trihedrals		Image Background	
		Short-term	Long-term	Short-term	Long-term
C HH/VV	Ampl	±0.8 dB	3.8 (±2.2) dB	±1.4 dB	2.5 (±1.7) dB
	Phase	±27.2°	-49.9 (±89.1)° (A)	±9.6°	-90.9 (±61.2)° (C)
C HV/VH	Ampl	±1.0 dB	0.4 (±1.4) dB	±0.6 dB	0 6 (±1.1) dB
	Phase	±56.0°	-40.1 (±74.4)° (B)	±10.1°	-40.1 (±71.8)° (D)
C HV/HH	Ampl	±4.0 dB	-28.0 (±3.7) dB	±2.0 dB	-10.1 (±1.3) dB
	Phase	-	-	±118.8°	-6.9 (±37.3)°

(A) -34.9 (±13.9)° if only May and June data included.
(B) -96.8 (±17.7)° if only May and June data included.
(C) -39.3 (±9.6)° if only May and June data included.
(D) -101.6 (±10.4)° if only May and June data included.

Table 9. Polarimetric Calibration Results for P-Band

		Trihedrals		Image Background	
		Short-term	Long-term	Short-term	Long-term
P HH/VV	Ampl	±2.4 dB	-0.2 (±2.5) dB	±1.4 dB	-0.5 (±0.6) dB
	Phase	±79.0°	51.1 (±31.8)°	±11.3°	55.9 (±17.7)°
P HV/VH	Ampl	±3.7 dB	0.3 (±3.0) dB	±1.8 dB	-1.4 (±1.9) dB
	Phase	±130.0°	95.4 (±125.3)°	±66.9°	148.2 (±15.3)°
P HV/HH	Ampl	±3.3 dB	-22.0 (±4.8) dB	±2.6 dB	-11.9 (±1.4) dB
	Phase	-	-	±88.0°	-2.5 (±55.5)°

Table 10. Cross-frequency Calibration Results in dB (from Trihedrals)

		Short-term		Long-term	
		Measured	Expected*	Measured	Expected†
L/C	HH	±3.5	±4.1	44.0 (±3.4)	46.5
	VV	±3.7	±4.2	47.1 (±3.4)	49 7
	HH/VV	±1.2	±1.2	-3.2 (±2.4)	-2.6
L/P	HH	±4.2	±3.8	-1.1 (±3.7)	3.4
	VV	±2.5	±2.6	-1.9 (±2.9)	5.1
	HH/VV	±2.7	±2.6	-0.6 (±2.3)	-1.6

* RSS of individual frequency short-term variations from Tables 11-14.

The Saskatoon SAR Calibration Experiment

T.I. Lukowski, R.K. Hawkins, B.Brisco, R. Brown, R. Ford
Canada Centre for Remote Sensing
2464 Sheffield Road
Ottawa, Canada
K1A 0Y7

P. Daleman*
Intera Technologies Ltd.
1525 Carling Avenue, Suite 600
Ottawa, Canada
K1Z 8R9

*Under contract to CCRS

ABSTRACT

During the spring and summer of 1988, a site near Saskatoon including the Kernan Agricultural Research Farm, managed by the University of Saskatchewan, was visited four times by the CCRS airborne SAR. The effort was made to assess relative and absolute radiometric calibration for the SAR facility in conjunction with temporal crop studies conducted in nearby locations.

Several active and passive point reflectors were used as well as large fields for distributed targets with support from airborne and ground-based scatterometers. Of special interest in the experiment was use of recirculating Active Radar Calibrators deployed in high scattering areas but repeating in regions of lower scatter.

Problems encountered and their implications for the relative and absolute calibration potential of the instrument are examined to suggest a practical methodology for continuing research and calibration requirements.

Keywords: SAR, calibration, radar, airborne, remote sensing

1. INTRODUCTION

Relative and absolute calibration of synthetic aperture radar images is an integral part of the evolving science of quantification of these data to meet scientific objectives. The Canada Centre for Remote Sensing has recently commissioned a new airborne synthetic aperture radar (SAR) system at X- and C- bands as described in [1] and [2]. Calibration is an important activity for understanding its performance and in characterizing its limitations as well as suggesting refinements of the system.

In order to develop our expertise, a project in calibration has been initiated at CCRS. As part of this work, during the spring and summer of 1988, a site near Saskatoon including the Kernan Agricultural Research Farm, managed by the University of Saskatchewan was visited by the CCRS airborne SAR. Up to four active and six passive point reflectors were deployed. Large fields were used for distributed targets. Support was also provided by airborne and ground-based scatterometers. Relevant descriptions of these systems are given in [3] and [4] respectively.

This paper will give the aims of the experiment, and describe our introductory analysis of this multi-temporal, multi-sensor data set.

2. EXPERIMENTAL DESIGN

The experimental site is located in Saskatchewan, near the city of Saskatoon in an agricultural region of the Interior Plains with topography showing only slight variation in altitude (varying between 490 and 535 metres above sea level) [5]. The area of interest includes fields sown with crops of wheat, barley, canola, alfalfa, and summer fallow. Corner reflectors and active radar calibrators employed as reference point targets were positioned in fallow fields or on the shore of Patience Lake as shown in the detailed site map (Figure 1). Also indicated here are the sites where observations were made using the ground-based scatterometer along with the flight lines for the airborne scatterometer and point target locations. The flight lines for the airborne SAR were chosen to be east-west. Preliminary SAR data were acquired in May for site selection purposes with imagery then obtained at three stages in the growing season as indicated in Table 1.

The active radar calibrators (ARC) originally described in [6] have been modified to include delay lines. These recirculating ARCs give an infinite series of repeated transmissions of attenuated signals at fixed time intervals [7]. ARCs have been developed with horn or microstrip (patch) antennas. Figure 2 is a photograph of a C-band ARC with complex horn antennas rotated to 45 degrees to receive and transmit both parallel and orthogonal polarizations.

ARCs and corner reflectors (CRs) were deployed with peak antenna gains appropriate to their imaged geometry and in locations where their received signals appear on a background of relatively low backscatter. A particularly favourable situation existed where the ARC was placed to recirculate over the dark background of Patience Lake.

The principal SAR imagery obtained as part of this data set was at C-Band with the Convair-580 flying approximately 6100 metres above ground level. X-Band data were also obtained. The C-Band data from the airborne scatterometer were obtained on separate passes over the test region at altitudes of about 460 metres. The selected scatterometer flight lines were in a north-south direction to place them within the SAR images. Thus, the SAR and airborne scatterometer had the same aspect angle overlapping the fields where ground-based

scatterometer data (at C,L, and Ku bands) were obtained during the overflights by the airborne microwave systems. Table 1 includes the depression angles in the SAR imagery at which the point targets were located and ground-based scatterometer data obtained. Elevation angles used for target pointing are also given here.

3. DATA ANALYSIS

This unique multi-temporal, multi-sensor data set has been found to have a number of features worthy of analysis. As a result, it is being utilized in a number of studies. The relative and absolute calibration analyses are described in the following sections.

3.1 Relative Calibration

The first purpose is to establish the use of relative radiometric calibration for intercomparison of multi-temporal data sets as defined by equation (8) in [8].

In this approach, we attempt to relate two images so that a line at fixed slant range appears as if it were imaged under the same conditions in both images. The multi-temporal imagery of a selected region consists of several scenes and includes both imagery and noise data. Background noise is first subtracted from each image. In order to bring the two images to a common grey scale, the ratio of the power received from a corner reflector

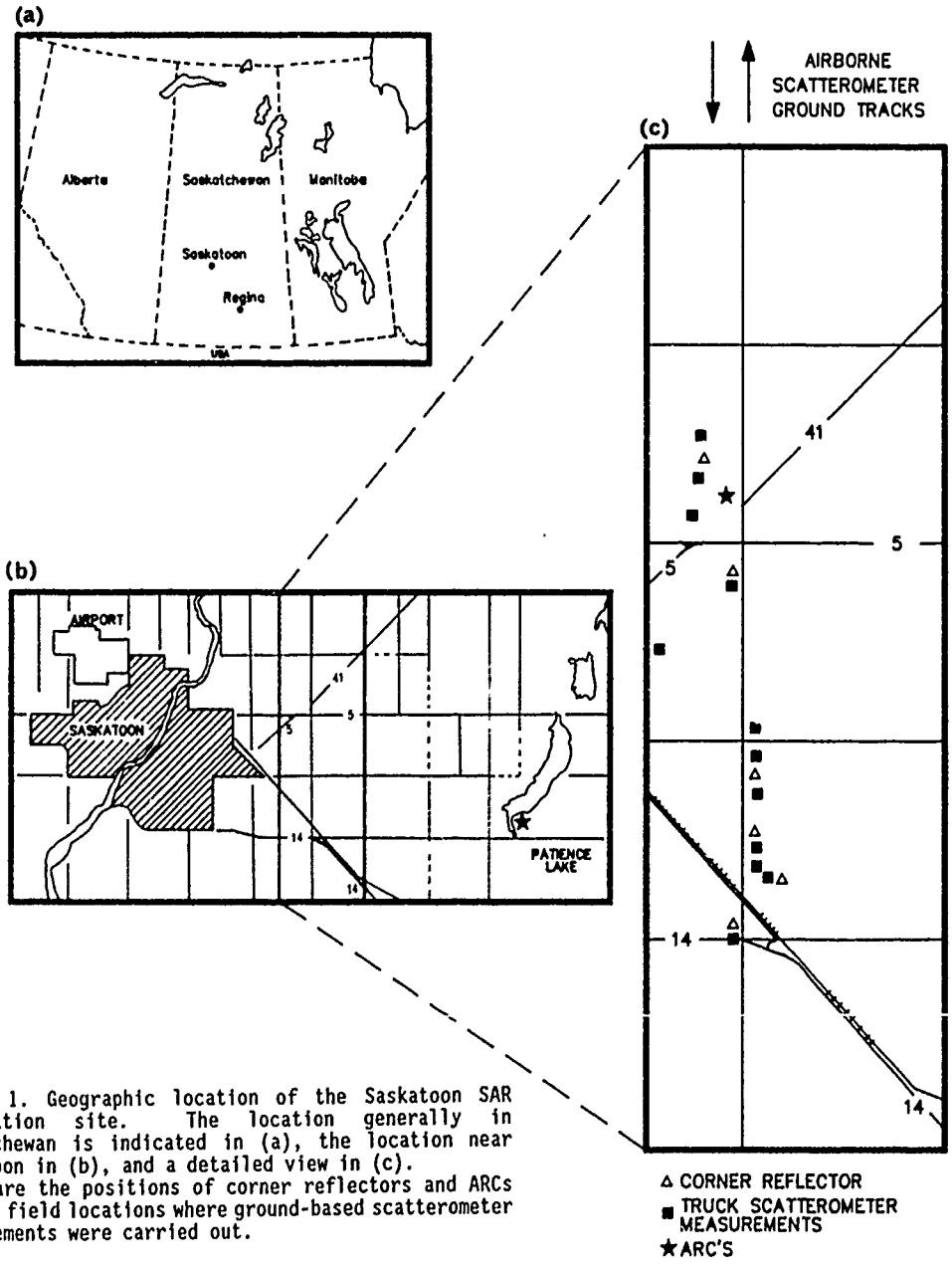


Figure 1. Geographic location of the Saskatoon SAR calibration site. The location generally in Saskatchewan is indicated in (a), the location near Saskatoon in (b), and a detailed view in (c). Shown are the positions of corner reflectors and ARCs and the field locations where ground-based scatterometer measurements were carried out.

TABLE 1:
Saskatoon Experiment SAR Data Collection Summary

The angles summarized are the following:

- θ_s = depression angle (from horizontal) at which the target appears in the swath [degrees]
- θ_b = target elevation angle of boresight as set by ground crew (above horizontal) [degrees]

The azimuth pointing angles of the point targets are not given in the chart because in all cases the point targets have been aimed with an azimuth accuracy of ±0.5° with respect to the actual heading of the aircraft.

BAND/POL is the frequency band and polarization of the SAR (eg. CHH is C-band (5.3 GHz), with horizontal polarization on transmit and receive)

DATE	LINE/ PASS	MODE	BAND /POL	CR θ_s	CR θ_b	ARC's θ_s	ARC's θ_b	ARC's TYPE	GROUND SCATT θ_s	
June 22	1/1	nadir	CHH, XHH	58.0	56.7	51.5	45.0	C-horn	58.7	41.6
				54.0	52.4				54.0	35.6
				49.2	47.5				52.1	32.3
				44.1	41.4				49.2	30.7
				32.6	29.3				44.6	29.3
				28.1	24.7				42.6	27.6
	2/2	narrow	CHH, XHH	—	39.6	43.6	45.0	C-horn	40.0	31.6
				46.1	36.8				44.8	28.8
				42.4	34.2				42.5	27.6
				38.5	30.3				38.0	26.4
				29.1	22.4				37.2	25.0
				25.4	19.2				36.3	
June 24	2/1	narrow	CHH, XHH	42.6	39.6	37.9	41.0	C-horn	43.3	31.7
				39.6	36.8				39.7	27.6
				36.7	34.2				38.7	25.2
				33.3	30.3				36.7	24.3
				25.5	22.4				33.6	23.2
				22.5	19.2				32.5	22.1
	1/2	nadir	CHH, XHH	55.5	56.7	48.7	53.6	C-horn	56.2	39.0
				50.9	52.4				50.9	33.1
				46.5	47.5				49.4	29.9
				41.5	41.4				46.5	28.6
				30.3	29.3				42.0	27.2
				26.0	24.7				40.1	25.6
July 20	1/1	nadir	CHH, XHH	57.7	45.7	51.5	53.6	C-horn	58.3	42.0
				53.5	41.1	29.8	34.0	C-patch	53.5	36.1
				49.2	29.0	29.8	34.0	X-patch	52.1	32.8
				44.4	22.9				49.2	31.3
				31.4	18.0				44.9	29.9
				26.7	13.7				42.9	28.2
	2/2	narrow	CHH, XHH	44.2[1]	39.6	41.1[1]	41.0	C-horn	44.6[1]	33.5[1]
				41.2[1]	36.8	25.1[1]	26.5	C-patch	41.2[1]	29.8[1]
				40.1[1]	34.2	25.1[1]	26.5	X-patch	40.9[1]	27.2[1]
				35.1[1]	30.3				40.1[1]	26.2[1]
				27.4[1]	22.4				35.5[1]	25.1[1]
				24.3[1]	19.2				34.1[1]	24.0[1]
July 21	2/1	narrow	CHH, XHH	47.0	39.6	41.7	45.0	C-horn	47.4	35.1
				43.7	36.8	25.8	26.5	C-patch	43.7	30.6
				40.5	34.2	25.8	26.5	X-patch	42.7	28.1
				36.9	30.3				40.5	27.0
				28.3	22.4				37.3	25.9
				25.0	19.2				35.9	24.6
	1/2	nadir	CHH, XHH	62.7	56.7	64.9	53.6	C-horn	63.4	44.9
				57.8	52.4	31.2	34.0	C-patch	57.8	38.3
				53.1	47.5	31.2	34.0	X-patch	56.2	34.6
				47.6	41.4				53.1	32.9
				34.9	29.3				48.2	31.4
				30.1	24.7				46.2	29.5
Aug 10	1/1	nadir	CHH, XHH	61.0	56.7	57.4	53.6	C-horn	62.4	43.6
				57.0	52.4	30.1	34.0	C-patch	56.7	36.8
				51.9	47.5	30.1	34.0	X-patch	55.3	33.4
				46.6	41.4				51.9	31.8
				33.7	29.3				46.7	30.1
				29.7	25.5				44.7	28.3
				29.1	25.0					
				28.7	24.6					
	1/2	nadir	CHH, XHH	60.6	56.7	53.3	53.6	C-horn	61.4	42.5
				56.0	52.4	29.6	34.0	C-patch	56.0	36.2
				51.0	47.5	29.6	34.0	X-patch	54.3	32.8
				44.7	41.4				51.0	31.0
				33.0	29.3				46.1	29.7
				29.1	25.5				44.0	27.8
				28.7	25.0					
				28.2	24.6					
	2/3	narrow	CHH, XHH	45.8	39.6	40.4	41.0	C-horn	42.0	29.5
				42.3	36.8	24.5	26.5	C-patch	41.7	27.1
				39.4	34.2	24.5	26.5	X-patch	39.4	26.0
				35.3	30.3				36.1	24.6
				27.2	22.4				34.7	23.6
				24.1	19.8				33.9	
				23.8	19.5					
				23.5	19.2					
Aug 11	2/1	narrow	CHH, XHH	41.1	39.6	23.2	24.5	C-horn	41.5	31.3
				38.6	36.8	23.4	26.5	C-patch	38.6	27.5
				35.8	34.2	23.4	26.5	X-patch	37.6	25.4
				32.4	30.3	23.3	25.5	C-horn	35.8	24.2
				25.6	22.4				33.1	23.5
				23.1	19.8				32.0	22.4
				22.8	19.5					
				22.5	19.2					
	1/2	nadir	CHH, XHH	59.5	56.7	29.2	34.0	C-patch	60.4	42.5
				54.9	52.4	29.2	34.0	X-patch	54.9	36.1
				50.2	47.5				53.3	32.7
				44.3	41.4				50.2	31.1
				32.9	29.3				46.7	29.7
				29.0	25.5				43.5	27.8
				28.6	25.0					
				28.2	24.6					

[1] approximate values, data to be reprocessed.

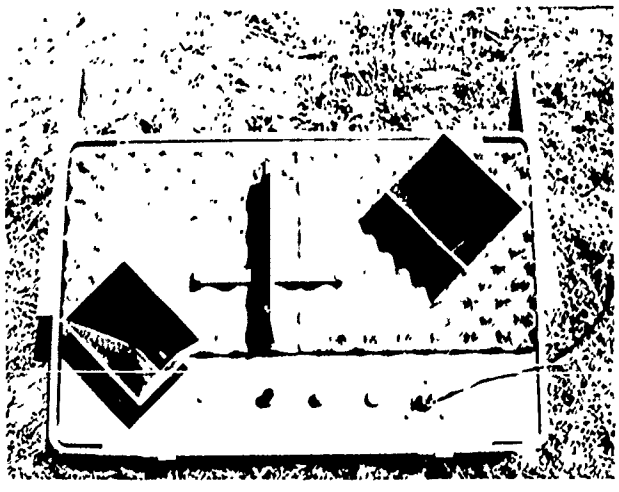


Figure 2. Field Deployment of ARC. The ARC is positioned in the swath, levelled, and oriented in both elevation and azimuth to ensure that it points perpendicular to the aircraft track and at the illuminating antenna.

to the target strength is determined for each image. A correction factor for each line at fixed range is used to multiply one image to obtain the common grey scale for the two. Each of the images also includes a sensitivity time control function (STC) as described in [1] and [8]. The STC is retained as a levelling factor so that the image can be processed on available image processing systems. All images are corrected to use the same STC law as the reference image. By modelling the gain patterns used in acquisition of each of the images, a correction is performed to give the same antenna gain pattern as a function of slant range (compensating for differences in antenna pointing in elevation). As a result, data at the same slant range in the reference and calibrated images can be compared.

3.2 Absolute Calibration

Absolute calibration can also be obtained from the same data set as explained in [8]. It is necessary to remove the effects of the terrain law part of the STC from both images. In this case, we will obtain a variable σ_0 (terrain scattering cross-section) as a function of the slant range. We can then compare the σ_0 within an image and between images, although the σ_0 depends on the incidence angle. Minor modification of the software used for relative calibration will enable absolute calibration.

4. RESULTS AND DISCUSSION

The data which have been processed to date have indicated the feasibility of relatively calibrating SAR imagery from this radar system. Initial processing has indicated that it is possible to account for the system variables. In particular, it is possible to create two images acquired at different times which appear to have been imaged under the same conditions.

In Figure 3, results from the first 10 echos for the ARC on the shore of Patience Lake are seen superimposed on the lake surface image. The first echo appears at the actual ARC location. The others, over the lake surface, exist in an area of very low backscatter, thus having excellent signal-to-clutter ratios. The full line shows the integrated response from the point target according to (5) of [8]. Each succeeding pulse from the ARC is both delayed and attenuated by a predetermined amount (accounting for the decrease in received power with detected replica). The dotted line is the clutter contribution to each recirculation and the dashed line is the response from the ARC when the average clutter is removed. (There is still an interference term [8] which cannot be removed by subtraction.) In this case, for most of the range (up to 7 reflections) the clutter contribution is inconsequential. The slope of the dashed line is related to the accumulated attenuation as the pulse from the radar is successively delayed and attenuated. It serves as a check on the linearity of the radar and a check on the properties of the ARC itself.

Based on these studies, the advantages of using a recirculating ARC have been found to include:

-The background clutter in the vicinity of the target may be eliminated by displacing the return.

-The calibration constant can be determined with much more certainty by combining data from many reflections since the interference term is averaged out.

-The response of the point target is gathered at several predetermined levels. Because the compression gain is high for point targets, saturation is often present in the processed product. Successive delay and attenuation allow the target to be captured at the highest signal to clutter ratio which does not saturate the signal processing system.

Considerable expertise in the deployment of ARCs and corner reflectors has been obtained as a result of this experiment.

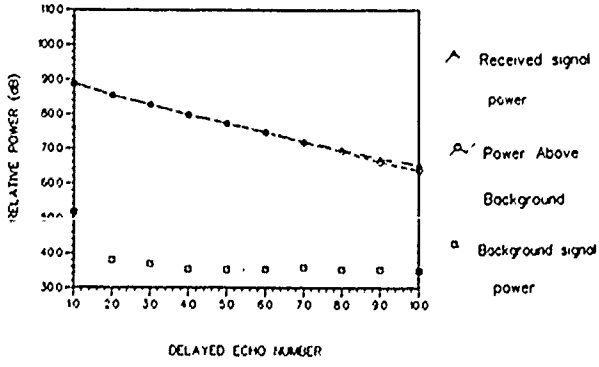


Figure 3. Repeating ARC Response Curve. Signal received from ARC showing delayed transmissions. The initial signal appears superimposed on the lake shore with the remaining echos over Patience Lake.

Experience gained in this experiment is allowing us to identify several problem areas and inherent difficulties in performing calibration of the C/X SAR data. For example, the positioning and characterization of the point targets is critical to the calibration. Differences in the actual flight line flown can result in noticeable shifts in target locations in the range (cross-track) direction. All of these factors will have to be carefully considered in further calibration studies.

5. CONCLUSIONS

Although much work remains to be carried out, this experiment has provided a major step in the calibration of the CCRS SARs. Initial results have shown the utility of the data set for calibration. Experience gained has led to identification of problem areas and improvements in future experiments. Active Radar Calibrators have been shown to be effective tools in this process. Further analysis of this data set for calibration remains to be performed. This experiment is part of an ongoing calibration study at CCRS and acquisition of more data for calibration purposes is planned at this site during the summer of 1989.

6. ACKNOWLEDGEMENTS

It is a pleasure to thank the pilots and crew of the Convair-580 for their skill in carrying out a careful data acquisition and also the University of Saskatchewan for use of their facilities.

7.REFERENCES

1. Livingstone C.E., A.L. Gray, R.K. Hawkins, R.B. Olsen, R.A. Deane and J.G. Halbertsma, "C-Band Airborne Radar - System Description and Test Results" Proc. 11th Canadian Symposium on Remote Sensing, pp. 503-518, 1987.

2. Livingstone C.E., A.L. Gray, R.K. Hawkins, and R.B. Olsen, "CCRS C/X-Airborne Synthetic Aperture Radar: An R and D Tool for the ERS-1 Time Frame" IEEE Aerosp. Electon. Syst. Mag. 3(10):11-20, 1988.

3. Hawkins, R.K., P.J. Farris-Manning, J.R. Gibson, K.P. Singh, "Calibration of the CCRS Airborne Scatterometers", Submitted to IEEE Trans. Antennas Propagat., 1988, 31p.

4. Sofko, G.J., A.G. Wacker, J.A. Koehler, M.J. McKibben, M.R. Hinds, R.J. Brown, and B. Brisco, "Ground Microwave Operations", to be published in Canadian Journal of Remote Sensing, April 1989.

5. The area of interest is located on map sheets 73A and 73B of the National Topographic System, Energy Mines and Resources Canada, 1976.

6. Brunfeldt, D.R. and F.T. Ulaby, "Active Reflector for Radar Calibration" IEEE Trans. Geosci. Rem. Sensing GE-22(2):165-169, 1984.

7. Brunfeldt, D.R. and S. Mukherjee, "SAR Response to Modulated Targets". In IGARSS'89 Remote Sensing: An Economic tool for the '90's, Vancouver, 1989.

8. Hawkins, R.K., T.I. Lukowski, A.L. Gray, C.E. Livingstone, "Calibration for Airborne SAR". In IGARSS'89 Remote Sensing: An Economic Tool for the '90's, Vancouver, 1989.

POSSIBILITY OF THE USE OF A TRANSPONDER AS AN ACTIVE SAR CALIBRATION TARGET

Nobuhiko Kodaira

Remote Sensing Technology Center of Japan
Tokyo, Japan 106

ABSTRACT

The transponder as an active SAR calibration target (Brunfeldt, 1982) is commonly used because of the wide antenna beam width and light weight which result easy to set up the system at any place, especially for the low frequency SAR such as L/P band. However it is necessary to select the site of very low back ground σ° area to obtain high accurate calibration. It is rather difficult to select such preferable test site in Japan where the mountain, cultivated area and city area are prevailed and there is no spacious flat place such as desert, dry up bottom of lake except the surface of the still water.

This paper describes a delay pulse transponder system which will shift an echo pulse, for example, from the shore of the lake to the middle of it on the surface of the water.

Keywords: SAR calibration, delay pulse, transponder.

1. Introduction

It is beyond question that the calibration of remote sensing sensors is absolutely necessary and important.

In case of small parameter SAR, especially single parameter JERS-1 SAR, it must be emphasized that the calibration is necessary for the comparison of a SAR data to another SAR system or differing temporal data. Calibration items will be 1) radiometric calibration such as transmitting power, receiver sensitivity 2) antenna radiation pattern and 3) ambiguity ratio.

Corner-reflector, transponder, and uniformly distributed target with a known σ° values are commonly used for space borne SAR calibration. The first two methods need to use the test site of sufficiently small background σ° such as a dry up bottom of lake, a flat desert, a surface of the still water or a large place of paved with asphalt. The topography of Japan is so complicated that it is very difficult to find out such a preferable test site except the surface of the still water.

The surface of the still water may be an ideal low back ground σ°, on the contrary it is rather difficult to set up those targets (corner reflector and/or transponder) on the surface of the water for the SAR calibration.

In this paper, use of a delay pulse transponder system is proposed. Which is installed on the shore of a lake and the echo pulse will be shifted as if it is installed in the middle of the lake. Thus a very low σ° back ground test site will be effectively realized.

2. Delay pulse transponder system.

Following items can be calibrated by the delay pulse transponder system:

1) Transmitting power of SAR.
2) Receiver sensitivity and linearity of SAR.
3) Range and azimuth antenna radiation pattern of SAR.
4) Ambiguity ratio.

There are two principal advantages of delay pulse transponder system. First, a low back ground test site can be realized without actually set up the instrument on the surface of the water. Second, easy to set up because its small physical size, light weight and broad antenna beam width.

On the other hand, disadvantages are as follows, a) transponder needs a power source and mainte-

nance, b) transponder is more expensive than corner reflector.

The main function of the delay pulse transponder is to receive the SAR signal, amplify, delay for several micro seconds but keep the doppler information and re-transmit the delayed pulse to the satellite.

In this delay pulse system, in order to keep the doppler information, received pulse is converted to I.F. frequency and A/D. Then delay for several micro second in digital form to shift the echo signal position. Fig. 1 shows the block diagram of the delay pulse transponder system.

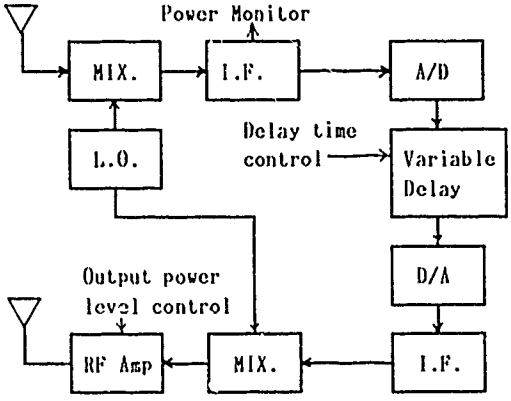


Fig. 1. Block diagram of delay pulse transponder.

The apparent slant range of delayed pulse differs from the original slant range and this cause a difference between the apparent range migration for delayed pulse and original one. Fig. 2 shows the relation between the range shift and an increment of range migration. Suppose the range shift value is kept less than 20 km, the increase of range migration is small enough (one tenth of range resolution) to be ignored as shown in Fig.2.

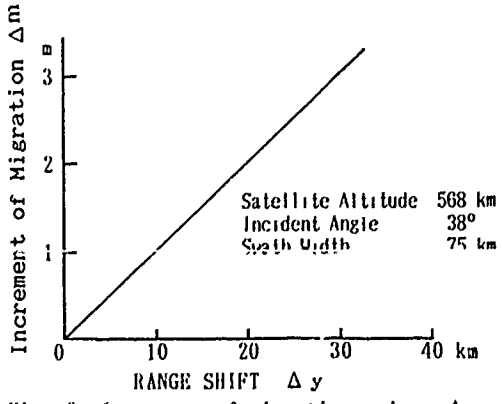


Fig. 2 Increment of migration value Δm as against range shift Δy. Refer to Fig.3 for Δm and Δy.

Power received by transponder Prt through the effective antenna receiving area of At is;

$$Prt = \frac{Pts\ Gs\ At}{4\pi R^2}$$

where Pst, Gs are the transmitting power and antenna gain of the SAR and R is slant range between SAR and transponder.(Fig.4) Transmitting power of transponder Ptt, which provides a minimum receiving power $Psrm$ at the satellite is;

$$Ptt = \frac{Psrm\ 4\pi R^2}{Gt\ As}$$

where Gr is the antenna gain of transponder and is related to At as $At = Gt\ \lambda^2 / 4\pi$. Actual transmitting power of transponder can be reduced by an improvement factor F produced by pulse compression. So that final transmitting power Ptt' is $Ptt' = Ptt / F$.

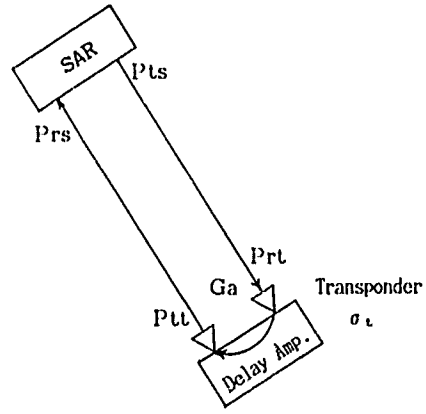


Fig.4. Transponder consists of delay amplifier of total gain Ga including antenna gain $(Gt)^2$

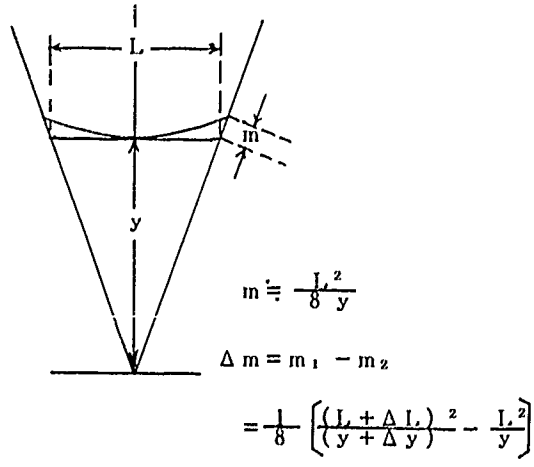


Fig.3. Range migration m at range y.

Accordingly radar cross section of the transponder σ_t is;

$$\sigma_t = (\lambda^2 / 2\pi)\, G_0$$

where G_0 is the overall amplification factor of the transponder, $G_0 = Ptt / Prt$.

Let us assume that we need the maximum measurement uncertainty of ±1 dB. The backscattering-to-calibration radar cross section ratio S becomes;

$$S = -20 \text{ dB}$$

the minimum background terrain scattering coefficient σ^0 is (Ulaby, 1982);

$$\sigma^0 = \sigma_t - A + S$$

Substituting radar parameters of Table 1 to these equations, minimum background σ^0 can be obtained;

$$\sigma^0 \leqq -31.4 \text{ dB}$$

Natural terrain σ^0 ranges between -15dB~-23dB for L_{hh} at 38° of incident angle. (JPL, 1983) σ^0 of less than -30dB can only be found at the surface of the water and desert as shown in Fig.5. (Long, 1975).

Table 1. SAR parameters.

Wave length, λ	23.5 cm (L-band)
Peak transmitting power, Pst	1100 w
Antenna gain, Gs	33.5 dB
Receiver noise figure, NF	4.1 dB
Psrm minimum,	-92 dBm
Receiver dynamic range	22 dB
PRF (center value)	1555 Hz
Off nadir angle, ϕ	35°
Band width of chirp signal	15 MHz
Pulse width, τ	35 μs
Satellite orbit: height, h	568 km (sun synch.)
Slant range (swath center) R	709 km
Speed of satellite, v	6.956 km/s
Ground resolution (3-look)	18m x 18m
Swath width	75 km
σ^0 *	-20.5 dB
Improvement factor, F	27.2 dB

* Noise equivalent back scattering coefficient.

Arrangement of transponder for the measurement of SAR antenna pattern in the range direction, at least 3 measurement points are necessary within swath width. For dynamic range and linearity of the receiver, transponder out put level will be changed together with a small additional delay, we can get successive variable intensity pulse train.

The return pulses of the transponder will appear at the different point in the SAR image but it came back through the same antenna beam position, so that the echo intensities indicate the receiver linearity excluding antenna pattern characteristics.

4. Conclusion

It can be concluded from the above discussion that the delay pulse transponder is useful especially in the complicated terrain area such as Japan Islands where it is very difficult to select a flat test site of low σ^0.

It should be investigated in the next step about the shift in the azimuth direction and another delay pulse system utilizing the time delay of magnetic wave propagation with a small corner reflector placed at a required distance from the transponder.

REFERENCES

1) Brunfeldt,D.R., and F.T.Ulaby, "An active radar calibration target" IGARSS '82, Vol.II, TA-6 5.1-5.5, 1982.

2) JPL, "Shuttle imaging radar-C" JPL Pub. 83-47, July 1983.

3) Long,M.W., "Radar reflectivity of land and sea" p.249, Artech House Inc. 1975.

4) Ulaby,F.T., "Microwave remote sensing" Vol.II, Addison Wesley Pub. Co., 1982.

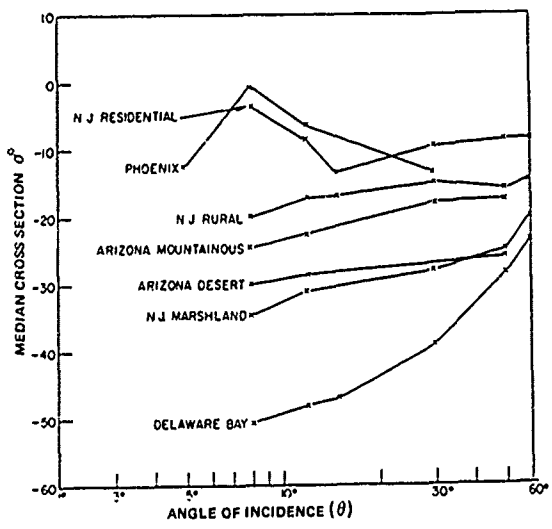


Fig. 5 Median σ^0 for various terrain at Lhh-band.(from Long, 1975)

EFFECT OF CALIBRATION ERRORS ON MEASURING THE RADAR CROSS SECTION OF DIFFUSE TARGETS WITH SAR

Christopher C. Wackerman
Eric S. Kasischke

Radar Science Laboratory
Advanced Concepts Division
Environmental Research Institute of Michigan
Ann Arbor Michigan 48107 USA

1.0 Introduction

One of the desired end-uses of synthetic aperture radar (SAR) data is to utilize the image intensity values as inputs into algorithms which estimate some biophysical or geophysical parameter about the surface being imaged. For instance, there is a concerted on-going research effort to develop connecting models so that spaceborne SAR data can be utilized in a variety of forest ecosystem models (Kasischke and Christensen, 1989).

The precision of the algorithms which utilize SAR image intensity data to generate a specific biophysical or geophysical parameter are going to be directly dependent on the precision of the SAR estimate of image intensity. In turn, the precision of the SAR image intensity measurement is dependent on three distinct terms: (1) radar fading or speckle noise resulting from the coherent nature of the SAR; (2) spatial variance in the physical radar cross section of the scene; (3) calibration errors introduced by system instabilities or uncertainties in knowledge of the SAR collection geometry or the SAR sensor parameters.

Although the errors associated with each of these terms has been researched independently (see, e.g., Goodman, 1976; Kasischke and Fowler, 1989; Ulaby et al., 1982, p. 476-494), no studies have been carried out which explore the combined effects of these terms. This paper attempts to fill that void by presenting a model for calculating the fluctuations expected in a SAR scene as a function of calibration error and spatial distribution parameters. The model is broken down into two parts, a model for the spatial distribution function and a model for the calibration errors, which are discussed in separate sections below. Expressions for the statistics of the resultant SAR image values are then calculated from these models and, in the final section, some conclusions are presented.

2.0 Spatial Distribution Model

Wackerman (1989) presents an overall model for determining the spatial distribution of diffuse scatterers in a SAR scene. In general, this model assumes that a diffuse scene can be broken down into a number of cells which have some radar cross section value that are rough with respect to the radar wavelength of the SAR. The backscattered energy from each cell will then vary with traditional speckle statistics about its radar cross section value. We will let the radar cross section values vary from cell to cell and assume that this spatial variation in mean radar cross section can be described by a density function $f_m(m)$ for each scene. Essentially we are assuming the type of scene (i.e. wheat, oak trees, etc.) can be parameterized by its density function, f_m, that describes the spatial variation in scattering sources. If we also assume the SAR image values are intensity values and that the image could, in general, have been multi-looked, then the speckle fluctuations can be described by a gamma density function, $\Gamma(s;c,1/c)$ where

$$\Gamma(x;c,b) = (x/b)^{c-1}[\exp(-x/b)]/b\Gamma(c) \ , \qquad (1)$$

where c is the shape parameter equal to the number of independent speckle values averaged together to generate the multi-looked data, b is a scale parameter, and $\Gamma(c)$ is the gamma function. If we consider the image values from the entire scene as samples from a single random variable, s, its density function would then be

$$f_s(s) = \int_0^\infty \Gamma(s;c,m/c) f_m(m) dm \ . \qquad (2)$$

Note that the model in Eq. (2) combines the effects of radar fading and natural scene variability into one term, $f_s(s)$.

3.0 Calibration Model

If we let the random variable s in Eq. (2) represent the radar cross section values for each resolution cell within the scene being imaged by the SAR, then the power recorded by the SAR system for each resolution cell, P_r, can be formulated as

$$P_r = G(\theta)s + P_n \qquad (3)$$

where P_n is the additive noise power for the system and $G(\theta)$ represents the overall system gain as a function of the incidence angle, θ, between the SAR and the resolutions cell. It is well established (see, e.g., Ulaby et al., 1982, p. 661-664) that

$$G(\theta) = \frac{H\lambda^3 P_t A^2(\theta)}{2(4\pi)^3 R^3 \sin(\theta)} \qquad (4)$$

where H is the overall SAR receiver and processor gain, λ is the radar wavelength, P_t is the transmitted power, $A(\theta)$ is the antenna gain as a function of

incidence angle and R is the range from the SAR to the scene cell.

The purpose of the calibration procedure is to transform the recorded power, P_r, into a radar cross section value, s, either in an absolute or relative sense (Kasischke, 1989). To do so we need to have an estimate of both the system noise, P_n', and the gain function, $G'(\theta)$. From Eq. (3) we estimate the radar cross section, s', by

$$s' = \frac{P_r - P_n'}{G'(\theta)} = \frac{G(\theta)}{G'(\theta)} s + \frac{(P_n - P_n')}{G'(\theta)} . \quad (5)$$

If we assume that either we have a precise estimate of the system noise, or that $P_r \gg P_n$, then we can ignore the second term in Eq. (5), and

$$s' \simeq \frac{G(\theta)}{G'(\theta)} s . \quad (6)$$

Any SAR system will have inherent instabilities in its hardware which limit the ability to estimate $G'(\theta)$.
These uncertainties can be estimated using either a propagation of error model (see, e.g., Kasischke and Fowler, 1989) or measured by imaging targets whose radar cross sections are known (Kasischke, 1989). In either case, for our combined model, we assume the system gain function has a mean value, $E[G(\theta)]$, and a variance, $var[G(\theta)]$ which can be estimated to some degree of accuracy. From these values we can calculate a coefficient of variation, C_G,

$$C_G^2 = \frac{var[G(\theta)]}{(E[G(\theta)])^2} \quad (7)$$

which will be used below. In general, both the mean and the variance of G might be functions of θ; however, here we assume that $var[G(\theta)]$ is constant over θ; i.e. that the system instabilities are inherent in the hardware and do not depend on θ.

In order to perform calibration we must generate G'; an estimate of G. From the above discussion, it is clear that we need to estimate $E[G(\theta)]$. Let

$$G'(\theta) = B_1 E[G(\theta)] + B_2. \quad (8)$$

The constants B_1 and B_2 represent any errors in the calibration procedure due to uncertainties in knowledge about the SAR collection geometry or the SAR system parameters specified in Eq. (4), for examples, errors in the range value or the antenna pattern gain. We assume a linear model for these errors since in most cases they are represented as a scale factor or a shift in value for any of the system parameters, although there are some non-linear relationships (how altitude errors effect range values, for example). Re-writing Eq. (6) as $s' = \alpha s$ where $\alpha = G/G'$ and using Eq. (8) we have

$$E[\alpha] = \left[B_1 + \frac{B_2}{E[G(\theta)]} \right]^{-1} \quad (9)$$

$$var[\alpha] = \left[\frac{B_1^2}{C_G} + \frac{B_2^2}{var[G(\theta)]} + \frac{B_1 B_2 E[G(\theta)]}{var[G(\theta)]} \right]^{-1} . \quad (10)$$

If we assume that the density function for α is f_α then the density function for s', $f_{s'}$, is

$$f_{s'}(s') = \int\int \Gamma(s';c,m\alpha/c) f_m(m) f_\alpha(\alpha) dm d\alpha. \quad (11)$$

Eq. (11) represents the most general result from our model. Given knowledge about the calibration procedure errors (B_1 and B_2), the system instabilities (E[G], var[G] and f_α), the type of field being imaged (f_m), and the type of speckle multi-looking that was performed (c), one can use Eq. (11) to calculate the expected density function for any diffuse scene in the SAR image. Comparing $f_{s'}$ from different scenes can generate estimates of how distinguishable they are.

4.0 More Specific Results

The results stated above are somewhat overgeneralized to be useful. In order to achieve more useful results, additional simplifying assumptions are required. First we assume that the additive error terms in our error model are small relative to the multiplicative errors, so that B_2 in Eq. (8) is essentially zero. Although the primary motivation for this is admittedly one of computational simplicity in Eqs. (9) and (10), it can also be justified by assuming that the imaging geometry is well known, the calibration procedure is essentially correct and that we can estimate all of the system parameters relatively accurately. If this is so, then the multiplicative factors in Eq. (4) are the key terms in the calibration proces Second we will assume that f_α is a Gaussian density function since the randomness in G comes from the combination of all the instabilities in each part of the radar. Third, simply for the sake of illustration, we will assume that f_m is also Gaussian with mean m_0 and variance v_0. Our first assumption allows us to re-write Eqs. (9) and (10) as

$$E[\alpha] = 1/B_1 \quad (12)$$

$$var[\alpha] = C_G/B_2^2 . \quad (13)$$

With our second and third assumptions substituted into Eq. (11) we can calculate the statistics of s' as

$$E[s'] = m_0/B_1 \quad (14)$$

$$var[s'] = \frac{1}{B_1^2} \left[\frac{c+1}{c} C_G (m_0^2 + v_0) + \frac{c+1}{c} v_0 + \frac{m_0^2}{c} \right]. \quad (15)$$

Eqs. (14) and (15) present give some more intuitive insight into the effect of the different perturbation terms. Eq. (14) shows that E[s'] is strictly dependent on the mean value of the radar cross section of the scene being imaged, m_0, and any scale factor errors in the calibration procedure. Eq. (15) shows that the variance term can be broken down into three components. The third term in the brackets of Eq. (15) represents the spread in SAR image values due to the inherent radar cross section of the scene; this represents the lower bound on the variance. The second term is the added variance due to any spatial variation in the radar cross section values. The first term is the increase in variance due to inherent system instabilities represented by the C_G term. Note that this first term also depends on m_0 and v_0 as is expected since any system variations will enhance any radar cross section variations. Finally, B_1 simply

acts as a scale factor, and thus could be normalized out. Eqs. (14) and (15) thus give a more direct, although less precise, method of predicting the ability to differentiate between two scenes given knowledge of the system instabilities (C_G), the type of scene being imaged (m_o, v_o), the type of multi-looking that was performed (c) and the calibration scale parameters (B_1). In addition, Eq. (11), with the two Gaussian assumptions, provides a method for numerically calculating the actual density functions so that probabilities of false alarm versus probabilities of missed detection could be calculated.

As a final illustration, we can derive a quick method of determining if two fields could be distinguished by a SAR as follows. Assume that we have single looked SAR data so that $c = 1$ and that the calibration procedure is correct so that $B_1 = 1$. Further assume that the spatial variation of radar crossection values for the SAR scenes can be parameterized by letting

$$v_o = p_o m_o \tag{16}$$

so that the p_o parameter specifies the degree of fluctuation for the scene. Eqs. (14) and (15) now become

$$E[s'] = m_o \tag{17}$$

$$var[s'] = m_o^2(2p_o[C_G+1] + 2C_G + 1) . \tag{18}$$

If we have two fields with mean m_0 and m_1 respectively and we assume that they are distinguishable if $m_1 - m_2$ is greater then q times the sum of their standard deviations, then we have the requirement

$$\Delta m > q(m_o\sqrt{2p_o(C_G+1)+2C_G+1} + m_1\sqrt{2p_1(C_G+1)+2C_G+1}) \tag{19}$$

where $\Delta m = m_1 - m_o$. Eq. (19) now provides a simple test for determining how system instabilities and spatial variation in the radar cross section of different field types can effect differentiation.

5.0 Conclusions

We have presented a model for estimating the fluctuations in a SAR image of diffuse scenes based on three error sources: (1) speckle fluctuations; (2) system instabilities and calibration errors; (3) spatial variation in the radar cross section values for the scene. A general model was presented along with a simpler version that gave more intuitive results that described how each term increased in SAR image value fluctuations. A simple test was derived to determine how system instabilities and spatial fluctuations in radar cross section effect differentiation of SAR scenes.

6.0 Acknowledgements

The authors would like to thank Rita Woods and Janice Ancuetil for their help in preparing this manuscript.

7.0 References

Goodman, J., Some Fundamental Properties of Speckle, J. Opt. Soc. Am. Vol. 66, No. 11, pp. 1145 - 1150, Nov. 1976.

Kasischke, E.S., Practical SAR Radiometric Calibration Experiments and Measurements, Proc. IGARSS '89, in press, 1989.

Kasischke, E.S. and N.L. Christensen, Jr. 1989. Connecting Forest Ecosystem and Microwave Backscatter Models, Int. J. Remote Sensing, in press, 1989.

Kasischke, E.S. and G.W. Fowler, A Statistical Approach for Determining Radiometric Precisions and Accuracies in the Calibration of Synthetic Aperture Radar Imagery, IEEE Trans. Geosci. Remote Sens., in press, 1989.

Ulaby, F.T., R.K. Moore and A.K. Fung, Microwave Remote Sensing, Active and Passive-Volume II: Radar Remote Sensing and Surface Scattering and Emission Theory, Addison-Wesley Publishing Co., Reading MA, 607 pp., 1982.

Wackerman, C.C., Calculation of the Spatial Distribution of Scatterers in a Diffuse Scene, Proc. IGARSS '89, in press, 1989.

INITIAL RESULTS OF A SYNTHETIC APERTURE MICROWAVE RADIOMETER

D. M. Le Vine* and M. Kao*, A. Tanner**, C. T. Swift**,
and A. Griffs**

*NASA/Goddard Space Flight Center, Greenbelt, MD 20771
**University of Massachusetts, Amherst, MA 01003

The Goddard Space Flight Center and the University of Massachusetts have been developing a synthetic aperture radiometer for microwave remote sensing of the earth. The radiometer measures the complex correlation of the voltage from pairs of antennas at many different baselines. Each baseline produces a sample point in the Fourier transform of the scene, and a map of the scene itself is obtained after all measurements have been made by inverting the transform.

An aircraft prototype to demonstrate this concept is being developed at L-band to measure soil moisture. The prototype is a hybrid which uses real aperture antennas to obtain resolution along track (stick antennas) and employs aperture synthesis to obtain resolution across track. The prototype was flight tested aboard the NASA P-3 in June, 1988, over the Delmarva peninsula south of NASA's Wallops Flight Facility. The map made during this flight shows the major land/water features and compares very favorably with a Landsat image of the area.
Key words. Microwave Radiometer, Synthetic Aperture

INTRODUCTION

Realizing the full potential of passive microwave remote sensing from space requires putting relatively large antennas in orbit. For example, antenna size is already the factor limiting deployment in space of a radiometer to monitor surface soil moisture. The soil moisture measurement is optimally done at a long wavelength (21 cm) with a spatial resolution of about 10 km (Murphy et al, 1987; Le Vine et al, 1989). This combination of wavelength and resolution requires an antenna on the order of 20 m x 20 m even in low earth orbit (e.g. as planned for the Eos system; Butler et al, 1987) and a very much larger antenna if the monitoring is to be done from geostationary orbit. Furthermore, soil moisture maps must be updated regularly (every 1 - 3 days) which means that this large structure must scan a very wide swath, even in low earth orbit.

Aperture synthesis may offer a way to achieve these requirements. Aperture synthesis is a technique in which the complex correlation of the output voltage from pairs of antennas is measured at many different baselines. Each baseline produces a sample point in the two dimensional Fourier transform of the scene, and a map of the scene is obtained after all measurements have been made by inverting the transform. Only one antenna pair is needed for each independent baseline and small antennas can be employed because the resolution is determined by how well the Fourier space has been sampled, not by the resolution of the actual antennas employed in the measurement. Consequently, a substantial reduction can be obtained in the antenna collecting area required in space as compared to a conventional imaging radiometer of the same resolution.

Aperture synthesis has been successfully employed in radio astronomy to obtain very high resolution maps of radio sources in what is called "earth rotation synthesis" (Swenson and Mathur, 1968; Brouw, 1975; Thompson, Moran and Swenson, 1986). The Very Large Array in Socorro, New Mexico is an example (Napier et al, 1983). It can also be done from space to look down on the earth for purposes of remote sensing (e.g. Le Vine, et al, 1989; Le Vine and Good, 1983; Schanda, 1979).. The principle advantages of aperture synthesis are its potential to form a high resolution map with small antennas and the fact that no mechanical scanning is necessary. However, because the technique also reduces the physical collecting area, a major issue to be addressed in applying aperture synthesis to remote sensing from space is achieving the required radiometric sensitivity, ΔT.

The sensitivity at each pixel in the image formed with aperture synthesis can be written in a number of equivalent forms (e.g. Napier et al, 1983; Ruf et al, 1988) one of which is the following (Le Vine, 1989):

$$\Delta T = \frac{T_{sys}}{\sqrt{B\tau}} \cdot \frac{A_{syn}}{n A_e} \qquad (1)$$

where T_{sys}, B and τ are the system noise temperature, the bandwidth and the integration time, respectively of the correlation receiver (i.e. assumed to be identical for all antenna pairs); and where A_{syn} and A_e are the equivalent areas of the synthesized antenna and the actual antennas used in the measurements; and where n is the square root of the number of independent antenna baselines in the measurement. Under a reasonable approximation (e.g. Napier et al, 1983; Le Vine, 1989), n the number of antennas used in the measurements, in which case, Equation 1 is just the sensitivity of a conventional, total power radiometer (i.e. $T_{sys}/\sqrt{B\tau}$) times the ratio of the effective area of the synthesized beam, A_{syn}, to the actual collecting area, $n A_e$, employed in the measurement.

Achieving adequate sensitivity is an especially critical issue for microwave remote sensors in low

earth orbit because the motion of the platform (about 7 km/s) limits the integration time available for imaging a particular scene. The problem is especially pernicious because high spatial resolution means larger A_{syn} and also less time spent by the spacecraft over the scene. In principle, the sensitivity can be improved by increasing the integration time-bandwidth product. However, for sensors in space, platform motion restricts the integration time, and requirements on the field-of-view and the problem of RFI restrict the bandwidth (Thompson and D'Addario, 1982). Sensitivity can also be improved by increasing the product, $n A_e$, for example by making many measurements at the same time; however, this generally means more hardware in orbit.

In order to demonstrate that aperture synthesis can be employed for microwave remote sensing from a moving platform and to study some of the critical issues such as mentioned above. we have assembled a prototype instrument for deployment on an aircraft. This instrument is a hybrid which uses real antennas to achieve resolution in one dimension and uses aperture synthesis to achieve resolution in the other dimension. This is a convenient configuration for testing the concept because aperture synthesis is only done in one dimension The configuration is also relevant because it is one which could be implemented in low earth orbit to do the soil moisture measurement (Le Vine, et al, 1989; Murphy et al, 1987).

INSTRUMENT DESCRIPTION

The concept is illustrated in Figure 1. The sensor consists of an array of stick antennas oriented with their axis in the direction of motion. The stick antennas produce a fan beam which is narrow in the direction of motion (along-track) and broad in the cross-track dimension. The fan beam is swept along the surface, as the sensor moves, and provides resolution along-track. Resolution across-track is achieved using aperture synthesis with pairs of the stick antennas.

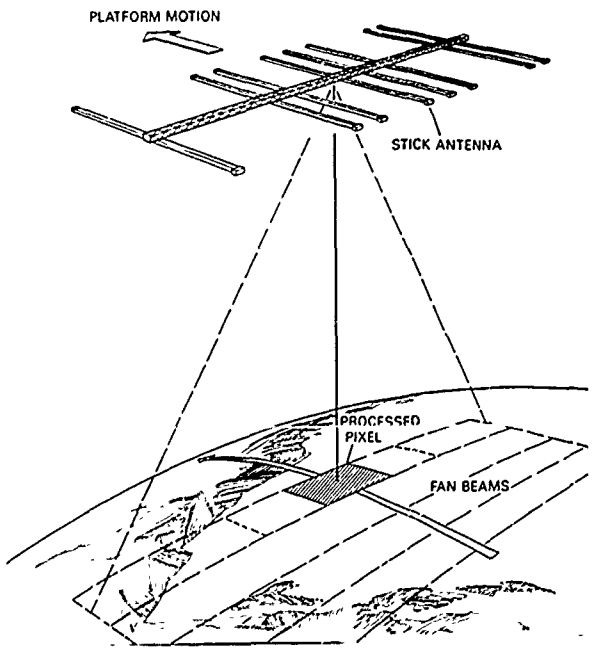


Figure 1: The instrument concept

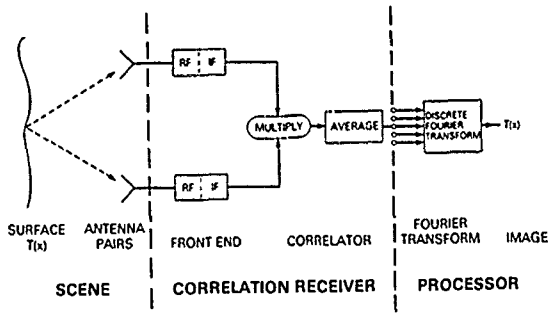


Figure 2: Block diagram showing the signal processing

All of the sticks are aligned to view the same cigar-shaped footprint and all the correlation-pairs are measured simultaneously. When all the measurements are processed an image is obtained of the fan-beam footprint with a resolution determined by the maximum baseline between antennas. By using a maximum baseline comparable to the length of the sticks, one obtains equivalent resolution in both dimensions. Since only those pairs which result in independent baselines are needed, only a sparse array of sticks is necessary. Furthermore, no scanning is necessary as the Fourier transform yields an image of the entire fan beam.

Some of the choices we made in deciding how to implement this concept were determined by the resources which were available at the time. In particular, by adapting an existing array of L-band dipoles to our purpose, we were able to save considerably in time and money compared to the costs of designing and fabricating new stick antennas. The existing array consisted of 64 crossed dipoles arranged in a rectangular array of eight elements on a side and spaced 1/2 wavelength apart. By combining the dipoles into rows, an array of eight parallel stick antennas was created. (Each stick is a linear array of eight dipoles.) Each of the resulting sticks has a beam approximately +/- 45 degrees wide in the cross track dimension and about +/- 6 degrees in the along track dimension. With these antennas it is possible to measure baselines with integer multiples of $\lambda/2$ from zero to 7. The measurement at zero spacing provides the constant (dc) in the Fourier series. Although only a minimal amount of thinning is achieved with this particular array, the number of sticks needed decreases dramatically for a larger array. For example, a 20 m x 20 m array suitable for measuring soil moisture from space would be more than 80% empty.

Figure 2 is a block diagram of the signal processing required in this system. In the first step (receiver front end) the signal is amplified, filtered (center frequency of 1413.5 MHz) and mixed down to a convenient IF. Two IFs were used. In particular, the signal from each antenna was split and mixed to 113.5 MHz in one path and 143.5 MHz in the other. Then, the multiplication was done by summing signals with different IF and passing them through a square law detector and filter. This produced a signal centered at 30 MHz which was coherently mixed with a reference oscillator at 30 MHz to produce the inphase product and with the oscillator shifted 90 degrees to obtain the quadrature product.

The measurement for zero spacing must be done slightly differently because in this case the two arms

to be multiplied are identical and the system noise is correlated. A null feed back radiometer was used for this measurement. This is a total power radiometer in which a portion of the output is compared with a reference source (noise diode) and the gain in a feedback loop is adjusted until the output is zero. The gain provides a measure of the actual input noise. This scheme was adopted because it permitted using the same front end for both the correlation channels and the total power channel.

This receiver yields 15 independent measurements: I and Q output for each of the 7 baselines plus the total power out at the zero spacing. These data were averaged for about 1/4 second using a low pass filter at the correlator output, the averages were then digitized and stored. In addition, a reference signal at 1413.5 MHz is periodically switched in at the antenna terminals. This provides a reference to keep track of changes in phase and gain in each of the channels of the correlator. Also, thermistors were placed throughout the system to monitor temperature.

DATA REDUCTION

Under ideal conditions (e.g. neglecting fringe washing and other potential problems: Thompson and D'Addario, 1982) the correlator output can be written as a Fourier transform of the scene, T(x) (Thompson, Moran & Swenson, 1986; Swenson and Mather, 1983). In one dimension the result can be written in terms of the direction cosine, x, of the vector from the antenna to the scene, as follows (Le Vine, 1989):

V(k) = CORRELATOR OUTPUT

$$= C(k) \int_{-1}^{+1} P(x)T(x)\exp[-j(2\pi d/\lambda)x]dx \qquad (2)$$

where k is an integer denoting the baseline (k = 0, 1, 2 ... 7 in the case at hand), C(k) is a (complex) constant which depends on the gain and phase of the relevant paths in the receiver, and P(x) is the normalized antenna pattern (Kraus, 1966) of each "stick" divided by cos(x). The objective in aperture synthesis is to obtain enough measurements, V(k), to invert this equation and obtain a map of T(x). This can be done analytically using either a discrete Fourier transform on the set {V(k)} or it can be done numerically by "discretizing" the integral in Equation 2 and then numerically doing a matrix inversion to obtain the brightness. Both approaches are being developed for obtaining T(x), but most success has been achieved to date using the "discretized" integral The approach has been to seek and optimum inversion using test scenes to optimize the basis matrix and to determine unknown parameters of the system (i.e. offsets and gains such as the C(k) in Equation 1). The "data" used to obtain this optimization included measurements of the antenna pattern, the response of each channel to a point source and several reference scenes such as an image of the horizon. At this time, the inversion is yielding an image; however, work is still being pursued to refine the inversion and to fully understand the sources of error.

FLIGHT TEST

The instrument was assembled in January, 1988, and tested during a brief check-out mission aboard the NASA P-3 the following February. The hardware performed well during this test; however, significant problems were encountered with radio frequency interference (RFI). A full scale test of the instrument took place several months later (June, 1988) after changes had been completed to reduce the effects of RFI. The purpose of this flight was to collect data to determine if an image could actually be formed. It was decided to perform this test over the Delmarva peninsula because the multitude of land/water boundaries in this area provided a region of high radiometric contrast which the instrument should clearly detect. Figure 3 is a map showing the region imaged. The flight originated at Goddard's Wallops Flight Facility (WFF) and consisted of a series of East-West passes from the Chesapeake Bay across the peninsula to the Atlantic ocean and back. The region of the peninsula which was mapped is indicated by the dashed box. Figure 4a is a Landsat image made of this same area in November (1982). It is included here as a reference to identify physical features of the peninsula. In particular, notice the islands and sandbars which mark the extreme East coast of the peninsula and the marsh that exists between these islands and the mainland. One expects that in this area the L-band radiometer which is sensitive to moisture in the soil may give a different image than the Landsat (infrared) image.

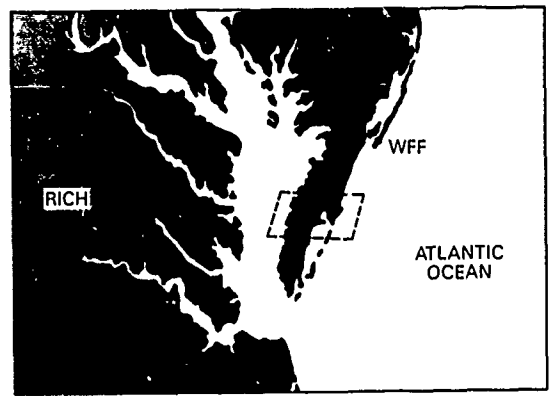


Figure 3: The area mapped during the June, 1988 flight.

Figure 4b is the image obtained with the ESTAR radiometer during the June flight. Notice, the rivers in both images on the West coast and the sandbars on the extreme Eastern edge at the Atlantic coast. The color code used in the ESTAR image has blue at the coldest and progresses through green, red, yellow and white as the temperature increases. Notice the temperature gradients in the coastal areas and the relatively uniformly warm temperatures over the central land areas. This flight was at about 1670 m (5,000 ft) which gives a resolution cell on the ground of about 350 m. The resolution degrades at the extremes of the fan beam, but the edges have been trimmed to provide an image with a reasonably uniform resolution.

Although, the sky was mostly clear during this flight a rather large cumulus cloud lay over the area in the upper left of the ESTAR image. This cloud was dense enough to prevent optical photography of the ground (and give a rough ride in the airplane) but clearly did not prevent microwave imaging of the surface.

CONCLUSIONS:

The long term objective of this research is to develop aperture synthesis as a tool for microwave remote sensing from space. The initial step has been to demonstrate the technique by making soil moisture

measurements from an aircraft. A beginning has been made and our success as documented in this paper indicates the potential of the technique.

However, to be a scientifically useful tool, the instrument must be calibrated. That is, an actual temperature needs to be assigned to the colors in Figure 4b. We are currently working to establish this calibration. This work includes flights over "known" scenes and flights with a reference instrument (such as a thermal infrared sensor) and refinements of our image inversion. We are also changing the hardware (the antennas and filters) to facilitate the calibration and to help cope with a persistent problem with RFI.

ACKNOWLEDGMENT:

We wish to acknowledge the contributions of C. Ruf to the initial design of the electronics and to T. Hiett for his work in various stages of electronics and algorithm development. This work supported from grants from NASA (NAS5-29385) and USDA Agricultural Research Service (USDA-#58-32U4-7-064.

REFERENCES:

Brouw, W. N., "Aperture Synthesis", in Methods in Computational Physics, Vol 14, B. Adler, Ed , pp 131-175, Academic Press, 1975.

Butler, et al., "From Pattern to Process: The Strategy of the Earth Observing System", Eos Science Steering Committee Report, Vol II. 1988.

Kraus, J. D., Radio Astronomy, McGraw-Hill Book Company, 1966.

Le Vine, D. M. and J. Good, "Aperture Synthesis for Microwave Radiometers in Space", NASA TM-85033 (Avial NTIS #83N-36539), 1983.

Le Vine, D. M., T. T. Wilheit, R. Murphy and C. Swift, "A Multifrequency Microwave Radiometer of the Future", IEEE Trans. on Geosci & Remote Sensing, Vol. 27 (2), pp. 193-199, 1989.

Le Vine, D. M., "The Sensitivity of Correlation Radiometers for Remote Sensing from Space", NASA TM-XXXXX, 1989.

Murphy R. et al., "Earth Observing Systems Report". Vol. IIe, report of the "High-Resolution Multifrequency Microwave Radiometer Instrument (HMMR) panel, NASA tech report, Washington, D.C., 1987.

Napier, P. J., A. R. Thompson and R. D. Ekers, "The Very Large Array: Design and Performance of a Modern Synthesis Radio Telescope", Proceedings IEEE, Vol. 71 (11), pp. 1295-1320, 1983.

Ruf, C. S., C. T. Swift, A. B. Tanner and D. M. Le Vine, "Interferometric Synthetic Aperture Microwave Radiometry for the Remote Sensing of the Earth", IEEE Trans. Geosci & Remote Sensing, Vol. 26 (5), pp. 597-611, 1988.

Schanda, E., "High Ground Resolution in Passive Microwave Earth Observation from Space by Multiple Wavelength Aperture Synthesis", Congress, Internat. Astronautical Federation, Anaheim, CA, October 1976.

Swenson, G. W. and N. C. Mathur, "The Interferometer in Radio Astronomy", Proceedings IEEE, Vol 56 (12), pp. 2114-2130, 1968.

Thompson, A. R. and L. R. D'Addario, "Frequency Response of a Synthesis Array: Performance Limitations and Design Tolerances", Radio Science, Vol. 17 (2), pp. 257-370, 1982.

Thompson, A. R., J. M. Moran and G. W. Swenson, "Interferometry and Synthesis in Radio Astronomy", J. Wiley & Sons, New York, 1988.

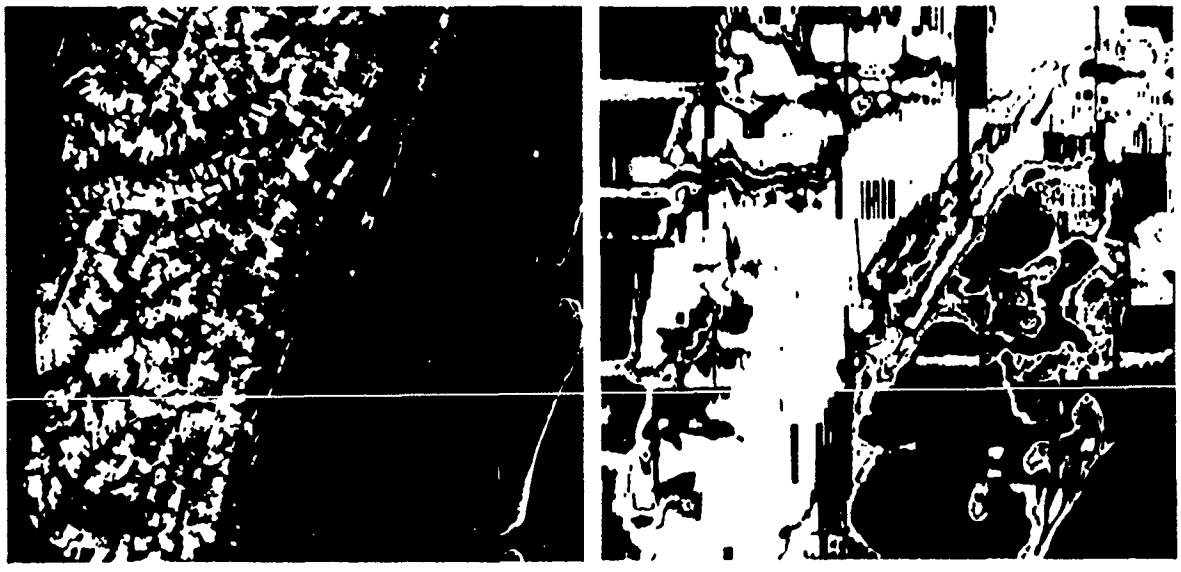


Fig. 4a

Fig. 4b

Figure 4: LANDSAT image (left) and ESTAR image (right) of a portion of the Delmarva peninsula. The ESTAR image is from June, 1988.

Beam compressed microwave scanning radiometer

Kohei Arai* and Tsutomu Suzuki**

* Earth Observation Center, National Space Development Agency of Japan
1401 Ohashi, Hatoyama, Saitama 350-03 Japan
** Dept. of Applied Electronics, The University of Electro-Communications
1-5-1 Chouhugaoka, Chouhu, Tokyo 182 Japan

Abstract

The beam compressed microwave radiometer taking into account the capability of installation is proposed for an improvement of spatial resolution. In principle, it is possible to improve the spatial resolution by some integer values. For instance, by adding one reference antenna to the main antenna, the equivalent -3dB beam width is reduced by 2.

Keywords

Microwave Radiometer, Beam compression techniques, Narrow beam, Capability of installation

1. Introduction

For many years, the microwave radiometers onboard spacecrafts such as SMMR, NEMS, SAMIR, SCAMS, ESMR, MSR of MOS-1, etc have provided usefull data for the investigation on salinity of the ocean, soil moisture, ocean wind, water vapor in the atmosphere, cloud liquid content, and so on (Ref.1 - 8). On the other hand, Users' requirements have been gotten high in terms of the spatial resolution and lower frequency channels for the estimation of soil moisture and salinity of the ocean. In order for it, ESTR of HMMR has been proposed (Ref.9). Since spatial resolution, or -3 dB beam width of microwave radiometer depends on the diameter of the antenna, and wave length, high resolution and low frequency microwave radiometer requires a large antenna reflecter. For example, ESTR composed with the 17 m x 17 m antenna.

Meanwhile the beam compression techniques are getting more popular for high resolution of radar system such as Mills Cross Type Antenna, Multiplicative or Product Type Antenna, Frequency Multiplier Type Antenna, Self Focusing Type Antenna, Temporal and Spatial Modulating Type Antenna, etc (Ref. 10, 11). By applying the techniques to the passive microwave radiometer, the beam compressed radiometer is proposed taking into account capability of installation.

At first, principle of the beam compression technique will be described followed by the results from a configulation trade off study.

2. Priciple of the beam compression techniques

Figure 1 shows a principle of the beam compression techniques, in particular Multiplicative type of Antenna system. In this figure, antennas A and B are Reference and Main antennas, respectively. Distance between centers of both antennas, D should be equal to the dimension of the main antenna.

Emitted microwave energy is acquired by both main and reference antennas simultaneously. In this case, the equivalent - 3 dB beam width at the output of the multiplication becomes about half of the - 3 dB beam width of main antenna Wm as is indicated in Figure 2. Therefore the dimension of the main antenna is also about half of the required antenna dimension without consideration of the beam compression techniques. In general, the beam width of the reference antenna does not matter for the beam compression factor.

As is hown in Figure 2, the beam compression techniques allow us to realize not only high resolution microwave radiometer but also supression of the side-lobe level results in an improvement of S/N ratio.

3. Installation

In order to realize a high resolution microwave radiometer, generally, Casegren type of antenna system is used. By applying the beam compression techniques, the reference antenna (horn type and broad beam of antenna) will be installed beside the reflecter of the main antenna. For instance, Figure 3 shows a configulation of the beam compressed microwave radiometer with mechanical scanning. On the other hand, Figure 4 also shows that

of electric scanning type of microwave radiometer. In both cases, however, delay time and S/N degradation due to the distance between main and reference antennas should be considered. To avoid such a situation, the following configulation was proposed. Figure 5 shows a configulation of the beam compressed microwave radiometer taking into account a capability of installation, with electric scanning. In this case, the delay line corresponding to the distance d indicating in the figure, should be inserted between the reference antenna and multiplication electronics as is shown in Figure 6, to adjust the phases between main and reference antennas.

Although the aforementioned example shows the possibility on the beam compression with compression factor of 2, by adding another antenna systems, arbitrely compression factors are realized.

4. Concluding Remarks

In order to realize high resolution and low frequency microwave radiometer, a large antenna system will be required. A large antenna would generate a large moment torque results in an influence to spacecraft attitude and orbit changes. Even if electric scanning type of antenna is adopted, a large reflecter will be required. To overcome this stuation, the beam compressed high resolution microwave radiometer taking into account capability of installation was proposed. Further investigation is required for realization of this radiometer. The results will be reported in another opportunity.

References

(1) Njoku, E.G., 1982, Passive Microwave Remote Sensing of the Earth from Space - A Review, Proc. of th IEEE, vol.70, no.7, p.728 - 750.

(2) Grody, N.C. et al., 1980, Atmospheric Water Content over th Tropical Pacific Derived from the NIMBUS - 6 Scanning Microwave Spectrometer, J. of Applied Meteorology, vol. 19, p.986 - 996.

(3) Staelin, D.H. et al, 1976, Remote Sensing of Atmospheric Water Vapor and Liquid Water with the NIMBUS - 5 Microwave Spectrometer, ibid, vol. 15, p.1204 - 1214.

(4) Prabhakara, C. et al, 1982, Remote Sensing of Precipitable Water over the Oceans from NIMBUS - 7 Microwave Measurements, ibid, vol.21, p.59 - 68.

(5) Liou, K.N. et al, 1979, Atmospheric Licquid Water Content Derived from Parameterization of NIMBUS - 6 Scanning Microwave Spectrometer Data, ibid, vol.18, p.99 - 103.

(6) Gohil, B.S.,et al, 1982, Remote sensing of atmospheric water content from Bhaskara SAMIR data,Int. J. Remote Sensing, vol.3, no.3, p.235 - 241.

(7) Arai, K.. et al, 1984, Some results on the field experiments in Marine Observation Satellite - 1, MOS - 1 Verification Program, the 14th ISTS, m-1-1.

(8) Arai, K.,1988, A preliminary assessment of radiometric accuracies for MOS-1 sensors, Int. J. Remote Sensing, vol.9, no.1, p.5 - 21.

(9) Murphy, R. et al, 1987, High-resolution mutifrequency microwave radiometer, NASA Earth Observing System Report vol.IIe.

(10) Kraus, J.D., Radio Astronomy, McGraw- Hill book Co., p.152 - 227.

(11) Blommendaal, R., 1964, The Radio and Electronic Engineer.

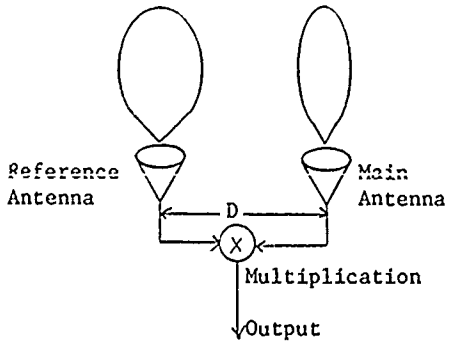


Figure 1 Principle of the beam compression technique of Multiplicative Type of Antenna

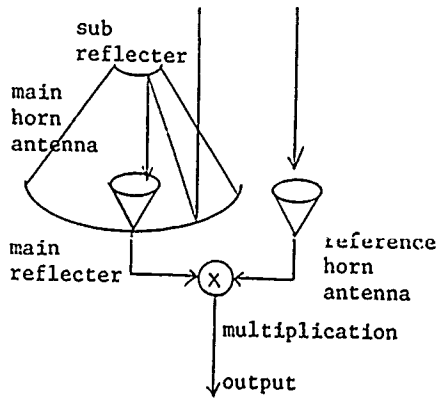
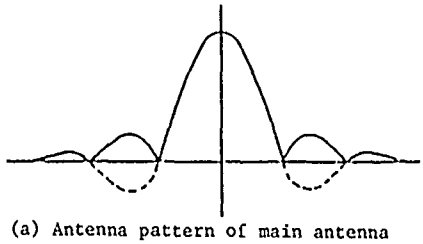
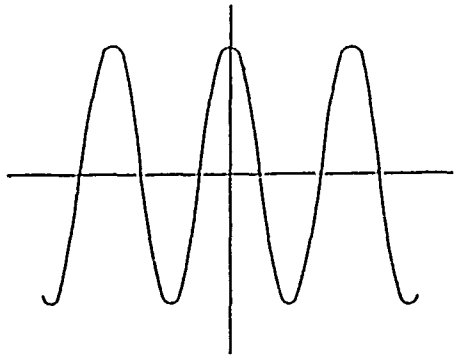


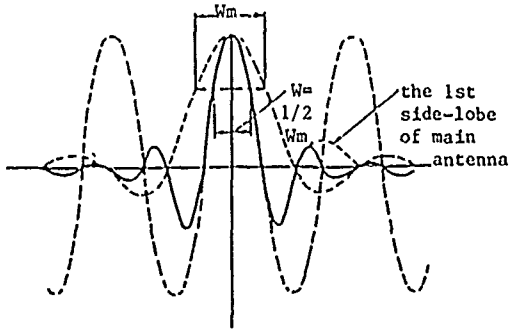
Figure 3 A configulation of beam compressed microwave radiometer with mechanical scan



(a) Antenna pattern of main antenna



(b) Antenna pattern of reference antenna



(c) Associated antenna pattern by multiplying the patterns of both main and reference antennas

Figure 2 Beam compression technique

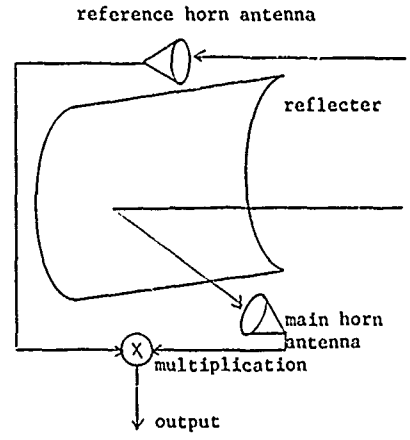


Figure 4 A configulation of beam compressed electric scanning type of microwave radiometer

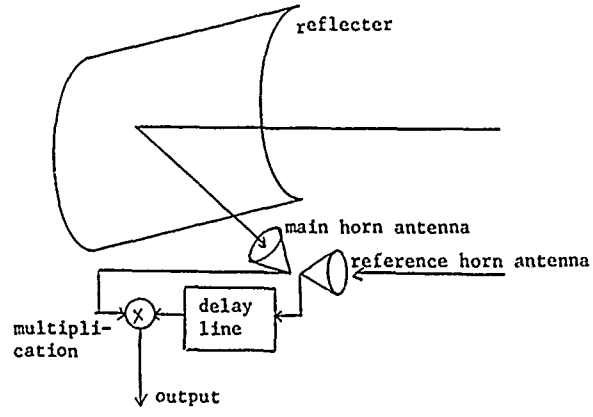


Figure 5 A configulation of beam compressed electric scanning type of microwave radiometer with consideration on capability of installation

TOPEX MICROWAVE RADIOMETER SYSTEM CALIBRATION; REFINING THE SMMR HERITAGE

Christopher S. Ruf, Michael A. Janssen, and Stephen J. Keihm

Jet Propulsion Laboratory
California Institute of Technology
Mail Stop 168-327
4800 Oak Grove Drive
Pasadena, CA 91109
tel. (818) 354-7257
FAX 818-354-3437

ABSTRACT

A modified version of the Scanning Multi-channel Microwave Radiometer (SMMR) will be used for wet tropospheric path delay corrections to the TOPEX/POSEIDON radar altimeter measurements. A number of the sources of calibration problems encountered by SMMR on board the SeaSat and Nimbus-7 platforms have been identified, and appropriate corrections have been attempted. Calibration hardware corrections include a more representative modeling of the microwave losses and reflections and a reduction in the thermal gradients expected across this hardware, through the use of radomes and sun shades and by the choice of pertinent orbit parameters. Antenna calibration corrections include a post launch fine tuning of the antenna pattern correction algorithm to accommodate small errors in the pre-launch antenna pattern measurements. This is accomplished by overpasses of ground based, upward looking water vapor radiometers. An absolute calibration accuracy of $\leq$ 1.0 K is anticipated.

KEY WORDS: SMMR, TOPEX/POSEIDON, Microwave radiometer, Absolute calibration

BACKGROUND

The SMMR has flown successfully on the SeaSat and Nimbus 7 satellite instrument platforms. The radiometer consisted of channels at 6.6, 10.7, 18.0, 21.0 and 37.0 GHz with either vertical or horizontal polarization at each frequency. SMMR produced images of the radiometric brightness temperature distribution over the earth with a mechanically scanning antenna. The instrument provided a wide variety of geoscience information about the atmosphere, sea surface, and sea ice (Njoku, et al, 1980a).

The upcoming TOPEX/POSEIDON ocean topography experiment will determine the sea surface height along the ground track of the satellite by precision radar altimetry. Corrections for the path delay of the radar signal due to atmospheric water vapor play a dominant role in the overall error budget of the altimeter. Atmospheric water vapor is monitored by the TOPEX Microwave Radiometer (TMR), a coincident downward looking radiometer with channels at 18.0, 21.0, and 37.0 GHz. The TMR is presently being designed and assembled at the Jet Propulsion Laboratory. It is based in large part on the SMMR instrument and much of the hardware is inherited from that system.

SMMR SYSTEM CALIBRATION HERITAGE

SMMR system calibration for both SeaSat and Nimbus followed a two step procedure. The antenna temperature referenced to the input of the antenna was derived from the measurements by inverting a model for the radiative transfer of the signal through the front-end hardware (Swanson and Riley, 1980). This step was especially critical given the SMMR design because the on-board system calibration reference brightness temperatures (an ambient load and a cold sky horn) were viewed by the radiometer through different front-end hardware. The second step involved a partial deconvolution of the SMMR antenna pattern to estimate the brightness temperature of the earth near the antenna boresight from the integrated antenna temperature over the complete pattern (Njoku, et al, 1980b). This process, too, was critical because of the relatively low antenna beam efficiency.

SMMR SYSTEM CALIBRATION PERFORMANCE

The SMMR data products have a long, well-documented history of significant calibration problems (SeaSat-JASIN, 1980), (Milman and Wilheit, 1985). The antenna temperature calibration suffered from temperature gradients along the front-end hardware due to limitations in the radiative transfer model. The accuracy of the brightness temperature calibration was limited by the low beam efficiency. The cold sky horn measurements were also contaminated by the sun. The antenna pattern correction algorithm required pre-flight antenna range pattern measurements with unrealistically high accuracy. These patterns were needed to determine antenna beam efficiencies over different

earth and cold sky viewing solid angles in order to estimate that portion of the measured antenna temperature which was due to the brightness temperature of the earth lying in the antenna main beam.

TMR SYSTEM CALIBRATION REFINEMENTS

A number of improvements are being incorporated into the TMR design to facilitate the system calibration. The multi-frequency feed horn (MFFH) in the offset parabolic reflector antenna is being modified to reduce the edge illumination on the reflector and to improve the impedance match of the feed over the three TMR frequencies. The change in aperture illumination should increase the antenna beam efficiency, and the reduction in mismatch reflections should improve the radiative transfer model and reduce the sensitivity to temperature gradients. Preflight thermal/vacuum tests will be modified from those used for SMMR to verify the radiative transfer model. A short section of low thermal conductivity waveguide will be added to the cold sky calibration arm to inhibit the propagation of thermal gradients. The pointing of the cold sky horn is being selected to minimize the effects of sun contamination. Finally, a post-launch verification program will be incorporated as part of the calibration program to fine tune the antenna deconvolution brightness temperature calibration.

TMR SYSTEM CALIBRATION RESULTS

A TOPEX/POSEIDON project requirement for the TMR is the correction for water vapor induced path delay with better than 1.2 cm accuracy. This corresponds to a one-sigma brightness temperature calibration accuracy of 1.0 K. Estimates of the effects of the modifications discussed above on the accuracy and precision of the TMR system calibration indicate that these modifications are a necessary and sufficient condition for meeting the project requirement.

The MFFH modification is expected to increase the antenna beam efficiency from approximately 90% to 95%. This corresponds to a decrease, from 1.3 K to 0.5 K, in the RMS error in the antenna temperature correction for the antenna sidelobes. This error is primarily due to uncertainties in the far sidelobes of the antenna pattern, which are typically 50-60 dB below the center beam and are, therefore, impossible to measure with sufficient accuracy. Lower far sidelobes will reduce the effects of errors in the pattern measurements.

An improved model for the radiative transfer of the incoming brightness temperature signal through the initial MFFH, waveguides, and ferrite switches in the TMR has been developed. The original SMMR model did not include the effects of impedance mismatches in the hardware, which were particularly significant at the waveguide-to-MFFH interface due to the wide bandwidth requirement on the antenna. This omission did not show up in the pre-flight temperature cycling tests which were run because no significant temperature gradients were introduced along the path of the signal flow. The different sources of hardware self-emission which showed up in the test data could, therefore, not be distinguished from one another. The redesign of the MFFH has improved its impedance match. In addition, more extensive, gradient inducing, temperature cycling is planned for the TMR. Temperature cycling models suggest that the improved radiative transfer model should be able to reduce the effects of a worst case 10 K hardware gradient on the accuracy of the antenna temperature calibration from 1-2 K error, in the case of the SMMR, to a 0.6 K error, in the case of the TMR.

Total system verification and calibration will be achieved after launch by intercomparisons of the TMR data with a set of ground stations. The ephemeris measured at these ground stations during TOPEX/POSEIDON overpasses will include surface temperature, pressure, humidity, and windspeed, as well as upward looking brightness temperature at 20.7 and 31.4 GHz, by a water vapor radiometer (WVR). The WVR frequencies are chosen to derive integrated water vapor content in the air column. This represents a "ground truth" test of the TMR estimate for wet troposphere induced path delay. In addition, the WVR data, together with the other surface ephemeris, can be used to estimate the expected differences between the brightness temperatures of the various TMR channels. The differences, rather than the brightness temperatures themselves, are used to reduce the effects of modeling errors regarding the surface emissivity. Discrepancies between the estimated differences and the TMR measured differences will be attributed to calibration errors and corrected accordingly. Models of this technique indicate a relative calibration accuracy between TMR channels of less than 1.0 K after correction.

ACKNOWLEDGEMENTS

This work was carried out at the Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA, under contract to the National Aeronautics and Space Administration.

REFERENCES

Milman, A.S. and T.T. Wilheit, "Sea Surface Temperatures from the Scanning Multichannel Microwave Radiometer on Nimbus 7", J. Geophys. Res., Vol. 90, No. C6, pp 11631-11641, 1985.

Njoku, E.G., J.M. Stacey and F.T. Barath, "The SeaSat Scanning Multichannel Microwave Radiometer (SMMR): Instrument Description and Performance", IEEE J. Oceanic Eng., Vol. OE-5, No. 2, pp 100-115, 1980a.

Njoku, E.G., E.J. Christensen and R.E. Cofield, "The SeaSat Scanning Multichannel Microwave Radiometer (SMMR): Antenna Pattern Corrections - Development and Implementation", IEEE J. Oceanic Eng., Vol. OE-5, No. 2, pp 125-137, 1980b.

SeaSat-JASIN Workshop Report Vol. I, "Findings and Conclusions", JPL Publ. 80-62, pp 5-81 - 5-86, 1980.

Swanson, P.N. and A.L. Riley, "The SeaSat Scanning Multichannel Microwave Radiometer (SMMR): Radiometric Calibration Algorithm Development and Performance", IEEE J. Oceanic Eng., Vol. OE-5, No. 2, pp 116-124, 1980.

Advanced Microwave Scanning Radiometer (AMSR): preliminary study on calibration approach

Yasuyuki ITO, Kazuo TACHI, Yuji SATO and Kohei ARAI*

Tsukuba Space Center, NASDA

1-1, Sengen 2, Tsukuba, Ibaraki 305 Japan
Phone 0298-52-2276, Fax. 0298-52-2299

*) Earth Observation Center, NASDA

Abstract

The Advanced Microwave Scanning Radiometer (AMSR) is a new type of spaceborne microwave radiometer which employs electrical beam scanning. The AMSR measures microwave radiation from the earth's surface and atmosphere at frequencies of 6.6, 10.65, 18.7, 23.8 and 36.5 GHz with the improved radiometric sensitivity, spatial resolution and swathwidth.

A specific feature of the AMSR is the new concept of receive-beam scanning of which principle is a beam switching technique. The antenna to realize the concept consists of feed horn arrays combined with switching circuits and a reflector. The number of the receive-beams is same as that of beam footprints on the earth's surface, and the beams are scanned by switching the feed horns. This beam scanning concept has great merits for microwave radiometer with high spatial resolution; first of all, little disturbance to the platform attitude for it has no mechanical antenna movements. Then, multi-frequency measurement, and conical scanning for preserving polarization purity are available. Furthermore, it makes possible multi-beam scanning by using two or more receivers per channel to achieve high radiometric sensitivity. Finally, a push-broom concept will be applicable in the future to achieve more improved performance.

The AMSR is one of the candidate instruments to equip the first NASA's polar platform (NPOP-1) scheduled in 1996. Currently, the overall analysis and definition of the instrument are being carried out as a phase A study. Accommodation studies to the platform are also being performed. As the results of conceptual design has already been discussed in [1], a preliminary consideration on calibration plan is described in this paper.

1. Introduction

An absolute accuracy of observed brightness temperature is one of the most important performances of a microwave radiometer as a remote sensing instrument.

The AMSR is the first challenge to employ electrical scanning using switching array technique. This is to achieve users' basic requirements, such as high spatial and temperature resolution, constant incidence angle over the swath, and high beam efficiency, without disturbing the platform attitude by rotating large antenna. The design is simple from mechanical point of view, i.e. no moving components, but complicated from electrical point of view.

The purpose of this paper is to describe the approach to calibrate the AMSR which employs new antenna concept. As the development phase is too early to define the detailed methods and analysis, only fundamental consideration regarding the calibration approach is discussed.

As for the definition, the following well-known definition is applied in the case of AMSR:
"A radiometer is said to be calibrated when an accurate relationship has been established between the receiver output voltage and the antenna integrated absolute brightness temperature." [2] An antenna pattern correction (APC) is outside the scope of this report.

2. AMSR calibration plan

A radiometer calibration process involves two steps: (1) calibration of the receiver, and (2) calibration of antenna and connecting network (e.g. waveguide). The former step is generally performed by using hot and cold stable reference sources. In the latter step, compensation is made for the attenuation and mismatch losses suffered in the antenna and connecting networks as well as thermal noise emitted by those structures. The data of components'

characteristics and the system tests using special techniques are used for the compensation. [2]

The AMSR is one of the Dicke type microwave radiometers which calibrate observed data by frequently comparing with high and low reference sources. Therefore the receiver calibration, the first step described above, is to be performed successfully in principle. The second step should be examined carefully because of the new antenna concept.

The conceptual flowchart of the AMSR calibration strategy is shown in Figure 1, which includes both steps. The major elements to affect the input noise signal are RF characteristics of the hardware such as loss, mismatch (i.e. VSWR), etc. and the temperature of the hardware. Therefore, in the first place, a temperature transfer analysis model is to be made taking account of these characteristics and the primary temperature transfer equation is to be defined. Then the primary temperature transfer equation is to be corrected, applying the results of system tests (temperature calibration tests). Finally, the final temperature transfer equation shall be defined after the last correction using flight data.

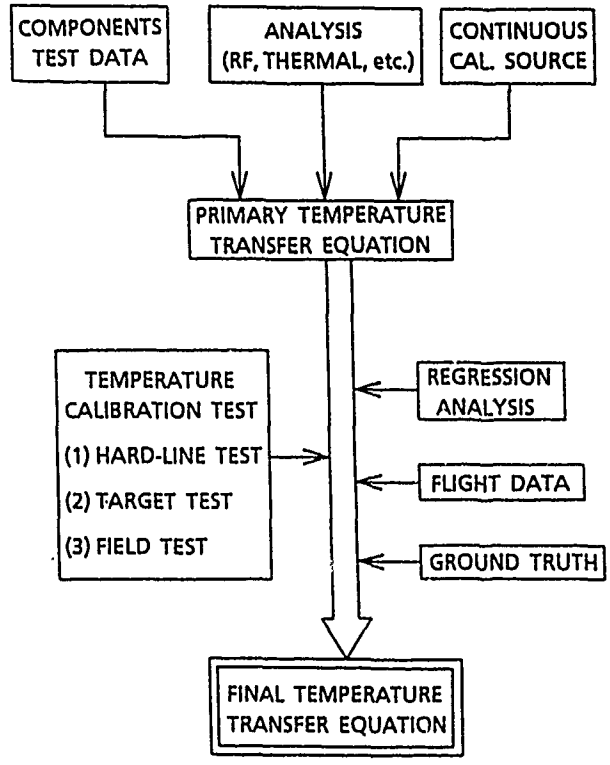


Fig. 1
Conceptual flowchart of AMSR calibration.

2.1 Temperature transfer equation.

The connecting network from AMSR antenna to the input port of the Dicke switch consists of several components such as waveguide, switches, Orthogonal Mode Transducer (OMT), etc. The sky horns and the reference loads are connected to the Dicke switch through those components as well. A temperature transfer analysis model is to be made considering the following elements which suffer the received signal:

(1) loss of each component,
(2) VSWR of each component,
(3) isolation of each component,
(4) temperature of each component.

When the calibration is done, components' data is to be used for (1)-(3) and telemetry data is to be used for (4). These figures are considered to be known with uncertainty of measurement.

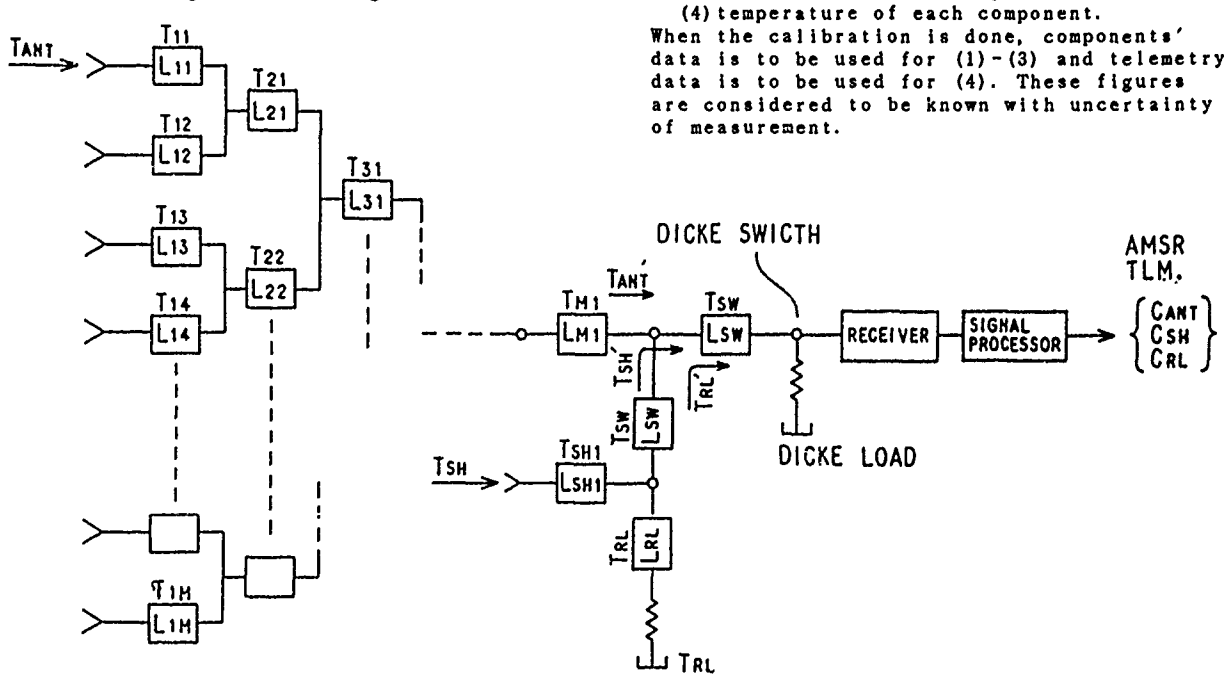


Fig. 2 A concept of temperature transfer analysis model.

Figure 2 is a concept of the temperature transfer analysis model.

Temperature transfer equations, which transfer the noise temperature at the antenna into the noise temperature at the input port of Dicke switch, are then to be defined based on the analysis model. The equations to transfer the noise temperature at the sky horn into the noise temperature at the input port of Dicke switch and to transfer the noise temperature at the hot reference load into the noise temperature at the input port of Dicke switch are also to be defined.

The antenna noise temperature at the input port of Dicke switch, Tant', is expressed as below:

$$Tant' = Tsh' + (Trl' - Tsh')\,(Cant - Csh) / (Crl - Csh) \qquad \text{Eq. 1}$$

where Tant':antenna noise temperature at the input port of Dicke switch, Tsh':sky horn noise temperature at the input port of Dicke switch, Trl':reference load noise temperature at the input port of Dicke switch, Cant:digital count of antenna noise temperature, Csh:digital count of sky horn noise temperature, Crl:digital count of reference load noise temperature.
Tant' is also expressed as below:

$$Tant' = f(Tant, L11, T11, L21, T21, \ldots, Tm1, Lm1, \ldots) \qquad \text{Eq. 2}$$

where Tant is the antenna noise temperature to be derived, and Lxx and Txx are loss and temperature of each component of the connecting network, as shown in Figure 2.

The sky horn noise temperature at the input port of Dicke switch, Tsh', is expressed as below:

$$Tsh' = g(Tsh, Lsh1, Tsh1, Lsw, Tsw, \ldots) \qquad \text{Eq. 3}$$

The reference load noise temperature at the input port of Dicke switch, Trl', is expressed as below:

$$Trl' = h(Trl, Lrl1, Trl1, Lsw, Tsw, \ldots) \qquad \text{Eq. 4}$$

The primary temperature transfer equation is derived from the Equations 1, 2, 3, and 4.

$$Tant = F(Tsh, Trl, Cant, Csh, Crl, Ll1, Tl1, \ldots) \qquad \text{Eq. 5}$$

As figures of the right side in Equation 5 are known, the antenna input noise temperature which is needed as calibrated observation data can be calculated.

2.2 System tests.

The primary temperature transfer equation is not supposed to be accurate enough because of errors in data base. Several pre-launch tests at system level are planned to get more accurate data base to correct the primary temperature transfer equation.

(1) "Hard line" Test

The conceptual diagram of the "Hard line" Test is shown in Figure 3(a). This is a test to acquire the calibration data of whole system but the feed horns and the reflector.

(2) "Target" Test

The conceptual diagram of the "Target" Test is shown in Figure 3(b). This is a test to acquire the calibration data of whole system but the reflector.

(3) Field Test

The conceptual diagram of the Field Test is shown in Figure 2(c). This test has a merit to acquire the calibration data of whole system.

After the launch, the temperature transfer equation is to be corrected using the telemetry data such as instrument temperature and the ground truth data.

2.3 Error sources left uncalibrated.

Potential error sources of the AMSR are shown in Figure 4. Where each source is categorized in the following three, considering the discussion in the previous section;

- *1) can be calibrated by analysis based on design and components' data,
- *2) can be calibrated by system tests,
- *3) may have residue of errors after above two calibration.

Therefore it is concluded that the absolute accuracy consists of the errors categorized "*3". In addition, the errors correspond to the accuracy of tests and analyses methods are unavoidable.

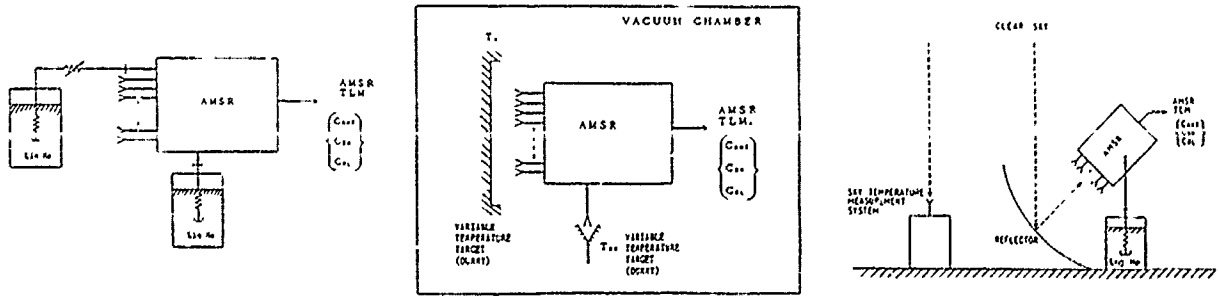


Fig. 3 Conceptual diagram of system tests.
(a)"Hard line", (b)"Target", (c)Field test

The residue of errors, i.e. contents of absolute accuracy, are are divided to the following two elements:

(1) "Bias" error (independent on time, therefore on temperature)
(2) "Random" error (dependent on time, therefore on temperature)

Random errors can be reduced by accurate measurement of components characteristics and temperature monitoring, while bias errors by accurate system tests.

3. Conclusion

The fundamental consideration regarding the calibration approach for the AMSR has been discussed.

In practice, it is understood that a precise thermal control of the instrument in orbit is the key to a reasonably complex calibration scheme.

The preliminary thermal analysis assuming passive thermal control shows that the feed array and switches are expected to be kept within a range of about 12 degrees peak to peak. Random errors caused by the feed array and switches in the output noise temperature are estimated to be about 0.28 K, assuming state-of-the-arts measurement. This factor is the only one which is included in AMSR but not in other Dicke type radiometers.

As for the bias errors, 0.4 to 0.7 K has experienced in the "Target" test of Microwave Scanning Radiometer (MSR) which equipped the the Marine Observation Satellite (MOS-1). [3] Therefore similar or better accuracy is expected for AMSR.

By implementation of high beam efficiency antenna, precision thermal control and temperature monitor, and efforts in tests and analyses, the calibration of the AMSR is expected to be reasonably good.

Acknowledgment

The authors wish to express their thanks to Drs. T. Suzuki, A. Shibata and T. Wilheit for their discussions. Technical contribution by MELCO staff is also appreciated.

References

[1] K. Tachi et. al., "Advanced Microwave Scanning Radiometer (AMSR) : Requirements and Preliminary Design Study", IEEE Trans. GE-RS, Vol. 28, No. 2, March 1989

[2] Manual of Remote Sensing, Vol. 1, American Society of Photogrammetry, pp. 475-516

[3] Y. Ishizawa et. al., "Microwave Scanning Radiometer (MSR) for Marine Observation Satellite-1 (MOS-1)", Proceedings of the 14th International Symposium on Space Technology and Science, Tokyo, 1984.

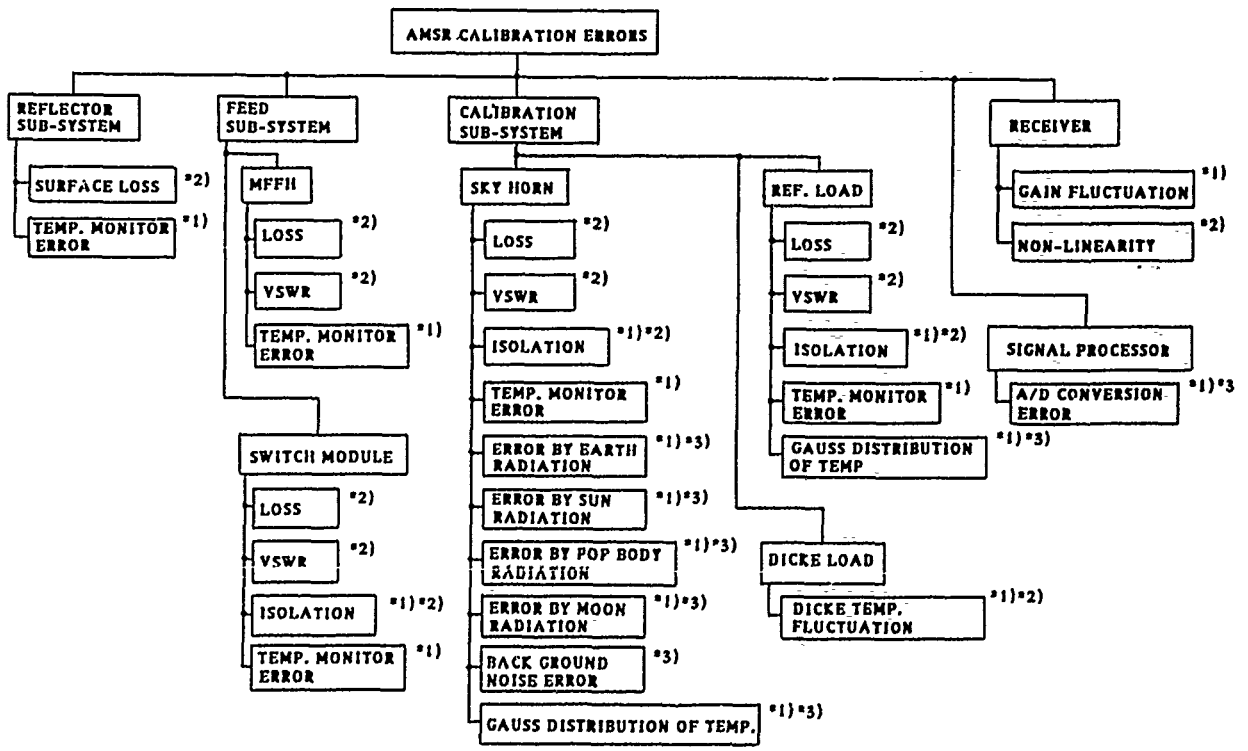


Fig. 4 Potential error sources of AMSR.

Millimeter-Wave Electronically Scanned Imaging Radiometer System For Shipborne Ice Navigation Application

G. Healy and C. Grant

COM DEV,
155 Sheldon Dr.,
Cambridge, ON
Canada N1R 7H6
Tel: (519) 622-2300
Fax: (519) 622-1691
Twx: 610-366-3164

Institute for Space and
Terrestrial Science (ISTS),
YorkUniversity,Petrie Building,
4700 Keele St., North York, ON
Canada M3J 1P3
Tel: (416) 739-4685
Fax: (416 739-4687

Abstract

Presented here is a new type of millimeter wave passive remote sensor, namely a 36-38 GHz electronically (frequency) scanned computer-automated imaging radiometer system. This system was designed, developed and tested as a joint COM DEV/ISTS research and development "proof-of-concept" project to address a potential Canadian requirement for shipborne sea ice detection and type discrimination. Radiometric signals, received from a given scene and processed under continuous calibration, are imaged in terms of radiometric brightness temperatures on a computer equipped with a high resolution colour monitor. Electronic tuning of the azimuth scan angle is readily achieved through employment of a double downconversion receiver design. A mechanical elevation scan capability is also built in as a prototype auxiliary capability to enable a complete sectoral image to be generated. Initial system radiometric resolution factors of the order of 3 K (typical) have been measured.

Keywords: mm wave radiometer, electronic frequency scanning, sea-ice detection

1. Introduction

A 36 38 GHz electronically-scanned computer-automated imaging radiometer system has been designed, developed and tested as a joint COM DEV/ISTS research and development "proof-of-concept" project Primarily the impetus behind the project was the need for an instrument for shipborne ice navigation in the near range region (within approx. 1 km). Currently operated radar systems appear to have problems with ice detection, classification and edge definition in this region. Secondly, the system's ability to perform azimuth scan without the use of moving parts appeared advantageous in the cold-weather conditions previewed. If used in conjunction with radar the system could ideally provide the real time on-board ice detection required by a vessel in arctic regions. The purpose of this paper is to fully outline this system's configuration and operation, beginning with a convenient subdivision of the system and subsequent individual sub-unit analysis and ending with preliminary system operational results together with a set of derived conclusions.

2. System Configuration Overview

The system, shown in Photo 1, has its configuration highlighted in Figure 1. It consists of two basic units - a radiometer subsystem and computer hardware/software - remotely linked by up to 200 feet of cable. The radiometer subsystem which resides in a radome shielded enclosure (42" x 15" x 23") may be divided into a number of sub-units listed as follows:

(a) Antenna
(b) RF hardware unit
(c) Video electronics unit
(d) Elevation motorized scan
(e) Data control and acquisition unit

The computer hardware/software controls system operation through a menu-driven user-friendly program, retrieves and interprets data returns and displays images on a high resolution colour monitor.

Each of the critical sub-units noted above will now be outlined in more detail.

2.a. Antenna

A serpentine antenna is used which, although following well known leaky-wave design theory [1], has rarely been applied to radiometry. The construction consists of slotted waveguide that is fed at one end and terminated in a matched load at the other. The interference pattern created between the electromagnetic radiation from each slot generates a central maximum or main beam in a perpendicular direction at centre frequency. Since the phase velocity within the slotted waveguide varies with frequency, the beam is effectively shifted toward the feed-point with decreasing frequency and away from the feed-point with increasing frequency. This feature provides a relatively simple means of performing electronic scanning without the need of multiple detectors, phase shifters and correlators. By selecting a specific receive frequency a corresponding beam direction follows. For our particular antenna a frequency scan from 38.02 GHz to 36.20 GHz corresponds to an azimuth beam scan from -33.5 degrees to +41 degrees with respect to the normal. An added advantage of the serpentine antenna is an almost constant beam-width throughout the scan. Although gain does drop off slightly with lower frequency due to the wavelength versus slot size relationship, the change is much smaller than for a dish.

The serpentine antenna has many positive attributes but there is one potential disadvantage when applied to radiometry. Since radiometric receivers are characteristically wide-band to optimize noise sensitivity, there is an inherent trade-off imposed by the serpentine design. A restriction on beamwidth corresponds to a restriction on RF bandwidth. For this particular system an antenna beamwidth of 2.5 degrees restricts the effective receiver bandwidth to 70 MHz.

2.b. RF space hardware unit

A general block diagram of the RF hardware subsystem, which is a total power radiometer [2] with continuos calibration, is shown in Figure 2. A finline pin diode SPDT Dicke switch enables the system to switch between the RF arm and a noise arm. This noise arm contains a calibrated noise source which may be switched on and off to provide two points of calibration for any given RF reading. A finline bandpass filter is then used to limit operation to the signal frequency band of interest (36-38 GHz) providing, in particular, greater than 35 dB isolation from the image band (30-32 GHz). A suspended substrate crossbar mixer fed by a DRO (Dielectric Resonator Oscillator)/multiplier L.O. subsystem provides the 1st downconversion stage to the 2-4 GHz range. Followed by an IF LNA (noise figure$<$1.3 dB, gain$>$30 dB) this mixer needed a state-of-the-art 5 dB noise figure to maintain critical receiver noise performance. A second downconversion stage contains the computer controlled frequency swept YIG (Yttrium Iron Garnet) oscillator which effectively selects the azimuth scan position as determined by the back-end bandpass filter (1.566 $\pm$ .035 GHz). This filter, which follows a lowpass filter and a number of gain stages, is of narrow bandwidth (70 MHz) corresponding to effective antenna beamwidth as stated previously. Following a final gain stage which optimizes signal power levels, a square law detector provides the system with its radiometric d.c. signal output. Tested separately this radiometer unit yielded noise temperatures of the order of 1200 K (typical) which, given the RF bandwidth and an integration time of 50 ms, corresponds to an effective noise temperature resolution of approximately 0.6 K.

2.c. Video electronics unit

The video electronics unit performs three distinct functions, all of which are summarized in the layout of Figure 3. Initially, digitally controlled gain stages enable linearization of the voltage output available across the band for a given azimuth scan. Following this a differential amplifier and offset reference is used to negate the major d.c. contribution of receiver noise. The offset voltage chosen is based on the lowest possible system voltage appearing at the amplifier which, in turn, is based on the lowest possible brightness temperature encountered. Finally, an operational amplifier in a low pass filter configuration is used for data integration This stage was fixed for preliminary testing to yield an equivalent integration time of 50 ms. Variable integration time is however achievable through multiple sampling of any fixed base integration value.

2.d. Elevation motorized scan

As an extra option a mechanical elevation angular scan capability was built into the system to enable computer controlled scanning of entire sectoral areas in front of the system without the need for ship movement. This option was intended for use for research purposes only and any future operational instrument would, in principle, not contain the same. A digitally controlled stepping motor allows the antenna to scan vertically up to $\pm$ 20 degrees about a fixed downward platform view (nominally 69 degrees with respect to vertical) with a stepping resolution of approximately 0.1 degree.

2.e. Control and data acquisition unit

The radiometer is controlled by several registers that are addressed across a shielded multi-conductor cable linking the sensor to the a computer. These lines are buffered by line drivers and receivers adhering to the RS-422 protocol. Without these RF and switching interference can cause spurious data to be written to control registers rendering operation unpredictable. Given the fact that the electrical environment on board a ship is quite harsh proper buffering is therefore imperative.

There are eight registers within the radiometer control and data acquisition unit. Two registers control a resident D/A board which selects the receive frequency through a voltage controlled YIG oscillator. Two more registers contain a digital value obtained by the latest sample of an accompanying A/D converter which, in turn, is connected to the output of the video electronics board. One register is devoted to adjusting the frequency of the timing signal that ensures a significant delay occurs between A/D samples such that statistical independence is preserved regardless of the CPU clock rate. Another register is devoted to basic hardware control including the dicke switch position, the on/off state of the noise source and the variable gain of the video electronics. Finally, two registers are devoted to system status, one for output and the other for input. Through these, the A/D can be reset and various operation complete flags can be monitored.

Most of the logic controlling the vertical stepping motor is housed in the computer. Therefore, the radiometer requires only a motor clock pulse signal and a rotation direction signal to operate. It returns an end-stop signal for both extreme motion limits. These signals are also RS-422 buffered and travel through the shielded cable.

2.f. Computer hardware and software

The radiometer operation is controlled by an IBM PC-AT compatible. The choice of such a machine includes several advantages like lower development cost, readily available documentation, portability, and multi-functional capability (data analysis). Six I/O addresses are used. Three addresses, namely data read, data write and address control, access the registers within the radiometer. Data is obtained from the radiometer via data read. Data is transferred to the radiometer through data write. Selection of a specific register in the radiometer occurs through address control. The three remaining addresses govern stepping motor operation. Two supply a count corresponding to a specific stepping motor position and the third initiates either motor movement or motor reset. Hard-wired logic compares the position count supplied with the actual position of the motor. When the motor movement pulse arrives the logic provides a locally generated pulse to the stepping motor until the two positions match.

The software is menu driven, providing the user with many options. One may choose vertical scanning mode where the main beam is controlled by the stepping motor only. This creates brightness temperature profiles leading away from the sensor. Alternatively, one may select horizontal scanning mode where the main beam is only swept electronically. This provides

profiles along a line perpendicular to the forward looking direction. The user can ultimately select imaging mode. This combines both the vertical and horizontal scanning modes so that a complete scene is mapped in two dimensions. The radiometric brightness temperature maps are created on a professional graphics screen in real time. At any instant, the operator can locate the footprint by observing the graphic outline. The radiometric sensitivity is also adjustable by increasing the integration time. As the integration time rises, more statistically independent noise voltage samples are collected and averaged. All data collected can be logged and replayed. The data is stored in ASCII format to permit easy import to statistical analysis programs such as Statgraphics or Lotus 1-2-3.

3. Preliminary Results

The system was initially tested in an indoor laboratory environment using the common hot cold load technique. The hot load used was a piece of absorber material at room temperature (295 K), whilst the cold load was the same material immersed in liquid nitrogen (77 K). Noise temperature resolution factors of the order of 3 K were recorded.

To test the system's imaging capabilities, following a sky/load calibration the radiometer was mounted on an antenna tower platform and connected through two hundred feet of cable tc the computer hardware/software which resided in an antenna hut (see Photo 2). Initial testing showed the system to be capable of creating crude images especially when high contrast objects were observed. Such an example is illustrated by the system scan of the scene shown in Photo 3 (two cars and an aluminum sheet on a grass background) which yielded the image shown in Photo 4. The three target objects were seen to be clearly distinguishable and both object size and range values appeared to correlate well.

Recently tested on board the CCGS Sir John Franklin in the marginal ice zone off Newfoundland during LIMEX'89 the system was found to be capable of discriminating between water, nilas (new ice) and first year ice.

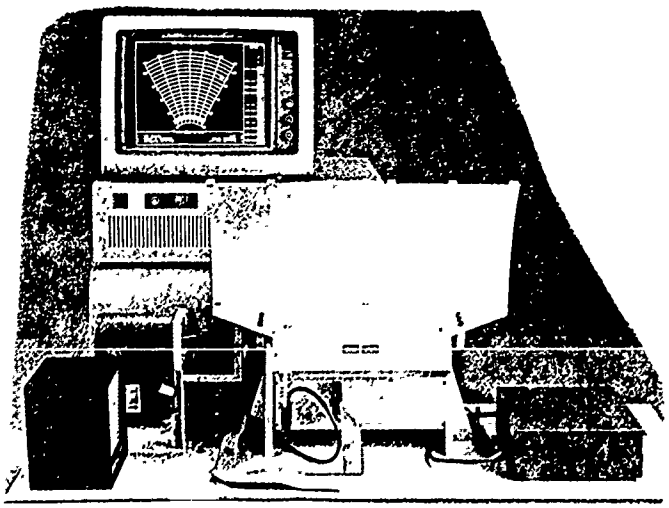


Photo 1 : Physical system layout.

The three preliminary test result experiences outlined above have essentially proved the concept undertaken in initial system prototype development. However two major limitations exist in this current system. The first of these is in the area of system resolution. Measured degradation in resolution from the RF hardware unit figure of 0.6 K to 3 K is directly attributable to the relatively lossy serpentine antenna which has an approximate equivalent loss of 3 dB. The second, an azimuth scan imaging time of approximately 20 s - as dictated by the integration time necessary to compensate for the narrow system bandwidth - is relatively slow. Both of these problems will have to be resolved in order for any "follow-on" instrument to satisfy shipborne operational requirements envisaged, which are the order of 1 K resolution and 5 s scan speed.

4. Conclusions

The approach to radiometer system design taken here is unique in that it employs a serpentine antenna to achieve electronic scanning. This design has the advantage of low cost and relatively simple construction with respect to other electronic scanning systems, and the obvious added advantage of "no moving parts" with respect to mechanically scanned systems. Unfortunately it displays limitations with regard to resolution and speed. Resolving these limitations will be the subject of future research before such an instrument becomes fully operational.

Acknowledgements

The authors would like to thank a number of members of the EHF group at COM DEV including Bill Thorpe, Robyn Buffet, Brian Vansickle, Rob Moore and Steven Low. The authors are also grateful to Dr. Rene Ramseier of ISTS for his guidance and encouragement throughout this project.

References

[1] Jasik H.: "Antenna Engineering Handbook," Chapter 16, McGraw-Hill, New York, 1970.

[2] Ulaby F. T., Moore R.K. and Fung A.K.: "Microwave Remote Sensing Active and Passive : Volume 1", Chapter 6, Addison-Wesley, Massachusetts, 1981.

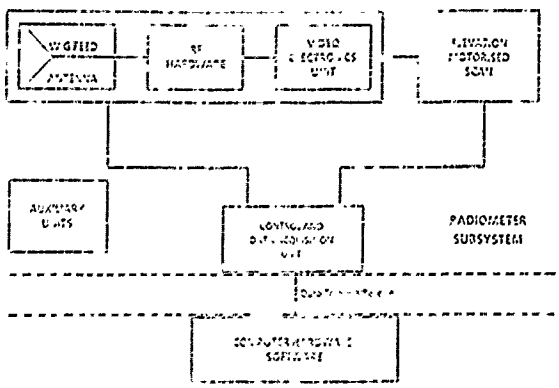


Figure 1 : System block diagram listing various sub-units.

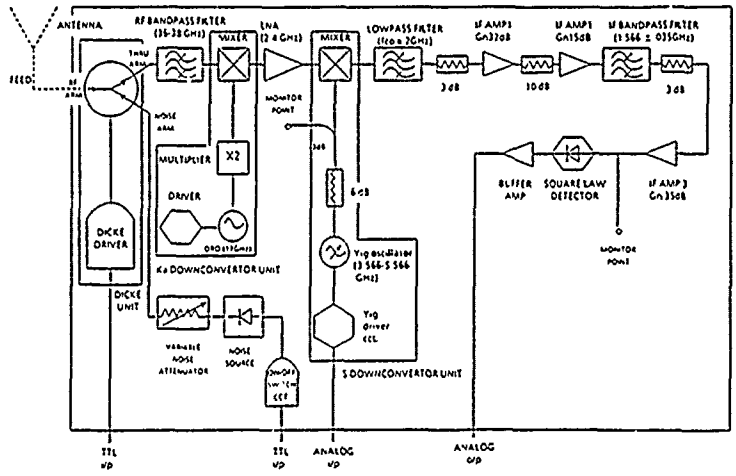


Figure 2 : RF hardware unit block diagram.

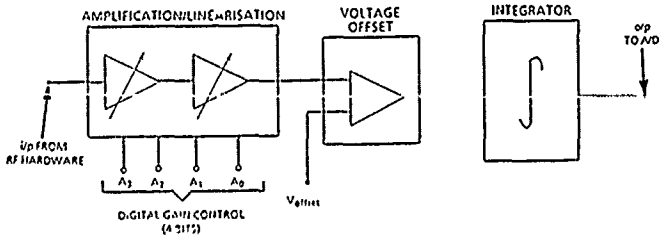


Figure 3 : Video electronics unit functional layout.

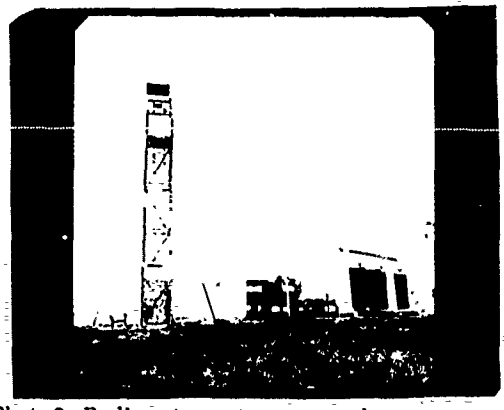


Photo 2 : Radiometer system mounted on 10 m high antenna tower positioner [left]. Antenna building containing computer hardware/software [right].



Photo 3 : Photo centre coincides with scan centre (approx.). Chevrolet Malibu, 5 m x 2 m (approx.) [left centre]. Sheet of Aluminum, 2.4 m x 1.2 m (approx.) [left below]. Renault Encore, 4 m x 1.2 m (approx.) [right centre].

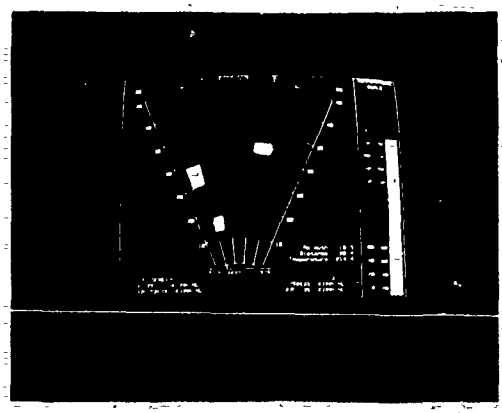


Photo 4 : Generated radiometer scan pattern (automatic scanning mode). Object size and distance correlation (approx).

DUAL FREQUENCY MICROWAVE RADIOMETER

XU PEIYUAN, JIN PEIYU, YAO ZHUANLIANG, FANG ZHENHUO, LI MINGXIANG, WANG ZHENRONG, ZHENG JIAHAO
(Insitute of Electron Physics, Shanghai University of Science and Technology, Shanghai, China)
ZHAO CONGLONG, LIN BIN
(Academy of Meteorological Science, State Meteorological Administration, Beijing, China)

ABSTRACT

A dual frequency microwave radiometer is developed to remotely sense water vapor and liquid in the troposphere. A pair of optimum frequency, 20.60 and 31.65 GHz, is adopted. Each frequency Dicke radiometer has its own antenna and receiver, but both radiometers are controlled by a microcomputer system and mounted on a rotatable pedestal. The sensitivity at 1sec time constant and the accuracy of the brightness temperature is 0.2K and 0.5K, respectively, at each frequency.

KEY WORDS

Microwave radiometer, Passive microwave remote sensing, Tropospheric water, Tropospheric liquid.

INTRODUCTION

The ammount of tropospheric water vapor and liquid is highly variable in both time and space. The microwave radiometry is the most promising method to continuously measure these variation in real time. A steerable dual frequency microwave radiometer is developed in China. A pair of optimum frequencies, 20.60 and 31.65 GHz, is adopted (Wu 1979, Hogg et al. 1983). The 20.60 GHz frequency is removed from the peak of the absorption line (22.24GHz) of water vapor, therefore at this frequency the change in absorption caused by the pressure broadening is minimize. The 31.65GHz frequency is situated in a transmition window of the troposphere and primarily senses the liquid in cloud. The brightness temperature at these frequencies incorporated in a pair of simultaneous equations to retrieve the amounts of water vapor and liquid.

The block diagram of this radiometer is shown in Fig.1.

ANTENNA SYSTEM

A 90° offset paraboloid reflector with a conical corrugated feed horn is adopted as antenna system for each frequency. The antennas are put on a rotatable azimuthal pedestal. The beams are synchronously scanned in azimuth by rotation of the pedestal about its vertical axis and in elevation by rotation of the reflectors about their offset axes. These axes are mechanically coupled, so only one stepping motor is used to drive the reflectors in elevation. The antennas are able to be scanned at 2°/sec in azimuth or elevation. The pointing accuracy is 0.1° in elevation and 0.5° in azimuth. The reflectors with tolerence of 0.1mm is machined from cast aluminum using profile modeling method. The parameters of the antenna systems are selected to ensure that the antennas have nearly similar beams (half power beam width 2.5°), low side lobes (-27dB) and high beam efficiency (90%) at both frequencies.

RECEIVER SYSTEM

The block diagram of the receivers is shown in Fig. 2. Both receivers are similar Dicke radiometers. A ferrite waveguide switch alternates the input to the mixer from the sky to a reference waveguide load. The switches and loads of both frequencies are assembled in a polyfoam package and temperature controlled at 40±0.05°C. The solid state noise diodes as the second calibration noise sources may inject their noise into the receivers through the -33dB couplers. The injected noise are primarily calibrated by the cold (liquid Nitrogen) and hot (ambient temperature) microwave absorbers which successively cover the antenna apertures. The Shottky diode mixer with the Gunn diode local oscillator and intermediate-frequency passband from 50 to 500MHz has 4.0dB noise figure. The digital demodulation is performed by the microcomputer system. In addition to the digital demodulation the analog synchronous detectors are also built in the receivers, which may be operated independently of the microcomputer system. The sensitivity at 1sec time constant and the accuracy of the brightness temperature are 0.2K

and 0.5K, respectively, at each frequency. Both receivers are assembled in one waterproof and thermally insulated enclosure. Its temperature is controlled at 35±1.5°C. This enclosure connected with the feed horns is put on the azimuthal pedestal. Because the offset angels of the reflectors are 90°, this enclosure does not rotate during elevation scan. The receivers have excellent mechnical and thermal stabilization to ensure the stable overall performance.

CONTROL AND PROCESSING SYSTEM

The block diagram of the control and processing system is shown in Fig. 1. There are two independent antenna operation modes, mode A and mode B, which can be selected by observer. In mode A the Z-80 microprocessor system is used to control the antennas. The antenna beams can be pointed to any position and scanned in azimuth or elevation simply by inputting suitable instructions using the keyboard on the panel of the control case. The operating programs are written in assembly languages and stored in an eprom. Antenna positions are displayed digitally at resolution of 0.1° on the panel. In order to ensure the pointing accuracy Hall sensors are used to calibrate the zero degree in reset for both of azimuth and elevation angle. In mode B the IBM PC/XT microcomputer system controls the antennas to point and scan the beams and to sample and store the antenna position angles. Independently of the antenna operation mode selection the microcomputer system also controls and processes the following items.

A. Switching the ferrite switches and noise diodes according to the modulating and calibrating sequences.

B. Sampling averaging and storing the following quantities corresponding with the sequences mentioned above. (a) The output of the converters. (b) The temperature of the reference waveguide loads. (c) The ambient air temperature, dew point and pressure. (d) The voltage of various power supplies.

C. Digital synchronous demodulating (Janssen 1985).

D. Calibrating, integrating, averaging and storing the demodulation output to get the brightness temperature.

E. Retrieving the integrated amounts of water vapor and liquid or the water vapor profile.

F. Displaying and printing the intermediate data for data quality checking and instrument self-diagnosis.

This microcomputer system provides considerable advantages in terms of continuous and unattended measurement of water vapor and cloud liquid in the troposphere.

APPLICATIONS

Since last October a prototype dual frequency microwave radiometer has been operated in the Academy of Meteorological Science at Beijing, China. The retrieved water vapor profiles have been compared with the radiosonde measurements (see Fig. 3). The operation experience has shown that this instrument is stable and reliable. As a water vapor profiler, it will join into the operational experiment on mesoscale meteorology in this summer.

The second radiometer is being assemblied in the Institute of Electron Physics at Shanghai, China. It is planned to join into the geodetic VLBI and GPS observation to promote the excess path length correction to the centimeter level for geodynamic and seismic research.

REFERENCES

1. Wu S.C., "Optimum Frequencies of a Passive Microwave Radiometer for Tropospheric Path-length correction", IEEE Trans. Antennas Propagat., vol.AP-27, pp233-239, 1979.

2. Hogg D.C.,Guiraud F.O.,Snider J.B.,Decker M.T.,and Westwater E.R.,"A Steerable Dual Channel Microwave Radio meter for Measurement of Water Vapor and Liquid in the Troposphere",J. Climate Appl. Meteor.,vol.22, pp789-806,1983.

3. Janssen M.A.,"A New Instrument for the Determination of Radio Path Delay Due to Atmospheric Water Vapor", IEEE Trans.Geosci. Remote Sensing,Vol.GE-23, pp485-490,1985.

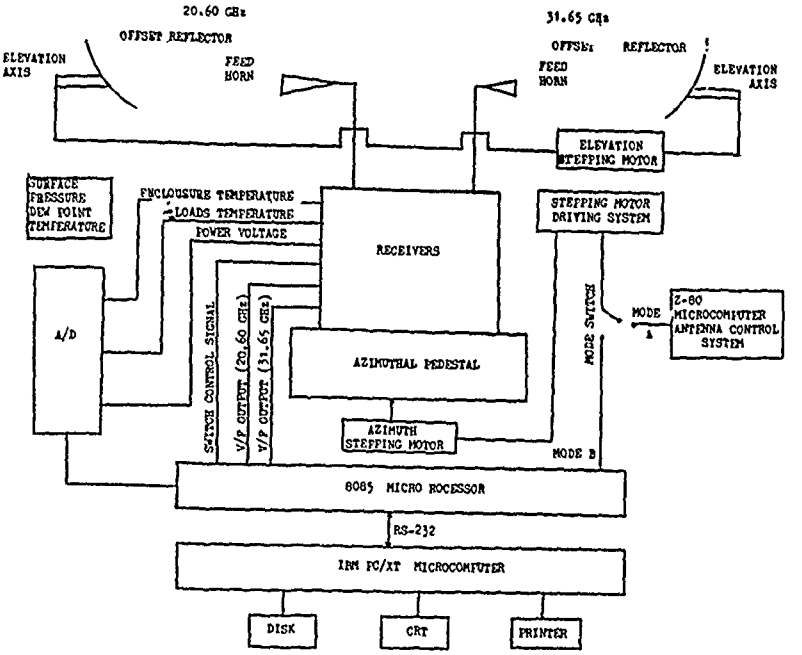


Fig. 1 Block diagram of the dual frequency radiometer

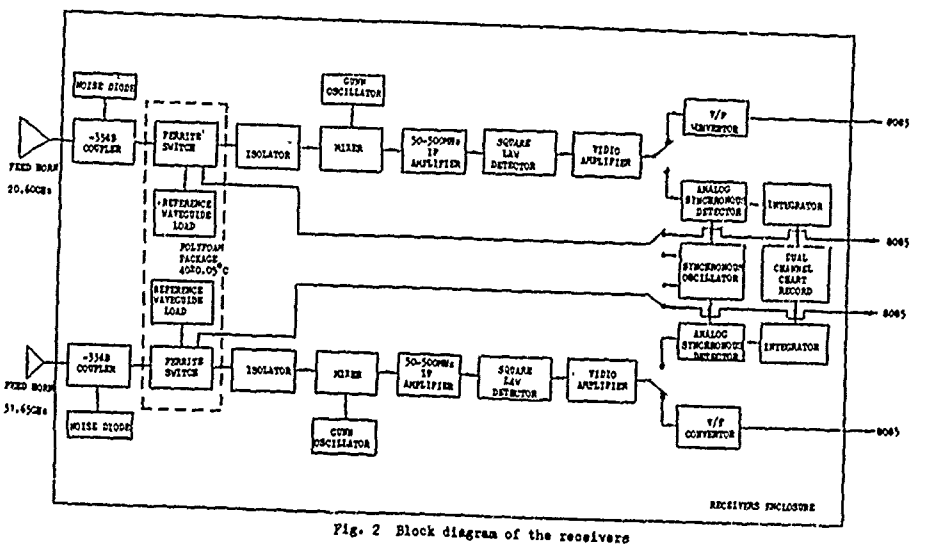


Fig. 2 Block diagram of the receivers

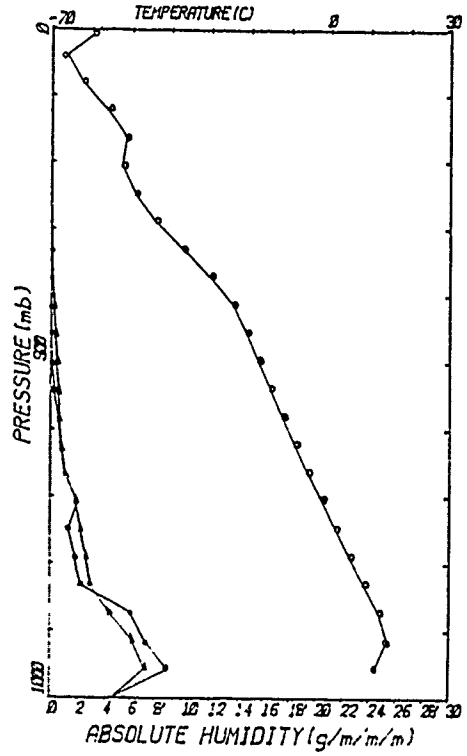


Fig. 3 Water vapor profile

VAPOR PROFILE RETRIEVALED WITH DATA FROM DUAL-FREQUENCY MICRO-WAVE RADIOMETER AND RADIOSONDE

FREQUENCY: 20.6 31.6 GHZ
TIME: 1988/10/19 7:15:00
LOCATION: BEIJING PRC

TEMPERATURE
VAPOR(radiosonde)
VAPOR(radiometer)

CAMS

SOIL MOISTURE ESTIMATION WITH L-BAND RADIOMETRIC MEASUREMENTS

J. R. Wang and J. C. Shiue
NASA/Goddard Space Flight Center
Greenbelt, Maryland 20771

E. T. Engman and T. J. Schmugge
USDA/Agricultural Research Service
Beltsville, Maryland 20705

ABSTRACT

During May-October of 1987, a total of four Intensive Field Campaigns (IFC) distributed over 57 days were conducted for the First ISLSCP (International Satellite Land Surface Climatology Project) Field Experiment (FIFE). FIFE took place at and around the Konza Priarie Long-Term Ecological Research site near Manhattan, Kansas. The objectives are to obtain the necessary data to permit interpretation of satellite observations to infer climatologically significant land surface parameters. One of several major elements in this cooperative effort is soil moisture mapping of the test area with the L-band pushbroom microwave radiometer (PBMR) aboard the NASA C-130 aircraft. There were a total of 11 complete flights for the four IFC's, each consisting of 15 flight lines. Twelve of these lines were long lines flown at an altitude of 600 meters, mapping an area of about 8 km by 14 km. Three short lines were flown at 300 meters over two adjacent watersheds (1D and 2D) with a total area of about 1 km by 1.2 km. Four additional flights of three short lines were made over these two watersheds during the first three IFC's. Intensive soil moisture sampling was made concurrently with all the aircraft flights in these two watersheds along the three short flight lines at locations separated by about 75 meters. Soil sampling was also made every other day during the IFC's at other thirty-four locations distributed over the entire experimental area.

We have estimated the surface soil moisture from these radiometric measurements over the entire area covered by the aircraft flights. This was done by making two basic assumptions. The first is that the ground was uniformly saturated after the 6 cm rain of May 27. Thus, for the radiometric measurements of May 28 the soils were known to be saturated and ground soil samples collected in the area gave an average volumetric moisture content of 43% with a standard deviation of 3%. This average value was assumed for the entire area covered by the aircraft flight on that day and, when combined with the radiometric measurements, a fixed point in the regression between radiometric response and soil moisture was determined for each radiometer footprint. The second assumption is that there is a common point of intersection for the regression lines corresponding to different vegetation conditions. This point of intersection was found from the regression lines for watersheds 1D and 2D at the emissivity of 0.892 and soil moisture value of 10.7%. The linear relationship between emissivity and soil moisture for each PBMR footprint is determined by this common point of intersection and the point at moisture value of 43%. The derived linear relation is then used to estimate soil moisture from other days of radiometric measurements. It was found that the estimated soil moisture values compared favorably with those derived from the independent gravimetric measurements. The estimated soil moisture was generally lower by about 1-2%, but this may be accounted for by the fact that most of the ground data were collected in the morning while the radiometric measurements were obtained in late morning or in the afternoon.

AN IMAGING PASSIVE MICROWAVE RADIOMETER FOR PRECIPITATION RETRIEVAL AND MESOSCALE STORM SYSTEM STUDIES

J. A. Galliano and R. H. Platt

Georgia Institute of Technology
Atlanta, GA 30332
404-894-3503
FAX # 404-894-3120

ABSTRACT

This abstract outlines the development and performance of an imaging passive microwave radiometer operating in the 10 to 85 GHz range specifically for precipitation retrieval and mesoscale storm system studies from the ER-2 high altitude research aircraft. The instrument is referred to as the Advanced Microwave Precipitation Radiometer (AMPR) and was developed under contract to Marshall Space Flight Center (MSFC) of the National Aeronautics and Space Administration (NASA). The AMPR's primary goal is the exploitation of the scattering signal of precipitation at frequencies of 10.7, 19.35, 37.0 and 85.5 GHz to unambiguously retrieve precipitation and storm structure intensity information in support of the spaceborne Special Sensor Microwave/Imager (SSM/I) during future airborne ER-2 data flights in 1989/1990.

Key Words: millimeter wave, radiometry, multifrequency, precipitation

INTRODUCTION

Scattering-induced brightness temperature depressions from precipitation are strong enough in the millimeter wave region to provide a meaningful contrast with the radiometrically warm land background (1,2). Higher frequencies (37 GHz and above) yield greater cloud penetration because of less sensitivity to small non-precipitating ice. Significant scattering has also been observed in low brightness temperatures at 92 and 183 GHz in oceanic convection (3). Lower frequencies (18 GHz and below) when used with the higher frequency channels allow an unambiguous separation of the rain signal from wet ground and water bodies, because the emissivity decreases with frequency for precipitation (volume scatterer), while the emissivity increases with frequency for water (emissive surface) (4,5).

Figure 1 provides evidence of how different frequencies of radiation might respond to different heights within a rain system. As the frequency decreases, the depth in the cloud from which most of the information is obtained increases. For precipitation measurements, one would like the response to be from a level as close to the ground as possible. However, the brightness temperature contrast between rain and the warm land background is small at such a low level. At the other extreme (highest frequency), the contrast temperature between the storm and land background is very strong; but it is not likely well related to the precipitation rate near the surface. Therefore, it is advantageous to select an intermediate frequency (such as 37 GHz) that has a relative strong signal due to attenuation by precipitation, and is still responsive to processes from deep enough in the cloud to be well related to rain rate.

Depending on the extent to which precipitation-sized ice in rain systems is important, the double valued character of brightness temperature versus rain rate relationship can lead to ambiguities in the interpretation of single-frequency data. An illustration of how a multifrequency algorithm can help alleviate this problem is shown in Figure 2. Note the double valued character of the 18 and 37 GHz curves for rain cells which extend to increasing heights as the rain rates at the surface increase. By taking the difference between the 18 and 37 GHz brightness temperatures, one obtains a useful relationship which is single valued throughout its range. This is one illustration of the possible advantages of utilizing multiple frequencies for the measurement of rain rates.

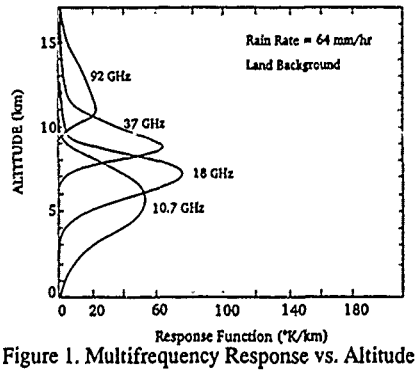


Figure 1. Multifrequency Response vs. Altitude

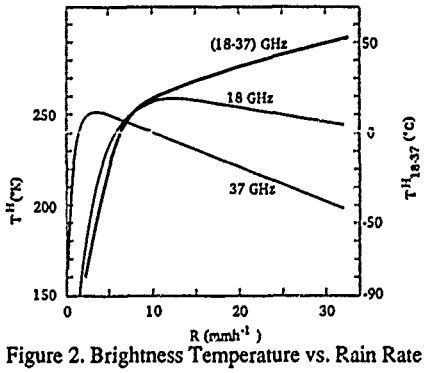


Figure 2. Brightness Temperature vs. Rain Rate

RADIOMETER SPECIFICATIONS

The AMPR is a total power radiometer operating at frequencies of 10.7, 19.35, 37.0, and 85.5 GHz and having the following primary characteristics: a total antenna aperture of fifteen inches, a cross-track scanning configuration, a scan rate providing contiguous coverage of the earth's surface at 85.5 GHz from an aircraft altitude of 20 km, temperature resolution less than one degree Kelvin on all data channels. The temperature accuracy goal for the sensor is less than two degrees Kelvin for brightness temperatures greater than 200 K. Brightness temperatures of 100 K to 200 K yield an accuracy goal of less than four degrees Kelvin.

An important requirement for the AMPR antenna system is the utilization of the multifrequency feedhorn design used on the SSM/I spaceborne imaging radiometer. Two separate lens antennas were designed with an aperture of 5.3 inches at 19.35, 37.0, and 85.5 GHz and an aperture of 9.7 inches at 10.7 GHz. These dimensions are based on a desire to have equal spatial resolution at 10.7 and 19.35 GHz and to have a total aperture of fifteen inches for both antennas.

The radiometer's temperature resolution for a total power design is given by:

$$\Delta T_{min} = T_0 F_{sys}\left[\frac{1}{\beta\tau}+\left(\frac{\Delta G}{G}\right)^2\right]^{1/2} \qquad (1)$$

where T_0 = ambient temperature = 290 K
$\Delta G/G$ = system gain variation = 0.01%
τ = integration time = 50 ms
β = system bandwidth
F_{sys} = system noise figure

The system bandwidth, noise figure, and the temperature resolution for each channel are given in Table 1. The temperature resolution specifications of 0.22 to 0.38 K for the AMPR channels are significantly better than the performance goal of 1.0 K.

Table 1. AMPR Temperature Resolution Specifications at 10.7, 19.35, 37.0 and 85.5 GHz

Ch (GHz)	β (MHz)	F_{dB}	F_{sys} (Note 1)	ΔT_{min} (Note 2)
10.7	100	3.7	2.344	0.31 K
19.35	240	5.7	3.715	0.33 K
37.0	900	6.2	4.169	0.22 K
85.5	1400	9.2	8.318	0.38 K

Note 1. F_{sys} = system noise factor = $10^{F_{dB}/10}$
Note 2. ΔT_{min} = temperature resolution calculated per equation (1).

An equally important specification for a radiometer used to perform meteorological research is the absolute temperature accuracy. A calibration procedure is required to determine the radiometer's absolute accuracy. The AMPR's temperature accuracy specification depends on the instrument's sensitivity, the physical temperature of the two calibration loads, and the observed brightness temperature of the scene itself. The equation used to determine the instrument's absolute accuracy (T_{acc}) is as follows:

$$T_{acc} = (\Delta T_{min} + \Delta T_x)\left[1+2\left(\frac{T_C - T_S}{T_H - T_C}\right)\right] \qquad (2)$$

where ΔT_{min} = temperature resolution
ΔT_x = thermistor accuracy = 0.1 K
T_C = cold calibration load temp. = 233 K
T_H = hot calibration load temp. = 343 K
T_S = scene brightness temperature

Table 2 summarizes the AMPR's absolute temperature accuracy for each data channel. The calibration load temperatures are based on a temperature controlled hot load (70°C) and an ambient (-40°C) cold load at an altitude of 20 km. Table 2 assumes four different scene temperatures in order to demonstrate variation in temperature accuracy. The largest value shown (1.6 K) is still less than the performance goal level of 2.0 K on all data channels.

Table 2. AMPR Absolute Temperature Accuracy (T_{acc}) At 10.7, 19.35, 37.0, and 85.5 GHz (Note 1)

Ch (GHz)	T_{acc} for T_s =			
	100 K	150 K	200 K	≥ 233 K (Note 2)
10.7	1.4 K	1.1 K	0.6 K	0.41 K
19.35	1.5 K	1.1 K	0.7 K	0.43 K
37.0	1.1 K	0.8 K	0.5 K	0.32 K
85.5	1.6 K	1.2 K	0.8 K	0.48 K

Note 1. Assumes that the scene temperature (T_s) is less than the cold load temperature (233K).

Note 2. The absolute accuracy is ($\Delta T_{min} + \Delta T_s$) for $T_s \geq 233$ K.

System Description

Figure 3 is a block diagram of the AMPR system depicting the multichannel receiver, scanner/calibration, and multifrequency antenna subsystems. A single reflector is used to scan across both antenna lens and a scanner processor is used to control the scan rate, scan angle, and the calibration cycle. A temperature control circuit maintains a constant elevated temperature for the hot calibration load, as well as, control of the instrument's baseplate temperature. The 10.7 GHz antenna consists of the 9.7 inch aperture rexolyte lens which illuminates a corrugated feedhorn. Each receiver's analog output is fed to its own integrate/dump circuit and then to a sample and hold circuit which maintains the output until sampled by the data processor. The data acquisition system is being developed by NASA's Marshall Space Flight Center.

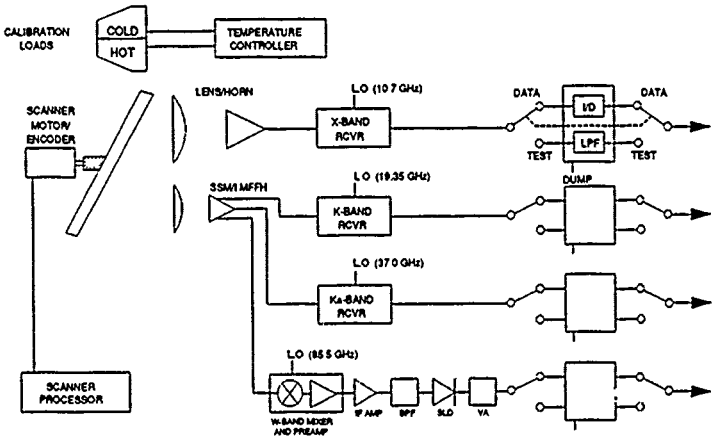


Figure 3. AMPR System Block Diagram.

Figure 4 is a photograph of the AMPR dual antenna configuration with the larger lens used at 10.7 GHz and the smaller (5.3 in. dia.) lens used at 19.35, 37.0, and 85.5 GHz. The smaller lens will illuminate the multifrequency feedhorn (MFFH) which has dual polarization ports at the higher three frequencies. Figure 5 is a photograph of the MFFH which was designed, developed, and fabricated by Microwave Engineering Corporation in North Andover, MA.

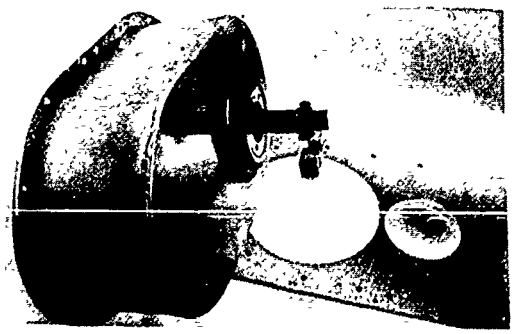


Figure 4. AMPR Dual Lens Antenna

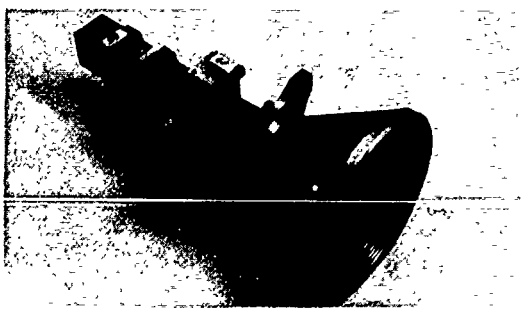


Figure 5. SSM/I Multifrequency Feedhorn

The X-band receiver consists of a solid state downconverter using a dielectrically resonating oscillator (DRO) operating at 10.7 GHz to pump the double-balanced schottky-diode type mixer. The mixer IF output is fed to a single integrated package containing IF amplification, square law detection and low pass filtering stages. Figure 6 is a photograph of the X-band receiver. A low noise amplifier (10.5 to 10.8 GHz bandwidth) is also shown in the photograph. This amplifier will be used between the antenna waveguide-to-coax adaptor and the mixer RF input to reduce the X-band receiver's overall noise figure. Similar receivers at 19.35 GHz and 37.0 GHz will be supplied by SPACEK LABS, Inc.

The W-band receiver is a downconverter consisting of a solid state Gunn Diode Oscillator (GDO) operating at 85.5 GHz used to pump a balanced mixer which uses GaAs beam lead diodes. The mixer IF output is followed by two low noise amplifiers operating over a bandwidth of 0.1 to 1.5 GHz. The mixer's conversion loss is 7.0 dB and the IF amplifier's noise figure is 1.5 dB resulting in a double sideband mixer/amplifier noise figure of 5.5 dB. Figure 7 is a photograph of the W-band receiver shown attached to the MFFH.



Figure 6. AMPR X-Band Receiver

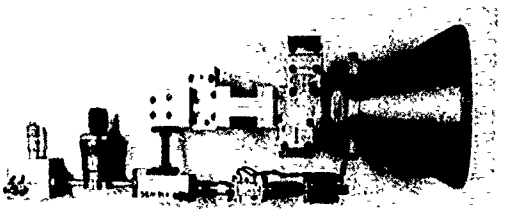


Figure 7. AMPR W-Band Receiver

Summary

High altitude research flights of the AMPR will be onboard NASA's ER-2 aircraft beginning in late 1989. Flights will cover precipitation over land, as well as variable surface effects such as wet ground and snowfields. Precipitation-oriented flights will include the simultaneous measurement of the rain systems by AMPR and a calibrated digital weather radar system. Future plans ca' - using lightning detection instruments together with the AMPR to investigate the relationship between storm ice production and the ele ication of the cloud. The AMPR, along with the lightning detector sensor and other multispectral atmospheric mapping sensors, will provide a complement of instruments suitable for the high-altitude flights over cloud and precipitation systems (6).

The post-flight analysis of the AMPR data will be supported by radiative transfer modelling efforts. These efforts include the transfer of multifrequency microwave radiation through both convective (deep with large ice) and stratiform (shallow with small ice) rain layers. These models will be run with varying surface and atmospheric temperature conditions and, in the convective case, will include the growing, mature, and dissipating stages of storm evolution. The multifrequency response of each case will be tested for separability and for information content on the rain rate at the surface.

Acknowledgements

The authors wish to acknowledge the valuable contributions and support provided by Roy W. Spencer, Hugh Christian, and Mark James of NASA/MSFC, Ken Green of Microwave Engineering Corporation, and by John M. Cotton, Rick W. Hoffner and Don J. Swank with Georgia Tech Research Institute. The continued development of the AMPR as well as the future scientific missions are being supported under NASA/MSFC Contract No. NAS8-37142.

References

1. Spencer, R. W. , et al. , "Heavy thunderstorms observed over land by the Nimbus 7 Scanning Multichannel Microwave Radiometer," Journal of Climate and Applied Meteorology, 1983, Vol. 22, pp. 1041-1046.

2. Savage, R. C., and J. A. Weinman, "Preliminary Calculations of the Upwelling Radiance From Rainclouds at 37 and 19.35 GHz," Bulletin of the American Meteorological Society, 1975, Vol. 56, pp. 1272-1274.

3. Wilheit, T. T., et al., "Microwave Radiometric Observations Near 19.35, 92, and 183 GHz of Precipitation in Tropical Storm Cora," Journal of Applied Meteorology, 1982, Vol. 21, pp. 1137-1145.

4. Spencer, R. W., et al., "Satellite Microwave Radiances Correlated With Radar Rain Rates Over Land," Nature, 1983, Vol. 304, pp. 141-143.

5. Grody, N. C., "Precipitation Monitoring Over Land From Satellites by Microwave Radiometry," Proceedings 1984 International Geoscience and Remote Sensing Symposium (IGARSS), Strasbourg, France.

6. Wang, J. R., et.al., "Profiling Atmospheric Water Vapor By Microwave Radiometry," Journal of Climate and Applied Meteorology, 1983, Vol. 22, pp. 779-788.

AN IMPROVED DESIGN FOR A MICROWAVE RADIOMETRIC PROFILING SYSTEM

J.R. Jordan, J.B. Snider, and T.A. Stermitz
Wave Propagation Laboratory, NOAA/ERL
325 Broadway, R/E/WP5
Boulder, CO 80303

Abstract

In 1977, the Wave Propagation Laboratory (WPL) developed a millimeter-wave radiometric system to measure atmospheric water vapor, and cloud liquid. Since that time, WPL has constructed or purchased six more radiometric systems utilizing frequencies ranging from 20.6 GHz to 90 GHz. All of these radiometers use an Automatic Gain Control (AGC)-corrected Dicke design which gives an instrument with the long-term stability and accuracy necessary for unattended operation. However, the electronics of these receivers has been virtually unchanged since 1977. A new prototype 53.85-GHz single-frequency radiometer incorporates the best features of the old design and improvements derived from 10 years of operational experience. The sensitivity of the 53.85-GHz prototype is improved by 70% compared with the original radiometer.

1. Introduction

In 1977, the Wave Propagation Laboratory (WPL) developed a millimeter-wave radiometric system to measure atmospheric water vapor and cloud liquid. Since that time, WPL has constructed or purchased six more radiometric systems utilizing frequencies ranging from 20.6 to 90 GHz. A six-frequency temperature profiler is operated routinely at Stapleton International Airport in Denver, Colorado (Hogg et al. 1983a). A steerable three-frequency system and several dual-frequency systems are used to measure water vapor and cloud liquid at various locations (Hogg et al. 1983b, Westwater and Snider 1987). All of these radiometers were constructed using an Automatic Gain Control (AGC) corrected Dicke design (Guiraud et al., 1979). The receivers use a three-way switching sequence to select either two different temperature waveguide terminations or the sky. The two terminations are used to remove receiver noise and gain fluctuations thus the instrument has the long-term stability and accuracy necessary for unattended operation. However, the electronics of these receivers has been virtually unchanged since the original radiometer was constructed in 1977. The old design used analog demodulation which is subject to small drift errors requiring hourly correction. This produces hourly gaps in the measurement, complicating any time series analysis of the data. In order to obtain continuous sampling and to modernize the existing design, a prototype 53.85-GHz single-frequency radiometer was constructed which incorporates the best features of the old design and includes improvements based on 10 years of operational experience. The proven AGC-Dicke technique is retained but a microprocessor is embedded to demodulate the video signals digitally and to control the instrument. Modern low noise components are used throughout to increase system sensitivity and to improve temperature stability. Better thermal control in the prototype has reduced temperature fluctuation errors in the data.

2. Factors Influencing Design

When the decision was made to construct the prototype radiometer, several factors were considered as constraints given the operational use of these systems. WPL run radiometers continuously at unattended remote locations. Therefore, the instruments require calibration stability for several months and excellent reliability. It was decided to retain the AGC corrected Dicke design, which has been successfully used in WPL for a decade. Since the current offset parabolic antenna design has the same beamwidth for different frequency channels and low sidelobe levels (Hogg et al., 1979), we chose to employ the same design in the updated radiometer. Frequency selection was restricted to one of the four channels operating at Stapleton International Airport, Denver, Colorado. Thus direct comparisons with the current hardware could be used to measure improvement. 53.85-GHz was chosen as the prototype frequency since it is off the center of the 50-GHz oxygen absorption line but not greatly affected by water vapor.

3. Design Improvements

A block diagram of the prototype radiometer is shown in Figure 1. The focus of the design is the triple-junction switched ferrite circulator. By controlling the state of the first two junctions, the microprocessor can select the sky signal from the antenna, a reference termination controlled to 318 K, or an AGC termination cooled to 250 K. The third junction in the switch is an isolator so the reflection coefficient seen by the mixer is constant. The polarization coupler allows two frequency channels to share the same antenna. Losses and reflections in the path between the sky and switch, and between cool load and switch, are accounted for by two calibration factors derived from radiosonde data. The noise power from the sky, reference termination, and cool termination are switched sequentially, integrated, and converted to digital format to be processed in the microprocessor. Brightness temperature is calculated using the integrated noise powers, physical temperature of various system components, and calibration factors. This calibrated brightness is then sent to a host computer. Design changes were implemented to improve the sensitivity, stability, and reliability of the prototype radiometer.

Sensitivity Improvements

Radiometer sensitivity (ΔT) is defined as the minimum detectable change in the radiometric antenna temperature of the observed scene (Ulaby et al., 1981). The sensitivity of an AGC-corrected Dicke radiometer is given in Eq. (1).

$$\Delta T = \sqrt{\frac{3}{B}\left[\frac{(T_r - T_a')^2[(T_{agc} + T_{rec})^2 + (T_r + T_{rec})^2]}{(T_{agc} - T_r)^2 \tau_{agc}} + \frac{(T_a' + T_{rec})^2 + (T_r + T_{rec})^2}{\tau}\right]} \qquad (1)$$

where

ΔT = Radiometric sensitivty (K)
B = IF bandwidth (Hz)
T_a' = Antenna brightness temperature (K)
T_r = Reference temperature (K)
T_{agc} = ACG termination temperature (K)
T_{rec} = Receiver noise temperature (K)
τ_{agc} = Averaging time for the AGC (s)
τ = Sky averaging time (s).

The equation is derived in a manner similar to Hach (1968) for a three-stage switching sequence. However, the equations are not identical due to differences in the switch sequence. Sensitivity is a function of the receiver noise temperature, averaging time, and intermediate frequency (IF) bandwidth. The total sensitivity improvement was accomplished by optimizing all three parameters. The receiver noise temperature is dominated by the mixer noise. The mixer/IF was purchased with the lowest double-sideband noise figure available at the time. Recent improvements in mixer design have reduced the noise figure by almost a decibel, which will improve sensitivity even further in future radiometers. The loss in the microwave components before the mixer also contribute to the receiver noise temperature. To keep this loss at a minimum, the three-junction switch was constructed with a configuration different from than the previous design, reducing the total loss to 0.9 dB from 1.8 dB. Waveguide lengths where also kept at a minimum to reduce their noise contribution. These changes reduced the prototype noise temperature to 845 K from the 1950 K of the original radiometer.

The second parameter for improving sensitivity is averaging time, τ. It is well known that ΔT varies inversely as the square root of the averaging time. However, Eq. (1) shows that the sensitivity is also a function of the ratio of the sky averaging time to the AGC averaging time. Figure 2 is a plot of Eq. (1) for two different ratios and the sky averaging time of 1 second. The plot displays the improvement in sensitivity obtained by averaging the AGC measurement 10 times longer than the sky measurement. The microprocessor applies an exponential averaging algorithm to smooth the measurements of the reference and cool terminations, averaging the AGC longer than the sky.

The last parameter to optimize for maximum sensitivity is the IF bandwidth. Equation (1) also shows that sensitivity varies as the square root of the bandwidth. Thus, use of the largest bandwidth possible is desirable. The radiometer is constructed as a double sideband receiver, which means that the power in the two sidebands is averaged to estimate the power at the local oscillator frequency. This requires that the sky's brightness temperature curve have a constant curvature as the temperature changes with frequency. Oxygen absorption near 60 GHz is a complex of many pressure broadened lines that merge into a generally smooth feature but has complex line structure. Local oscillator frequencies are chosen in the center of smooth regions of the

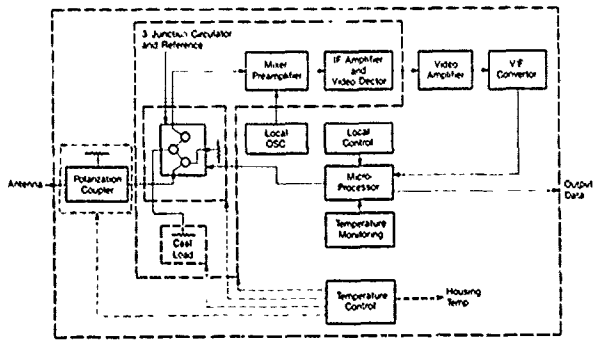


Fig. 1. Block diagram of the prototype radiometer

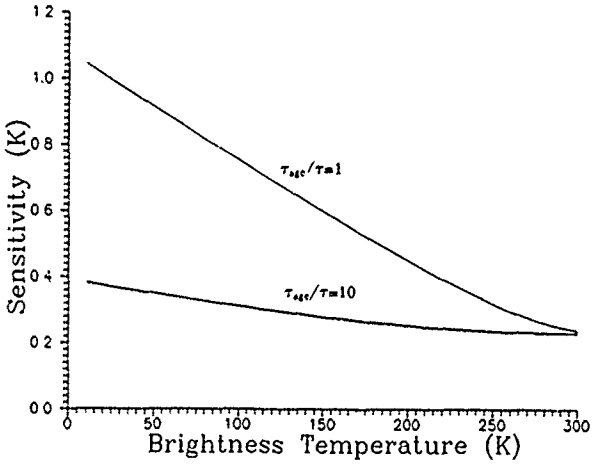


Fig. 2. Plot of Eq. (1) for two averaging ratios.

brightness curve, and the IF bandwidth is selected to span this smooth region. Simulations of the line shape showed that the IF bandwidth could be increased from 100 MHz to 150 MHz without affecting the double sideband accuracy.

The result of our changes is a reduction of ΔT to 0.22 K for a 1-second averaging time where the current radiometer has a ΔT of 0.76 K for the same averaging time. This represents about a 70% improvement in sensitivity compared with the current radiometer.

Stability Improvements

WPL operates radiometric systems continuously at unattended remote sites. This requires that the systems maintain calibration stability to better than 0.5 K for several months time. The AGC corrected Dicke design has been shown to fill this requirement (Snider et al. 1988) and has been retained in the prototype radiometer. Historically, sensitivity has been defined in terms of the system noise temperature since that was the dominant source of uncertainty. However, the advent of low-noise components has reduced that part of the total receiver fluctuations to the level of the systems short-term stability. Short-term is used here to mean periods of 24 hours or less. Therefore, sensitivity estimates must consider both receiver and short-term fluctuations. Prototype short-term stability has been improved by implementing a digital sampling scheme and enhancing the temperature control circuitry.

The sequential switching between three different noise sources at the input of mixer can be considered as a form of modulation. The composite signal must be sychronously demodulated at the output of the receiver to recover the input signals. The current design uses an analog technique for demodulation and the Dicke subtraction. These analog demodulation boards must have their gain and offset measured periodically. Gain and offset are assumed to remain constant between measurements. Besides introducing gaps in the data which complicate time series analyses, short-term drift may affect the data. The prototype design incorporates a digital sampling technique so the demodulation, Dicke subtraction, and AGC functions are accomplished in a microprocessor. As a result, the analog signal path is common to all three measurements, and any gain or offset drift is corrected for when brightness temperature is calculated. The digital sampling is accomplished by applying the amplified video signal to a linear voltage-to-frequency converter. The resulting pulses are counted for 110 ms, and the total count is read into the microprocessor before the switch to the next signal in sequence. This process integrates the noisy signal and gives a good estimate of the mean. Besides being simple, this technique does not require an antialiasing filter, which would increase the settling time when the signal is switched and decrease the percentage of integration time.

An important part of the short-term fluctuations is the system temperature stability. WPL has always operated the radiometers in a benign environment and provided temperature control of all important components. The prototype design also controls the temperature of the polarization coupler and local oscillators. Loss in the coupler and associated waveguide not only attenuates the incoming power but also adds noise due to thermal emission. Correcting for both of these effects and for reflections requires knowing the loss and reflection coefficients to a high degree of accuracy. WPL combines the effects into a single calibration factor derived from radiosonde data. Improved temperature control of the polarization coupler allows the calibration factor to be treated as a bias without any temperature dependence. Local oscillator frequency stability is im-

portant in the region near the 60-GHz oxygen absorption line since the slope of the brightness-temperature-versus-frequency curve is quite steep. The frequency of the Gunn oscillators used by WPL decreases with increasing temperature, causing the measured brightness to decrease as the temperature of the oscillator increases. This effect is opposite to the temperature dependence of the polarization coupler and is difficult to account for when brightness temperatures are calculated. Therefore, the temperature stability of the local oscillators has been significantly improved in the prototype.

Reliability Improvements

Reliability has always been a consideration in the construction of WPL's radiometric systems. Therefore, in the prototype it was decided to change the design of components that have caused trouble: the AGC termination, the computer, and the video amplifier.

The current radiometer design holds the AGC termination at 418 K, hot enough to cause corrosion failures in the heater element after extended periods of time. The material used for the waveguide termination was found degraded, causing the termination to crumble changing emissivity and reflection coefficient. Although its material has been changed to a higher temperature specification, there has been insufficient time to know the degree of improvement. It was therefore decided to cool the AGC termination (Fig. 3), instead of heating it. The termination is made of standard lossy waveguide material that is cut so the point lies against the edge of the waveguide to reduce thermal gradients. A sensor is embedded in the material so its temperature can be monitored. The cooled section of waveguide is a gold-plated copper block that is refrigerated by a thermoelectric cooler (Peltier element) capable of 70°C difference between its hot and cold surfaces. The monitor sensor is the feedback element in a servo circuit that holds the temperature of the termination at 250 K with 0.1 K peak-to-peak stability. The thin-wall stainless steel waveguide reduces heat flow but is gold plated for low electrical loss. This waveguide is filled with a low-loss foam (dielectric constant of 1.03) to prevent water from condensing on the termination. Silica gel placed behind the termination absorbs any water that is inside the waveguide. Another benefit of cooling the AGC termination is that the sky brightness at 53.85 GHz is close to the 250 K AGC temperature. This places fewer demands on the linearity of the radiometer than the 418 K termination that is currently used.

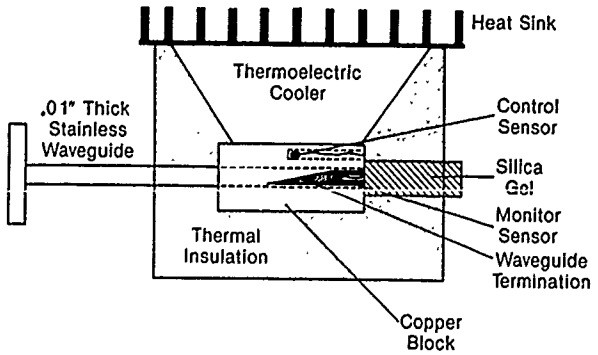


Fig. 3. Block diagram of cooled AGC termination.

The radiometric systems operated by WPL contain a minicomputer to collect analog voltages and compute the final brightness temperature. These minicomputers dating from the 1970's have been the most unreliable part of the system. Therefore, a CMOS microprocessor was incorporated in the prototype radiometer to perform all the functions previously performed by the minicomputer but with fewer components and more reliable packaging. This eliminated the need for an external computer with external analog cabling resulting in improvement in reliability.

The current radiometric design requires periodic adjustment of the video amplifier gain and offset to remain within the range of the analog-to-digital converter. If the output of the video amplifier drifts, voltage peaks in the signal can get clipped causing the output to be nonlinear. To correct this problem, the prototype video amplifier was designed with a microprocessor-controlled offset and gain adjustment. Whenever the output exceeds preset limits, the microprocessor realigns the amplifier automatically. The current analog radiometer also requires adjustment of all analog circuits during construction or component replacement. The prototype design requires no adjustments, which simplifies construction and operation.

4. Plans

The six-frequency radiometric system at Stapleton International Airport, Denver Co., is the only temperature profiler currently operated by WPL. The prototype system described here is the initial step in the creation of a second generation radiometric temperature profiler. Components have been purchased to add a 56.02 GHz channel to the 53.85 GHz prototype. In addition, a second dual channel radiometer with frequencies at 23.85 GHz and 31.65 GHz is now under construction which will measure water vapor and cloud liquid. These two radiometers will be installed into a movable building to form a four frequency temperature profiler

(Schroeder et al. 1989). The initial application will be to evaluate the accuracy of temperature profiles retrieved from the four-channel system. In the future, two more frequencies will be added to the radiometric system between 50 GHz and 60 GHz making it a complete six-channel temperature profiler.

5. Acknowledgment

Design and construction of the prototype radiometer is partially funded by the office of Naval Research University Research Initiative and Pennsylvania State University.

References

Guiraud, F. O.,J. Howard and D. C. Hogg, 1979: A dual-channel microwave radiometer for measurement of precipitable water vapor and liquid. IEEE Trans. Geosci. Electron., GE-17, 129–136.

Hatch, Johann-Peter (1968), A very sensitive airborne microwave radiometer using two reference temperatures. IEEE Trans. Micro. Theory and Tech., MTT-16,9, pp. 629–636.

Hogg, D. C., F. O. Guiraud, J. Howard, A. C. Newell, D. P. Kremer and A. G. Repjar, 1979: An antenna for dual-wavelength radiometry at 21 and 32 GHz. IEEE Trans. Antennas Propag., AP-27, 764–711.

Hogg, D. C., M. T. Decker, F. O. Guiraud, K. B. Earnshaw, D. A. Merritt, K. P. Moran, W. B. Sweezy, R. G. Strauch, E. R. Westwater and C. G. Little, 1983a: An automatic profiler of the temperature, wind, and humidity in the troposphere. J. Climate Appl. Meteor., 22, 807–831.

Hogg, D. C., F. O. Guiraud, J. B. Snider, M. T. Decker and E. R. Westwater, 1983b: A steerable dual-channel microwave radiometer for measurement of water vapor and liquid in the troposphere. J. Climate Appl. Meteor., 22, 789–806.

Schroeder, J. A., J. R. Jordan and M. T. Decker, 1989: Design considerations for a network of thermodynamic profilers. J. Atmos. Oceanic Technol. (in press).

Snider, J. B., M. D. Jacobson, M. J. Falls and J. R. Jordan, 1988: Comparison of temperature profiles measured by three colocated microwave radiometers and radiosonde. NOAA Tech. Memo. ERL WPL-151, 39 pp.

Ulaby, F. T., R. K. Moore and A. K. Fung, 1981: Microwave Remote Sensing, vol. I. Reading, MA: Addison-Wesley.

Westwater, E. R. and J. B. Snider, 1987: Microwave radiometer facilities at the wave propagation laboratory. Proceedings of NAPEX XI, Virginia Polytechnic Institute and State University, Blacksburg, Virginia.

CONVERTING DIGITAL PASSIVE MICROWAVE RADIANCES TO KELVIN UNITS OF BRIGHTNESS TEMPERATURE

L. DENNIS FARMER, DUANE T. EPPLER and ALAN W. LOHANICK

NORDA Polar Oceanography Branch Office, 72 Lyme Road, Hanover, NH 03755-1290

BACKGROUND

NORDA, in cooperation with the Naval Weapons Center (NWC) at China Lake, California, has been collecting passive microwave imagery with the Ka-band Radiometric Mapping System (KRMS) since 1980. With the exception of data collected in 1983, none of these data have been converted to brightness temperatures. The 1983 data were converted to brightness temperatures using an engineering method which required measuring gains and losses within the system and then scaling the resultant radiances to measured surface values for open water and first-year sea ice. Here we propose alternative methods. The KRMS system has a measured internal reference load which is related to brightness temperature and can be used as a warm reference temperature. A graph was produced at NWC by placing a thermocouple at the reference load, varying the reference load voltage and recording the resultant equivalent brightness temperature. This provides a warm reference point. Another possible method is to use a local ambient temperature at the surface, adjusted for the highest anticipated emissivity for sea ice (0.94), as the warm reference point. The cool reference point that is used for both methods is an assumed brightness temperature for calm open water at nadir of 135 K. This study explores the applicability of both methods.

INTRODUCTION

Data from three KRMS missons were used to evaluate the conversion procedures: March 1983 (Beaufort Sea), March 1987 (Labrador Sea), and March 1988 (Bering, Beaufort, and Chukchi Seas). The conversion procedures presented make several assumptions.

First, we asume that the lowest apparent brightness temperature in a scene is greater than or equal to that for smooth open water at nadir, approximately 135 kelvins (Hollinger 1973, Hollinger and Lo, 1984). This open water value is a surface measurement of calm sea water at nadir and does not take into account atmospheric effects.

Second, we assume that the highest brightness temperature in a scene can be estimated using either a known surface temperature (Tt) or by measuring the reference load voltage and computing the equivalent brightness temperature.

If these assumptions hold, then the range of digital radiance values present in a KRMS data set can be linearly scaled to brightness temperatures that fall within this range.

CONVERSION METHODS

The warm tie-point is determined by using either a local ambient temperature, adjusted for the highest anticipated emissivity, or the reference load equivalent sensor temperature.

The first method uses the ambient surface air temperature adjusted by the emissivity of first-year ice for the warm tie-point. Since Tb = E * Tt relates radiometric brightness temperature (Tb) of a body with emissivity (E) to its physical temperature (Tt) , and E ranges from 0.0 to 1.0, the radiometric temperature (Tb) of a surface should not exceed its physical temperature (Tt). The emissivity of any natural surface is less than 1.0, so the highest anticipated radiometric temperature in a scene is necessarily less than its physical temperature. Since first-year sea ice and some forms of young ice have the highest emissivity (0.94), they will display the highest radiometric temperatures observed in KRMS images of sea ice.

The second method uses the internal reference load and its equivalent brightness temperature for the warm tie-point. KRMS data are digitized across a twenty volt range, with the highest signal level set at +10 volts and the lowest set at -10 volts. When analog data are digitized, gain and offset applied to the analog signal are adjusted such that the reference load load voltage corresponds to a digital value 0. By deriving the brightness temperature that is equivalent to the reference load and a digital value of about 0. a warm tie-point is established. The cool tie-point of 135 K is set at a digital value of 2000, and the data scaled linearly between these tie-points.

Figure 1 provides a comparison of the engineering conversion used in 1983 and the corresponding conversion using a local adjusted ambient temperature and open water. An example of brightness temperature conversions obtained using both the reference load and the adjusted ambient

temperature for the same data set are shown as Figure 2.

The average slope of the conversion equation for the ambient temperature method is 0.0584, with a standard deviation of 0.00549. The average slope for the sensor temperature (reference load) method is 0.0728, with a standard deviation of 0.0027. These values are based on nine coincident samples and are shown graphically as Figure 3.

CONCLUSIONS

Both methods presented produce ranges of brightness temperatures that appear to be reasonable for the types of surfaces observed. The sensor temperature method produces values that compare more favorably with data collected by other passive microwave radiometers. The sensor temperature/water conversion method is more repeatable and is the recommended method to use with existing KRMS data. The ambient temperature method is not suited for use with existing data, due to the difficultyin obtaining reliable ambient temperature measurements along flight tracks.

The open water brightness temperature of 135 kelvin is an assumed value for a calm sea. The distance between the sensor and the surface are known, however, atmospheric contributions are ignored, as are the effects of surface roughness. Thus, this value is a possible source of error for both methods.

The variability of results obtained clearly support the need for system calibration for the KRMS and other passive radiometers.

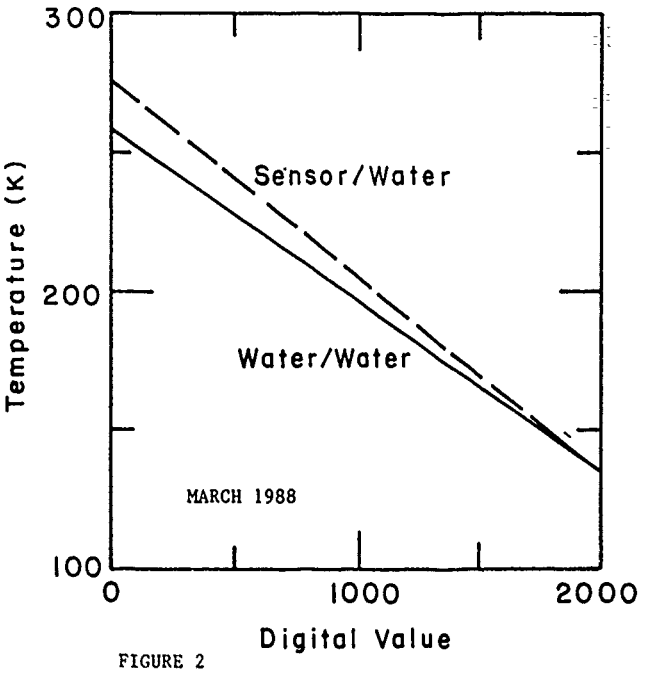


FIGURE 2

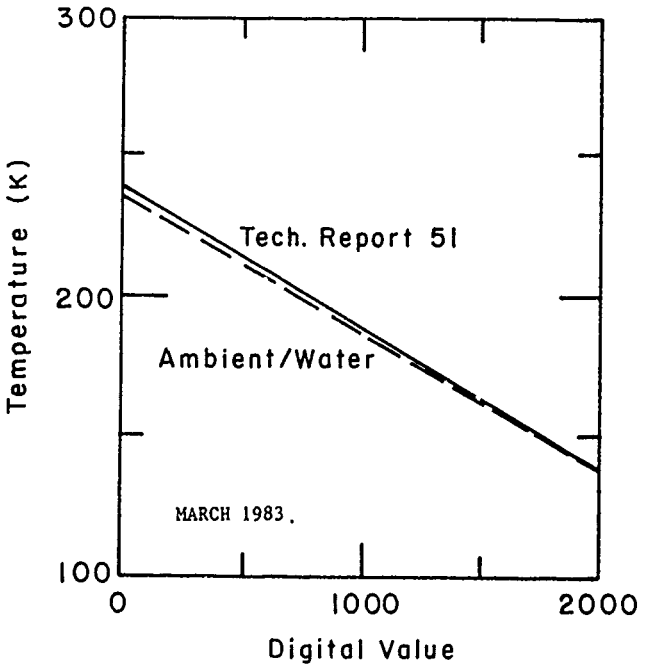


FIGURE 1

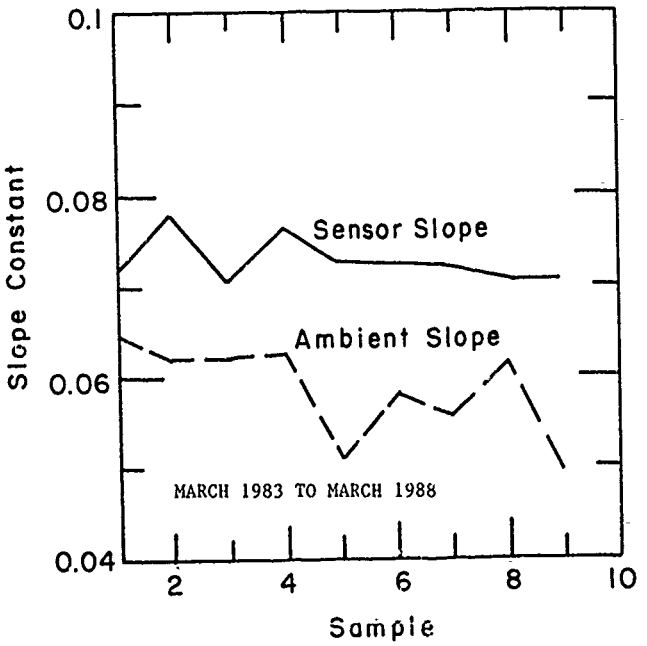


FIGURE 3

ADVANCED TECHNOLOGY PROCESSES FOR SATELLITE INSTRUMENTS

John P. Hackett, James B. Dorey, Roger S. H. Tong and Attia El-Shakweer

COM DEV LTD.

Cambridge, Ontario

Canada

Abstract

Space-borne scientific instruments usually have very demanding requirements and stringent constraints. This will often lead to a hierarchy of desirable features for the instrument design phase. Specialized design and fabrication procedures (including materials, processes and components selection) are often required to achieve a fully compliant instrument. This paper discusses some of these procedures with examples cited from a recently completed satellite mass spectrometer.

Topics that are discussed include instrument manufacturing processes for space qualification, such as; laser welding of thin foils, electron beam welding, photo-chemical machining, electrical-discharge machining, and electroforming. Design and manufacturing practices for multilayer circuits with radiation hardened electronics are also discussed. An example of mechanical and thermal analysis for a satellite instrument is presented.

This paper is intended to assist the instrument engineer in selection of specialized materials, processes and components. Familiarity with the performance of these processes and knowledge of associated problems will promote high performance instruments, with fewer implementation difficulties.

Keywords: Satellite, Instrument, Spectrometer, Materials and Processes

Introduction

Remote sensing and space science instruments for space-borne applications demand high performance and high reliability. Satellite instruments (versus shuttle or sounding rocket) are the most demanding due to the extended mission life and complete lack of serviceability.

A summary of general characteristics for satellite instruments is presented in Table 1.

Table 1: General Characteristics For Satellite Instruments

• High Reliability
• Low Mass
• Compact Size
• Low Power (Consumption and Heat Flow)
• Integral Structural and Thermal Design
• Compliance with Environmental Parameters
• Radiation Hardened Electronics
• Acceptable Material Outgassing
• Minimal Costs
• Minimal Risk (Schedule and Technology)
• Reasonable Data Rates

To meet these general requirements and instrument specific operational parameters, it is often necessary for the instrument engineer to utilize advanced design and fabrication procedures.

This paper summarizes several advanced manufacturing and design techniques that have been used on recent projects at COM DEV. Most of the examples are taken from a satellite-borne ion mass spectrometer called SMS [1].

Machining Techniques

The more conventional machining techniques include CNC milling, turning using lathe, grinding etc. These techniques have inherent limits with respect to dimensional accuracy, stress imparted on the work piece, and the minimum size of the part. For various materials and workpiece sizes there are alternative and more advanced machining processes.

Electrical-Discharge Machining

For metallic materials, electrical-discharge machining (EDM) can be used. There are two types of EDM processes, wire EDM and sink EDM. Wire EDM is conceptually similar to "bandsawing" where the blade is actually a 10 mil diameter wire (usually brass). The wire is at a moderate voltage (30-100 volts) from the grounded workpiece and a discharge (arcing) occurs causing the workpiece to erode in a precisely controlled manner. As a result, extremely tight tolerances can be

maintained while imparting virtually zero stress to the workpiece. An example of wire EDM is shown in Figure 1.

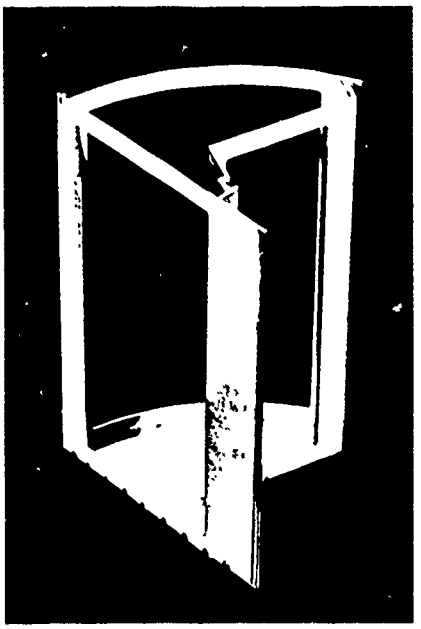


Figure 1: An example of EDM Wire Fabrication (The plate was welded to the bottom after the frame was machined)

EDM sink is a similar process except the wire is replaced by a custom electrode (usually copper) which makes an "imprint" into the workpiece.

For both types of EDM, the feed is computer controlled. The surface finish can be controlled by using the appropriate feed rate and current settings. The EDM process can be utilized to provide parts with low mass, high precision, high reliability and compact size.

Photo-Chemical Machining

Photo-chemical machining (PCM) is often used for thin parts which are machined from a sheet or foil. The material is usually a metal or metal on dielectric film. PCM is most often single sided etching, where the artwork image is photolithographically produced on one side of the sheet, and the back of the sheet is totally masked. For more demanding tolerances and small detail size, double sided etching can be used for metal parts. In this case two artworks are produced and aligned, then the workpiece sheet is inserted between the two artworks so that the photolithographic image is produced on both sides of the sheet. The "etch back" is reduced by a factor of two with double-sided etching. The limits on detail size are a function of material thickness for "etch away" areas, and are limited by the fragility of the part for "flat" areas.

The PCM process can be very cost effective for large quantities. PCM also provides the highest reliability in delicate part manufacturing.

Figure 2 shows a grid developed for the SMS project. The grid mesh is constructed out of 1 mil molybdenum foil, the strand width is 1.6 mil ± .5 mil.

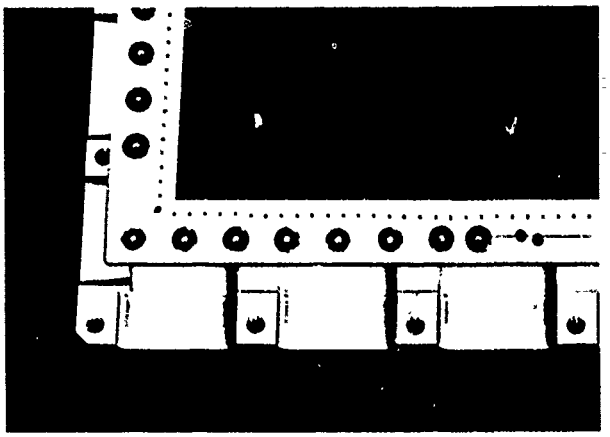


Figure 2: An Example of PCM Fabrication and Laser Welding of Thin Foils

Electroforming

Parts which have very intricate detail on irregular or inaccessible surfaces, such as millimeter wave corrugated horns, may be best manufactured by using electroforming techniques. The material selection is limited to; copper, nickel, silver or gold. Copper is the most common due to its low cost and non-magnetic property. Nickel is much stronger than copper and is not much more expensive, but the inherent magnetic property of nickel make it not suitable for some applications. Silver is sometimes used for fabricating very small waveguide sections for the millimeter wave band. Silver has the highest conductivity of all the pure metals. Gold is much more tarnish resistant than silver, but gold's conductivity is only 2/3 that of silver.

Welding Techniques

Most conventional welding techniques are not suitable for precisely toleranced parts or thin-walled low mass structures due to the high thermal distortion caused by the welding. However, electron beam welding (EBW) and laser welding (LW) can often be used in these situations since the heating is extremely localized and well controlled.

Electron Beam Welding

Although EBW can be performed in the atmosphere [2], the best results are obtained with vacuum systems. The part is mounted in the vacuum chamber and electrically grounded to the chamber housing. After a sufficient vacuum is achieved ($\approx 10^{-3}$ torr) a well focused beam of high energy electrons ($\approx$60 keV) is directed onto the workpiece. The depth of weld is controlled by the welding current (typically tens of milliamps) and the welding speed ($\approx$ 2-50 inches/min).

The EBW process is often used on aluminum alloy 6061 for space applications. A filler material must be used due to the high degree of brittleness of the 6061 alloy. The EBW military specification MIL-W-46132 states that aluminum alloy 4047 or 4043 must be used for welding alloy 6061. We have found the 4043 is more ductile than 4047 and the resulting welds are much less prone to cracking.

The SMS housing was welded together using EBW butt joint with backplate (4043 filler shim). The wall

sections were 0.063" thick 6061 aluminum that had previously been nickel plated. Tests performed on sample pieces demonstrated a 30,000 psi tensile strength which was more than adequate for the application.

Laser Welding

The achievable performance in laser welding will vary greatly between laser systems, however the following general statements apply. LW can provide spot or continuous welds. The depth of weld achievable is less than that of EBW. LW can be much faster than EBW since the workpiece is not contained in a vacuum. Aluminum and copper are difficult to work with for most laser systems.

For the SMS project it was necessary to weld 1 mil thick molybdenum foils to titanium frames. This was a very challenging problem since if even a minute gap existed between the foil and the frame, the foil would be evaporated at the laser beam focus. In addition, the grids were required to be very flat when welded. After some development we found that excellent welds could be obtained using a CO_2 laser and a modified process. To ensure that the grids (foils) were flat and in the proper position when welded, we mounted the foils on plexiglass and aligned the frames on the foils. These parts were then welded through the plexiglass, this allowed us to ensure that the foil was in intimate contact with the frame at the time of welding. An sample of this type of weld is shown in Figure 2. This approach also allowed us to weld foils onto curved (cylindrical) surfaces by first molding the plexiglass to the desired shape.

Surface Finishes

Metal components require a surface finish to enhance performance and/or to prevent uncontrolled surface oxidation.

The most common metal used in satellite instruments is aluminum alloy 6061. In addition to electroplating, this metal can be finished by chemical conversion techniques such as chromate conversion or by electrochemical techniques such as anodizing.

Most metals can be electroplated with a variety of metallic finishes. However certain metals are more difficult to electroplate, and intermediate adhesion layers are often required. Electrolytic plating can cause a heavy build up of the deposited material at sharp edges and corners due to increased current densities. Electroplaters that are experienced with high precision components can minimize the plating build up by using proper anode geometry.

Common electroplated surface finishes include silver nickel, gold and rhodium. Silver is highly conductive and solderable but suffers from tarnishing. Gold is tarnish resistant and conductive, but it is difficult for soldering since special care must be taken to prevent gold migration which could cause solder embrittlement. Nickel has low conductivity and is usually magnetic, but it can provide an economic and attractively passivated surface for noncritical components. Rhodium has moderate conductivity, high wear resistance, excellent soldering properties and an excellent (low) secondary electron emission coefficient. A rhodium surface finish (or flash) will improve RF component performance against multipaction breakdown.

Mechanical and Thermal Analysis

The mechanical and thermal analysis that is required for satellite-borne scientific instruments is a very essential stage in the instrument design cycle. This analysis of the design will ensure proper equipment operation, sound thermal performance and adequate structural integrity. The analysis should be performed prior to the fabrication of each model (engineering model, qualification model and flight model). Hence all design modifications that arise would be modeled prior to their implementation.

For simple shapes, the mechanical/thermal analysis can be performed adequately by using standard formulas for each subsection of the instrument. However for more precise results or for more complicated structures the analysis can be performed with the aid of a computer model. As an example, the SMS structure was analyzed using COM DEV's in-house Finite Element Modeling (FEM) capabilities. The modeling utilized a large scale Finite Element Analysis (FEA) computer program called NASTRAN, and a sophisticated pre and post-processing program called PATRAN 2. The FEA model used more than 400 elements to represent the entire structure. These elements consisted of plates, beams, springs and concentrated masses. The model was then solved for eigenvalues and frequency response to determine the critically stressed areas.

Of primary interest in this analysis was the structural response of the upper housing plate to the applied environmental loading. The FEM analysis indicated a low frequency resonance in the upper housing plate which could have lead to large deflections and accordingly high stresses.

A design change was then implemented by stiffening the plate. This exercise resulted in a highly reliable and structurally sound housing, without the added costs and schedule delays of fabricating another unit.

Instrument Electronics

Instrument electronics are governed primarily by minimal mass, power and volume constraints. Very often multilayer circuit boards and radiation hardened devices are used.

Multilayer Boards

Most space science instruments have a pre-defined shape for the electronics cavity determined by the science and minimal mass objectives. This imposes a stringent requirement on the packaging of the electronics. To maximize the area utilization of the printed circuit board one can use multilayer PCBs where components can be packed side by side with a spacing of 0.1 inches, this allows 80% utilization of area for components. The other advantage of using multilayer PCBs is that it allows separate ground and power supply planes for EMI shielding and low impedance power distribution, which is not possible on conventional double sided PCBs.

For satellite instrument applications, the preferred material used for multilayer boards is polyimide-glass. The polyimide-glass material possesses high heat resistance, good mechanical properties at high temperature, stable electrical values, superior chemical resistance and most importantly excellent z-axis dimensional stability. The above properties makes it extremely attractive for harsh environments with high

temperatures gradients. Figure 3 shows the coefficient of thermal expansion for different materials. The polyimide material matches the coefficient of thermal expansion for copper up to a temperature of 240°C. Therefore polyimide-glass minimizes the stress developed in the plated through holes during soldering (=360°C) and during the temperature cycling.

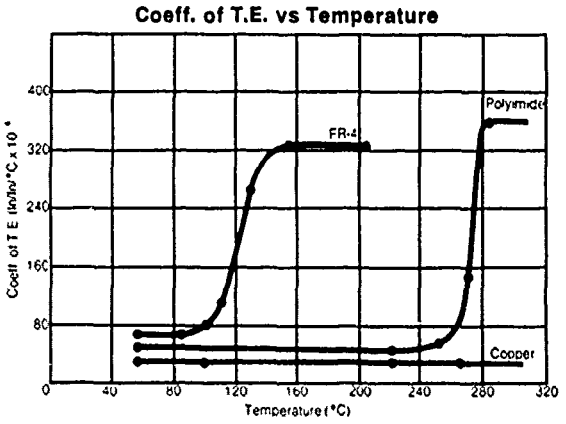


Figure 3: Thermal Coefficient of Expansion for Various PCB Materials

For multilayer board design and manufacturing it is necessary to pay special attention to the following:

1. layer to layer registration should be specified on the master drawing rather than quoting MIL-P-55110, since MIL-P-55110 allows for a maximum registration error of 14 mil and minimum angular ring of 2 mil. The worst case angular ring of 2 mil does not allow an adequate margin of safety for thermal cycling or soldering stresses.
2. microsectioning should be performed aside from the manufacturers supplied sections which are normally done on larger pad sizes. The sectioning chosen should be on the smallest size holes used. The drill diameter and plating thickness should be measured to ensure that the manufacture has used the proper sized drills rather than increasing/decreasing the plating thickness to meet the dimension of the plated through hole.

Special precautions are also required in the assembly of polyimide multilayer boards. Prior to assembly the board must be baked at 100°C for a minimum of 1 hour to remove moisture absorbed by the material (polyimide has a water absorption of 30% by weight). During the assembly process the board should be maintained at ≈100°C using a hot plate and should be soldered by a special type of solder iron with a large thermal mass to minimize thermal gradients during soldering which will minimize the soldering time. This is especially important for multilayer PCB with power and ground planes, since these planes tend to distribute heat away from the plated through hole which prevents the solder from flowing nicely.

Radiation Hardened Electronics

Space-borne instruments will inevitably be exposed to some level of radiation due to cosmics rays, solar wind and the Van Allen radiation belts. The radiation dosage will depend on the orbit parameters, the mission life and the amount of shielding provided by the satellite.

One of the primary concerns about the radiation environment is its impact on the lifetime of semiconductor devices. A special series of semiconductor devices that were developed for high radiation environments, called "radiation-hardened" devices, are available off the shelf.

Rad hard devices have been specially designed and use specialized processing to ensure that the components will meet all of the data sheet parameters after receiving the specified radiation dose. Due to the increasing requirement for rad hard components military standard MIL-M-38510 has included a Radiation Hardness Assurance (RHA) designator for all JAN microcircuit part numbers as shown in Table 2 [3].

For a very long design life, annealing of the semiconductor devices should be considered.

Table 2: MIL-M-38510 Radiation Hardness Assurance Designator

DESIGNATOR	TOTAL DOSE RAD (Si)	NEUTRON FLUENCE LEVEL (n/cm²)
/	No RHA	No RHA
M	3000	2×10^{12}
D	10^4	2×10^{12}
R	10^5	$>10^{12}$
H	10^6	$>10^{12}$

Though most manufacturers of rad hard devices claim that their devices are functionally compatible with commercial or even MIL-STD devices, there are minor differences which might affect the overall circuit performance if not considered at an early stage of the design. Typical parameters which vary are;

- input capacitances of radiation hardened devices are usually 20% higher.
- the input high voltage has a much smaller range than commercial devices (e.g. V_{DD}-0.5V for a logic high on an 80C85RH from Harris).
- most devices from Harris have regenerative latches at their output pins which can cause problems in some cases when interfaced to another family of devices. In other cases some manufacturers over specify their worst case parameter, which could create a problem when performing worst case analysis, and this causes the system to be over designed (e.g. the worst case timing parameters for RCA CD4000 are twice their typical parameters).

As the radiation hardened processing technology matures, some of these problems will become more visible to circuit designers and some of the above problems could be eliminated.

Assembly Techniques

Some of the assembly techniques which can promote high performance and minimize cost and schedule risk are presented in this section. These techniques include; press fit assembly, solderless electrical contacts, and specialized handling procedures.

Press Fit Assembly

Press fit assembly techniques are valuable for use in spacecraft instruments because this design philosophy permits the instrument mass to be reduced, ensures cleanliness by avoiding adhesives and allows easy replacement of parts. Conventional satellite instrument design uses screw fastenerc to fix internal components in place. For applications where many components must be tightly packed into a small volume, press fit assembly techniques (similar in concept to commer .ial plug-in PCB cards) can be used to effectively increase component densities. This eliminates both the mass impact of using threaded fasteners and the associated volume to allow access to mechanical fastening points. Press fit assembly techniques circumvent the need to use adhesives to hold components in place; this is a very important consideration when items that are contamination sensitive, like detectors, are in close proximity. Since fasteners and adhesives are not required to fix components in place, replacement of these components is facilitated. This is an important consideration for applications that require iterative calibration adjustments or select-on-test component installation. An example of press fit assembly used on the SMS instrument is shown in Figure 4. The photo shows ceramic circuit boards mounted on metal guard rings that provide component interconnections between the guard rings. The guard rings are press fit into polycarbonate support pieces, and this assembly is itself press fit into the housing. Figure 5 shows the completed SMS instrument with the top cover removed, the components are press fit into the housing.

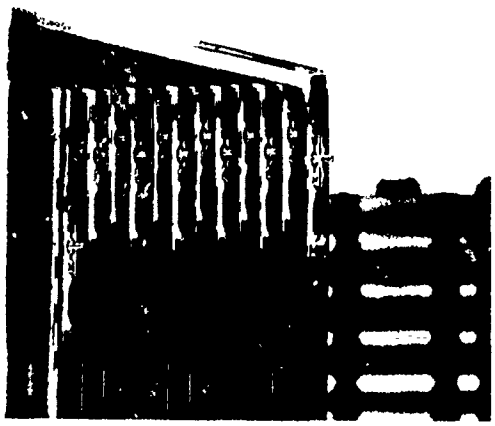


Figure 4: An Example of Press Fit Assembly and Solderless Contacts

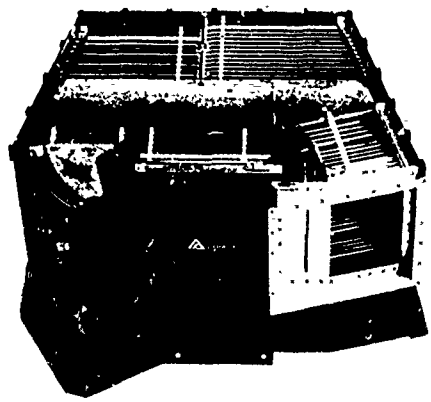


Figure 5: The SMS Instrument With The Top Cover Removed

Solderless Electrical Contacts

For most instruments it is desirable not to use soldering on final assembly for two reasons; the solder flux would contaminate the instrument, and the assembly/disassembly procedure would be very cumbersome.

The SMS instrument required more than 70 solderless electrical contracts. These were implement by using bellows shaped contact springs. This type for contact is highly reliable since the spring force is generated by a continuous surface [4]. Figure 6 shows a guard ring with a ceramic circuit board which contains a resistor, a chip capacitor and the contact spring.

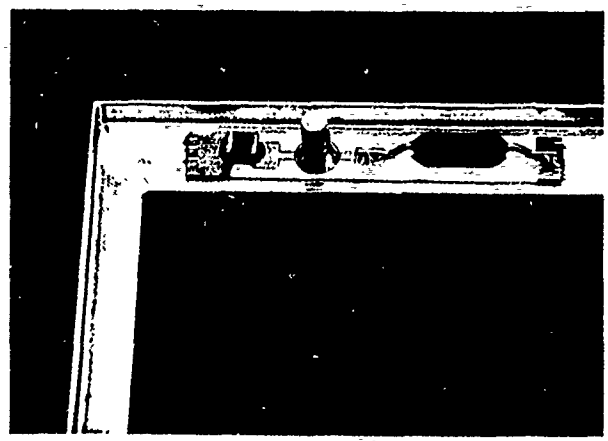


Figure 6: A Solderless Contact Assembly

Handling of Delicate Components

Many satellite instruments use components that require special handling procedures due to component fragility, electrostatic sensitivity or contamination sensitivity.

The handling procedures are very important in the instrument design since the cost and delivery times associated with replacement of critical instrument components would cause project cost and schedule penalties.

For the SMS instrument, fixtures were designed and used to protect the delicate grid subassemblies prior to, and during the instrument integration. All integration activities at the instrument level were performed on a laminar flow bench to minimize particle contamination that could degrade the performance of the SMS micro-channel plate detector.

Conclusions

We have discussed numerous processes and design methods that can be used for satellite-borne instruments. These techniques will promote higher instrument reliability, improved instrument performance and lower schedule and cost risks.

Acknowledgements

For the elements of this paper which relate to the SMS project, the authors acknowledge Dr. Brian A. Whalen and Dr. R.H. (Bob) Hum at the National Research Council of Canada for their technical support and encouragement.

References

[1] The example instrument, called SMS (Suprathermal and energetic ion Mass Spectrometer), is a state of the art satellite mass spectrometer developed by NRCC, and represents the most sensitive satellite mass spectrometer to date.

[2] Lowry, J.F., Fink, J.H., and Shumacher, B.W., "A major advance in high power electron beam welding in air", Journal of Applied Physics, Jan. 1976.

[3] MIL-P-55110 Military Specification, Printed-Wiring Boards, General Specification For.

[4] " Servometer Gold Plated Bellows Contact Springs" Servometer Corporation, Cedar Grove, NJ

EXPERIMENTAL INVESTIGATION OF IMAGING RADAR TECHNIQUES

R. F. SCHINDEL AND A. J. BLANCHARD

WAVE SCATTERING RESEARCH CENTER
UNIVERSITY OF TEXAS AT ARLINGTON
ARLINGTON, TEXAS 76019

ABSTRACT

Much attention has been given to imaging radar. To date, the primary focus of this attention is on imaging theory. It has been well established that three dimensional tomographic imaging is theoretically possible. However, only a very limited experimental data set in the microwave frequency range is available. This paper presents some experimental imaging results for both metallic and dielectric bodies. Some three dimensional data has also been recovered although only two dimensional projection images are presented at this time.

Key Words: ISAR, SAR, Inverse Imaging

INTRODUCTION

Microwave imaging theory is constantly being developed. However, the microwave inversion theory is still inadequate to describe experimental results. It has long been realized that inversion theory at its present developmental stage accounts only for first order scattering. However, it is clear that the present inversion theory is deficient on a more fundamental level. The experiments indicate that a microwave image is a map of various scattering mechanisms. Theoretical inversion can only produce an image of the support volume of the scatterers. This is accomplished by introducing an object function which has a value of one inside the scattering volume and zero outside the scattering volume. Furthermore, it will be shown that an image can be formed from the imaginary part of the inverted scattered field; whereas, the object function is purely real. Using several different targets, microwave imaging is investigated experimentally in this paper.

In the following section, a review of microwave imaging theory is presented. Next, a description of the various experiments is described, followed by the results of these experiments. The conclusions which can be drawn from the experimental results are presented in the final section of this paper.

THEORY

There are two fundamental interpretations of general inverse scattering theory for microwave experiments. The most common interpretation is borrowed from the ultrasonic theory developed by Porter - Bojarski (Langenburg, 1983) and Devaney,(1980). This formulation makes use of a Holographic field developed from application of the Scalar Green's theorem to the geometry shown in figure one. This formulation results in the well known Porter-Bojarski equation (Langenberg, 1983):

$$\theta_H(\bar{R},\omega) = -\iint\left[\phi_s(\bar{R},\omega)\frac{\partial G^*}{\partial \hat{n}'} - G(\bar{R}-\bar{R}',\omega)\frac{\partial \phi_s}{\partial \hat{n}'}\right] d\bar{S}'$$

$$= j2\int\int\int q_c(\bar{R}',\omega)\,\mathrm{Im}\{G(\bar{R}-\bar{R}',\omega)\}d^3\bar{R}' \qquad (1)$$

This equation relates backpropagated scattered field measurements, ϕ_s, and their normal derivatives on an arbitrary measurement surface, to secondary sources residing in the scattering volume V. This equation is valid under the scalar, Kirchhoff, and Born approximations. An equivalent formulation for the vector fields is possible. However, mathematical inversion results in the recovery of the gradient of the object function. The secondary sources in the scalar Porter-Bojarski equation, which ultimately provide the image of the target, are non-unique. Uniqueness is achieved if the so-called object function describing the geometry of the target is introduced and either angular or frequency diversity are employed. Within the far-field approximation, the well known Fourier inversion formulas shown below are obtained (Langenberg, 1983):

$$O_{LP}(\bar{R}) = \frac{R}{2\pi^2}\iiint\left[\frac{\phi_s^{far}(\bar{K})\,e^{-j\bar{k}_r\cdot\bar{R}}}{k^2}\right]e^{j\bar{K}\cdot\bar{R}}\,d^3\bar{K}$$

$$O(\bar{R}) = \frac{R}{\pi^2}\iiint\left[\frac{\phi_s^{far}(\bar{K})\,e^{-j\bar{k}_r\cdot\bar{R}}}{k^2}\right]e^{j\bar{K}\cdot\bar{R}}\,d^3\bar{K} \qquad (2)$$

From an experimental point-of-view, the theory is inadequate to describe the physics of inverse scattering. Although the theory leads to an inversion formula which works on experimental data, the inversion does not reproduce the object function of the target. Indeed, this type of formulation, which involves inversion to a non-physical object function, is not consistent with the IEEE definition of an image (Mensa, 1984), which states:

> "An image is a spatial distribution of a physical property such as radiation, electric charge, conductivity or reflectivity, mapped from another distribution of either the same or another physical property."

In an attempt to resolve this conflict of theory and experiment, Mensa has reformulated the inverse scattering problem for the microwave spectrum using object reflectivity. This formulation also results in a Fourier inversion of the scattered field measurements, however the inversion results in a map of the physical object reflectivity function instead of the non-physical object function. The inversion equations have the form (Dolaty and Blanchard, 1988):

$$g(x,y,z) = \iiint_{-\infty}^{\infty} H(F)\,G(F)\exp\left[2\pi j\left(f_x x + f_y y + f_z z\right)\right] df_x\, df_y\, df_z \tag{3}$$

The drawback of this formulation is that it only accounts for perfectly conducting targets. The Porter-Bojarski formulation is valid for perfectly conducting targets, as well as dielectric, or penetrable targets.

Both the Porter-Bojarski and Mensa formulations allow for tomographic reconstruction of a three-dimensional target. The tomograms are formed by projecting the scattered field spectrum onto a desired two-dimensional plane and then Fourier transforming to get the spatial distribution of the sources (object function or reflectivity) in the desired plane. The projection of the scattered field spectrum onto a given z-plane, β, is accomplished using the equation:

$$g(x,y,z=\beta) = \int\int_{-\infty}^{\infty} \exp\{j2\pi(p_x x + p_y y)\}\int_{-\infty}^{\infty} H(p_x,p_y,p_z)G(p_x,p_y,p_z)$$

$$\exp\{-j\sqrt{p^2 - p_x^2 - p_y^2}\,\beta\}\, dp_x dp_y dp_z \tag{4}$$

Although inconsistent with the theory, the Porter-Bojarski formulation can be used to show internal structure of penetrable targets using this tomographic technique. However, the Mensa formulation, if consistent, should only reveal the outline of the target in the desired plane. It is easily demonstrated experimentally that neither interpretation of inversion theory is adequate.

DESCRIPTION OF EXPERIMENTAL IMAGING

To obtain images of real targets, the inversion formulas require measurement of the scattered field over an arbitrary surface. The scattered field data processed to obtain the images presented in this paper were obtained at the bistatic radar cross-section measurement facility at the University of Texas at Arlington. The geometric model of the measurement facility is shown in figure two. An HP8510 microwave measurement system allows for broadband frequency diversity, and an azimuth-over-elevation positioner allows for angular diversity experiments. The data is processed into an image using a VAX 8700 and the images are plotted on a SUN 3 workstation. Several imaging experiments at various stages of completion have been designed for the measurement facility. The results of the completed experiments are presented in the next section.

The first image presented is of a 50:1 scale model of a B1-B aircraft, which has been metalized using nickel paint. This target allowed the evaluation of the inversion algorithm by providing a perfectly conducting object. An image of a rough surface is also presented to investigate imaging of extended targets. Finally, a target comprised of a dielectric cone with an imbedded three in radius spherical void has been designed. A graphical depiction of this target is shown in figure three. A three inch metallic sphere is also place in the void to provide for a composite target comprised of a weak dielectric, and a perfect conductor. Since the cone has a different cross-section at each cut perpendicular to its axis, tomographic reconstruction can be demonstrated.

EXPERIMENTAL RESULTS

Figure 4. like pole B1-B
Figure 5. imaginary B1-B
Figure 6. surface
Figure 7 cone

The results of the experiments described in the previous section are discussed in this section. Figure four shows a like-pole image of the metalized B1-B model aircraft. Obviously, the image is not a map of the object function of the target. This is clearly indicated by the fact that only the edges of the wings of the aircraft are visible in the image. A much more plausible explanation of the image of the wings is a mapping of the edge diffraction from the wings. This is also indicated by the poor image of the fuselage. Another inconsistency between theory and experiment is the ability to reconstruct an image from the imaginary part of the inverted measurements. An image produced from the imaginary part of the inverted data is shown in figure five. If the image was a map of the object function, the image should be purely real. An image of the random surface is shown in figure six. Again, it is apparent that the object function is not reproduced. The back edge of the surface appears as a deterministic structure. This is most likely due to surface waves propagating around the surface(Langenberg, 1983). The final image, shown in figure seven, is the dielectric cone with the metallic sphere. The dielectric interface with free-space and the sphere are clearly visible. This indicates that the internal structure of

penetrable scatterers can be obtained from inversion. This is not accounted for in the reflectivity formulation developed by Mensa, since the wavenumber remains constant throughout the object volume.

CONCLUSIONS

Inverse imaging theory for microwave scattering problems is inadequate to describe experiments for very fundamental reasons. Mathematical inversion results in a map of either the object function or the reflectivity function of the target. However, it is clear from experimental results, that the image is actually comprised of several scattering mechanisms. The experimental results contained in this paper indicate that a microwave image is actually a map of physical scattering mechanisms such as edge diffraction, and surface waves. Experimental results also indicate that the formulation in terms of the object reflectivity function is inadequate to describe the imaging of the internal structure of penetrable scatterers.

REFERENCES

A. J. Blanchard, M. Dolaty, " Bistatic Frequency Diverse Imaging of Complex Scattering Targets," Proceedings of IGARSS '88 Symposium, Edinburgh, Scotland, 13-16 Sept., 1988.

A. J. Devaney, "A New Approach to Emission and Transmission CT", 1980, *Ultrasonics Symposium*, pp. 979-983.

K. J. Langenberg, G. Bollig, M. Fischer, D. Bruk, "Transient Methods in Electromagnetic Imaging", *Inverse Methods in Electormagnetic Imaging*, NATO ASI Series, Series C: Mathematical and Physical Sciences, Vol. 143-Part I, 1983, D. Reidel Publishing Co., pp.87-110.

D.L. Mensa, *High Resolution Radar Imaging*. Artech House, Inc. Dedhom, MA. 1984..

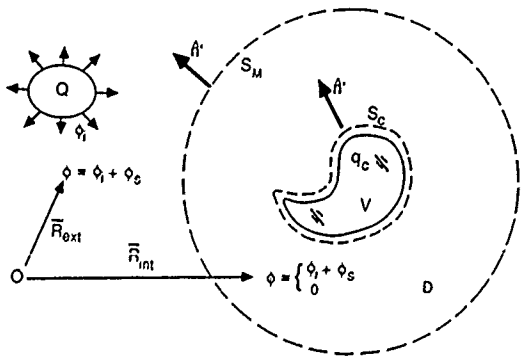


Figure 1. Geometry use in the development of the Scalar Green's Theorem.

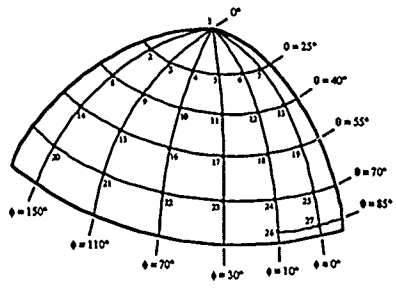


Figure 2. Measurement Facility

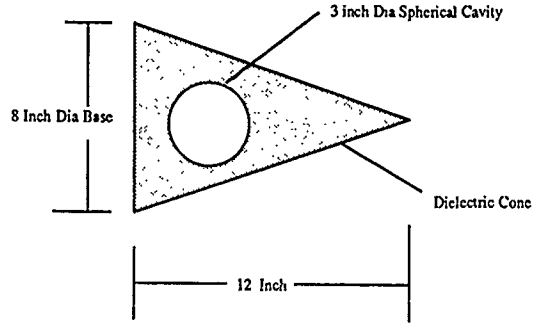


Figure 3. Dielectric target geometry

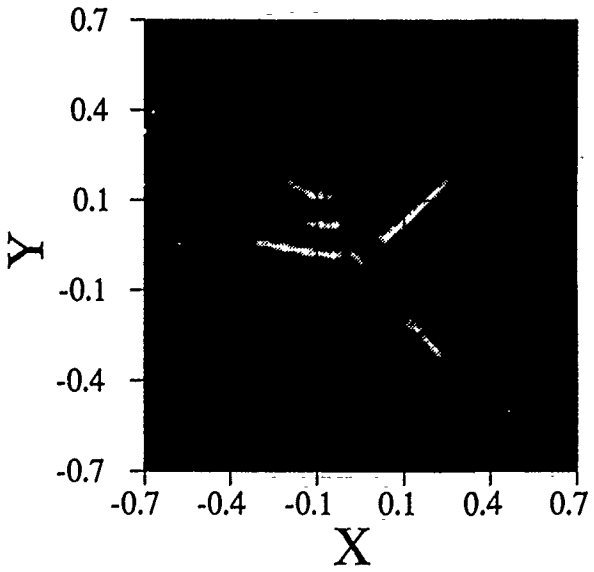


Figure 4. Like polarized (HH, Magnitude) image of a metalized aircraft

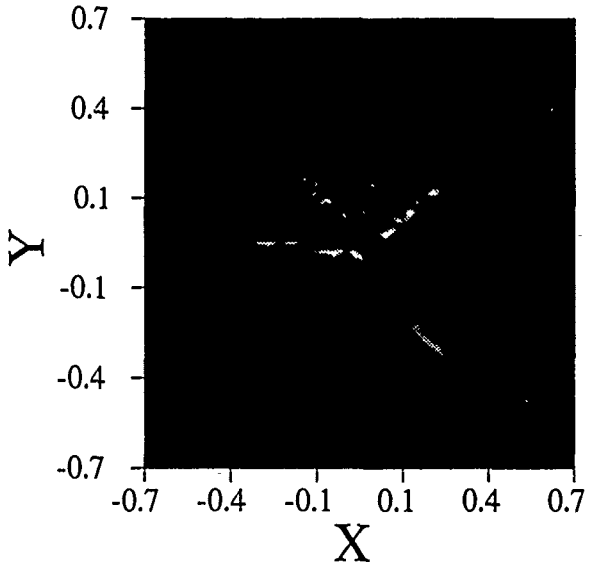


Figure 5. Like polarized (HH, Imaginary) image of a metalized aircraft

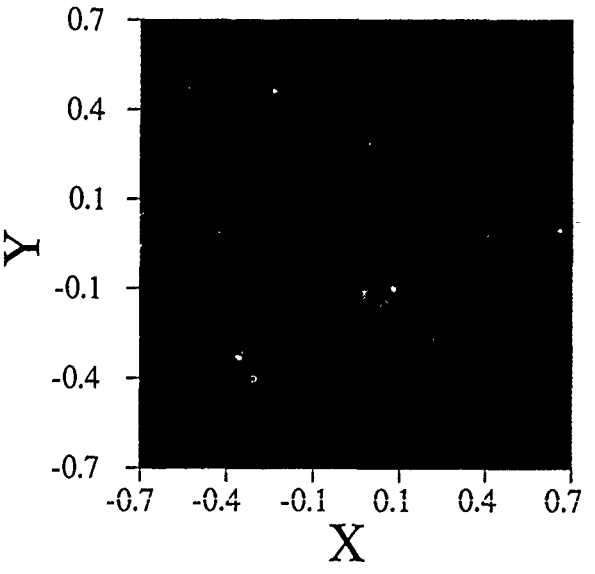


Figure 6. Like polarized (VV, Magnitude) image of a faceted surface

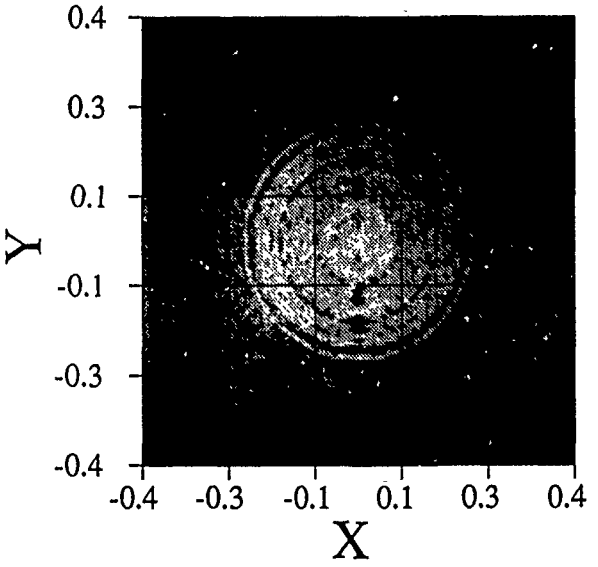


Figure 7. Projection image of the dielectric cone.

A COAXIAL-CIRCULAR WAVEGUIDE FOR DIELECTRIC MEASUREMENT

M.R. Taherian[*], *D.J. Yuen*[+], *and T.M. Habashy*[*]

[*] Schlumberger-Doll Research
Old Quarry Road
Ridgefield, CT 06877-4108

[+] Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology
Cambridge, MA 02139

Two-port dielectric cell measurements have been extensively used in literature to help in understanding the dielectric properties of a variety of materials. These dielectric measurements are also useful in developing a data base for the interpretation of dielectric measuring devices for geophysical applications. Most of the two-port dielectric measurements that exist in literature, employ a coaxial line cell. In this paper we introduce an alternative two-port dielectric cell which is a symmetric cylindrical sample holder connected to a coaxial line connector at each end.

We have developed a full-wave model that can predict the response of the cell. The model predictions agree satisfactorily well with the experimental measurements. We have also developed an inversion algorithm based on a modified Newton minimization approach to invert simultaneously for the dielectric constant and conductivity of the sample from the four S-paramter measurements supplied by a network analyzer.

In this paper we describe the measurement procedure, the full-wave forward model for predicting the response of the cell and the inversion algorithm. We show the reasonable agreement between theory and experiment and demonstrate how well the inversion procedure operates. We also compare the sensitivity of the cell proposed in this paper to that of the coaxial line cell.

MAGNETOTELLURIC INVERSION USING APPROXIMATE INVERSE MAPPINGS

D.W. Oldenburg and R.G. Ellis
Department of Geophysics and Astronomy
University of British Columbia
Vancouver, B.C.
V6T 1W5

Beginning with earth surface measurements of electric and magnetic fields the goal of an inversion solution is to find a conductivity structure which adequately reproduces these data. When the number of parameters in the earth model and the number of data are both in the order of thousands, the regular approach of linearizing the problem, and computing the Frechet derivatives to find how the data are changed when the model is perturbed, yield a formidable amount of numerical computation. We demonstrate that the technique of approximate inverse mappings can be used to generate a geophysically reasonable conductivity model with considerably less computation. Reduced computation is possible when the approximate inverse mapping is chosen to represent the dominant physical process involved. The mapping is used iteratively to converge to a true solution.

For the 2-dimensional magnetotelluric inverse problem the dominant physical process is the 1-dimensional attenuation of the electromagnetic fields in the conducting earth. This leads to an approximate inverse mapping based on 1-dimensional flattest model inversions at each measurement site. This choice of approximate inverse mapping yields a 2-dimensional inversion scheme which is fast and robust. An important feature of the approximate inverse mapping technique is the convenient manner in which information from different sources may be included in the solution. For example, in the 2-dimensional magnetotelluric problem there are four common measurements: apparent resistivity and phase for TE and TM modes. We demonstrate how information from these four sources can be used to provide a consistent model. Inversion results will be presented for a number of different data sets.

FUZZY CLASSIFICATION WITH APPLICATIONS TO GEOPHYSICAL DATA

BY

Bijan Lashgari

Unocal Corp.

Abstract- The theory of fuzzy classification with applications to geophysical data is explored in this paper. The proposed technique is motivated by inaccuracies and unreliabilities that often exist in the geophysical data which adversely affect the performance of the Bayesian classifiers. Fuzzy classification, through its membership function, offers an elegant solution to the problem. The algorithm consists of a supervised and unsupervised components. These two components interact in a hybrid fashion depending upon the level of uncertainty associated with the control samples. The algorithm takes into account the uncertainty in the data by assigning continuous membership grades to the samples with respect to the classes. Examples from seismic data illustrate the geophysical applications of the method.

The development of the fuzzy classification stems from the need to characterize the inaccuracy that may exist in the measured data or the training samples. Such imprecision is often encountered in geophysical data due to noise caused by measuring instruments or data processing steps. In the data preparation step, a great deal of decision making processes take place that can introduce error in the final interpretation, such as imprecision in the subsurface velocities, poorly estimated reflector dips, etc. {7,9} . Final interpretation of the data is further limited by the uncertainty resulting from the lack of signal resolution as well as the nonuniqueness nature of the geophysical problem.

There are also other types of inaccuracies which are due to the erroneous statistics caused by insufficient data samples. This type of error, for instance, has paramount effect on the classification result where a specific probability distribution is assumed, like in the case of the Bayesian classifiers.

While fuzzy classification offers the conventional classification for "hard" clusters it also handles the representation for "fuzzy" clusters resulting from the unconsolidated samples not having any specific membership in a particular group. Such a fuzzy behavior of the samples could be due to the inaccuracies in the input attributes or lack of sufficient information about the training samples. Fuzzy classification through its continuous class membership provides a measure for handling these uncertainties. The samples that are in the core of the clusters have membership grades close to one, while the samples at the periphery have membership grades approaching zero. Moreover, the membership grades are further influenced by a set of training samples which may be unreliable or "fuzzy" in character.

DIAGNOSTIC ELECTROMAGNETIC SEAM WAVE IMAGING OF ANOMALOUS GEOLOGIC DISTURBANCE ZONES IN LAYERED FORMATIONS

by: Larry G. Stolarczyk
Stolar, Inc.
Raton, New Mexico
505-445-3607 (Tele)
505-445-9659 (Fax)

ABSTRACT

Continuous Wave (CW) Radio Imaging Method (RIM) technology has been developed and applied in the last five years in more than 150 diagnostic surveys of geologic disturbance zones in layered formations. Reconstructed tomographic images of faults, dikes, sills, roof/floor paleochannels and rapidly thinning seams have been acquired and compared with in-mine geologic mapping data obtained in mining near or through geologic disturbance zones. The images have been used by mining companies to guide mining machines away from hazardous zones. The RIM technology is based upon the transmission and measurement of electromagnetic wave propagation constants along multiple ray paths in natural waveguides in the earth. Waveguides are formed whenever less conductive seams of coal, trona, potash, gilsonite, quartzite or oil/gas sandstones are surrounded by more conductive rock layers. The predominate seam wave is a zero order mode transverse electromagnetic (TEM) wave with the electric field vertically polarized between the surrounding more conductive layers while the magnetic field component is horizontally polarized in the seam. When an EM seam wave intersects a geologic anomaly, the propagation constants change from the values determined for an undisturbed waveguide. Interpretation of image results follows from the analytical determination of the attenuation and phase shift rates for the prevailing electrical parameters of the layered formation (J.R. Wait, Radio Science, April 1976), (P. DeLonge, Peter Peregrinus, LTD., 1982), (D. Hill, Journal of Research of the National Bureau of Standards, September-October 1984). This paper compares the propagation constants determined by the tomography algorithm and analytical formulations with in-mine geologic mapping results.

INTERPRETATION OF THE DEPOLARIZING EFFECTS IN VECTOR (POLARIZATION) DIFFRACTION TOMOGRAPHY

Nabil A. Soliman and Wolfgang-M. Boerner

University of Illinois at Chicago
UIC-EECS/CL, M/C 154
840 W. Taylor St., SEL-4210
Chicago, IL 60680-4348
Tel: +1(312)996-5480
Fax: +1(312)413-0024

The major difficulty of evaluating possible practical applications of vector diffraction tomography is that experimental data are still sparse, and that data generated involve usually either the BORN or RYTOV scalar wave approximations. In order to test the depolaring effects in electromagnetic wave diffraction tomography and especially for the inverse problem of image formation, oblique incidence on a circular cylindrical scatterer is investigated. It is shown that reconstruction of the exact fields at oblique incidence deteriorates rapidly as the angle of incidence deviates further from normal incidence. And it is demonstrated how the entire copolarized (HH and VV) and cross-polarized (HV and VH) components must be carefully integrated into the image reconstruction process in order to obtain a complete image according to the "span" - invariant. By analyzing the obvious but simple model of a cylindrical conducting and/or dielectric scatterer at oblique incidence, the role of wave depolarizing effects in image formation in vector diffraction tomography is clearly interpreted.

RECENT ADVANCES IN THE DEVELOPMENT OF THE PHYSICAL OPTICS INVERSE SCATTERING THEORY FOR NON-SYMMETRIC, CONDUCTING, CLOSED SHAPES THE DEPOLARIZING EFFECTS FOR THE MONOSTATIC AND BISTATIC CASES

Bing-Yuen Foo and Wolfgang-M. Boerner

University of Illinois at Chicago
UIC-EECS/CL, M/C 154
840 W. Taylor St., SEL-4210
Chicago, IL 60680-4348
Tel: +1(312)996-5480
Fax: +1(312)413-0024

Advances made in the electromagnetic inverse scattering problem for plane wave incidence on a perfectly conducting closed, convex, but not rotationally symmetric scatterer under the physical optics solution are considered for the monostatic and bistatic cases.

Using Kennaugh's target ramp response formulation, Kennaugh's formula is extended to the bistatic case; and in order to account for the polarization characteristics of vector electromagnetic inverse scattering in both monostatic and bistatic cases, Bennett's polarization correction to physical optics is applied and extended to obtain asymptotic solutions between the phase difference of the co-polarized and the co/cross-polarized elements of the bistatic scattering matrix, and the principal curvature difference at the specular point.

In addition, the effects of the torsional term in the Minkowski-Hurwitz equation of vector differential geometry, which is asscciated with this problem, are considered and the properties of Huynen's helicity factor are interpreted.

SPHERICAL APPROXIMATION THEORY OF INVERSE SCATTERING

Taner ŞENGÖR, MEMBER, IEEE
Department of Electrical Engineering,
Yıldız University, Yıldız, 80750 İstanbul, Turkey

Abstract.

A method to obtain the configuration of perfect conducting scatterers is determined. Most of the methods addressed to inverse scattering use physical optics current approximation that assumes the induced surface current on the dark region is zero. However, some nonvanishing current constitutents flow on the whole of scatterer[1].

The aim of this investigation is to treat the surface wave traveling on a three dimensional curved surface having edges. Scattering by inflection, tip, and corner points were excluded.

We trace a way beginning with canonical problems cited in (1) with a different method introduced here. The explicit expressions for the scattering of rotationally symmetric high frequency fields by a perfect conducting spherical sheet are obtained via surface currents induced on the cap. The contributions of various diffraction phenomena to surface current are given explicitly. The results explain the contribution of surface currents to scattered field and introduce a method to obtain the configurations of scatterers and these results can be extended to obtain source configuration. The canonical formulas are accompanied with numerical applications. These results give a method which is used in inverse problems. We call spherical approximation theory of inverse scattering (SATIS) the introduced method. The physical and geometrical results were extended to surfaces including no inflection, tip, and corner points in (2) by taking into account the physical behaviour of phenomena. The details of the method for calculating the various constitutents of surface current and expressing the backscattered field in terms of these constitutents are explained in (3).

To have information about a target we assume that every point M on the target is approximated by a point on a sphere with radius a_M since high frequency diffraction is a local phenomenon with respect to the postulates of GTD.

The results based on the following theorem[3,4,5]:

If the incident field at an observation point P is obtainable then the total magnetic field which is radiated by the source pairs K and K' at P containes $n_{max}(P)$ unknown terms which give sufficient knowledge about the target where $n_{max}(P)$ is defined as

$$n_{max}(P)=2+\sum_{K}(n^{K}_{max}(P)-2+4\overline{K}n^{T}_{KP}(M)) \qquad (1)$$

where

$$n^{K}_{max}(P)=83n^{illum}_{KP}(M)+75n^{shadow}_{KP}(M)+77n^{shadow}_{K'P}(M)+62n^{dark}_{KP}(M)+2 \qquad (2a)$$

$$n_{KP}^{T}(M)=n_{KP}^{illum}(M)+n_{KP}^{shadow}(M)+n_{K'P}^{shadow}(M)+n_{KP}^{dark}(M) \qquad (2b)$$

$\overline{N}$ is the finite number of the roots μ of $J_{\mu}(ka_M)=0$ where $Re\mu>0$. The measurements of the magnetic field at the same point P with different values which has the number of $(1/2)n_{max}(P)$ of wave number k give the opportunity of obtaining sufficient knowledge about the perfect conducting target.

A method for target sensing by surface current constitutents induced on perfect conducting obstacles is obtained. The method determines the diffraction phenomena occured on the scatterer and gives the coordinates of the points on surface approximately.

Refefrences.

(1) M. İdemen and T. Şengör, "High-frequency surface currents induced on a spherical cap", SIAM J. on Appl. Math., Vol. 46, No.1, pp.99-117, Feb. 1986.
(2) T. Şengör, "Cross section of perfect conducting bodies by surface currents" (submitted).
(3) T. Şengör, "Surface current approach to inverse scattering and scattering by spherical sheet" (submitted).
(4) T. Şengör, "Imaging of perfect conducting bodies by surface currents" 1988 IEEE AP-S Symposium & URSI Radio Science Meeting, Syracuse (published).
(5) T. Şengör, "Some numerical results related with inverse scattering by surface currents", Proceedings of ISRP'88, Int. Symposium on Radio Propagation, Beijing (China), April 18-21, 1988 (published).

AN AUTOMATED METHOD FOR ESTIMATING OCEANIC FLOW FIELDS FROM SATELLITE IMAGERY

Shawn D. O'Keefe and Melba M. Crawford

The Center for Space Research
and
Department of Mechanical Engineering
The University of Texas at Austin
Austin, Texas 78712-1085

ABSTRACT

Satellite imagery interpretation, *in situ* studies, and mathematical simulations have all been used extensively to monitor eddies and infer the corresponding near surface circulation patterns. Although these studies have yielded invaluable information about these oceanic features, tracking techniques are still inadequate. While *in situ* tracking is limited by the expense-constrained spatial and temporal gaps in the data, other methods utilizing remotely sensed data are restricted by their computational and man-hour requirements and limited range of applicability.

A new feature based approach is being developed to automatically estimate the advective velocity fields for an eddy. Linear features are identified in each image and combined pairwise to form corners. Hypothesis integration, an artificial intelligence evidence-accumulation technique, is used to efficiently match similar corners from different images, resulting in tentative hypotheses which are clustered in the transform space. Dense clusters represent likely global transforms which are verified or refuted; verified transformations provide an estimate of feature movement. The methodology has been applied to co-register images by matching coastlines. Current research is focusing on extending the algorithm to estimate both the translational and angular velocities of an oceanic eddy from the thermal channels of AVHRR satellite imagery. The new method is potentially superior to correlation methods which are more computationally intensive as well as theoretically limited to tracking translational movement.

Keywords: Feature Tracking, Feature Detection, Hypothesis Integration, Eddy Tracking

INTRODUCTION

Remotely sensed imagery is an important source of information for detecting and monitoring changes in environmental phenomena. Common applications include tracking storm systems by estimating cloud motion, tracking sea ice, and monitoring changes in vegetation caused by deforestation. Oceans are perhaps the most important and least understood link in the heat and mass transfer cycle of the earth. The vast expanse of ocean, the remoteness of many areas, and the massive scale of many oceanic features make it necessary to monitor the ocean through remote sensing. Scientists modeling the circulation of the world's oceans and seas need synoptic data to develop, validate, and update their models. No other means exists to obtain data from many of the locations, and certainly no alternative is available to obtain nearly continuous coverage over such wide areas.

Recently satellite imagery has been used in tracking mesoscale oceanic features such as eddies. Sea-surface advective velocity estimates are obtained from thermal image sequences either entirely by hand or by working interactively with a computer and an imaging system [Emery et al, 1986; Vastano and Reid, 1985; Vastano and Borders, 1984]. Cross-correlation can be used to estimate the movement of an area by correlating the intensities of pixels between windows in the images -- the shift which yields the highest correlation is used to determine the displacement of the feature [Emery et al, 1986]. Vastano subjectively tracks specific eddy flow features from the first to the second image in a manpower intensive approach. Clearly inappropriate velocities are removed subjectively for both approaches The resulting velocity vectors show good correspondence when compared with historical data and simultaneous information obtained from buoys. In a related application, both techniques were found to be effective in estimating sea-ice travel from synthetic aperture radar images [Vesecky et al, 1988].

Although the effectiveness of both tracking techniques has been demonstrated, each has serious dificiencies. Subjective tracking requires a high degree of training, and the task is time consuming and leads to inconsistent results -- not only between different operators, but even for the same expert in repeated trials [Cornillon, 1988; Vastano and Borders, 1984]. Cross-correlation methods are highly computer intensive and theoretically limited to tracking translational movement. Rotational movement must be inferred by applying these methods in local windows where translation is a good approximation or where the images are rotated artificially relative to each other [Emery et al, 1986].

Thus, current tracking techniques are inadequate While *in situ* tracking is limited by the expense-constrained spatial and temporal gaps in the data, other methods utilizing remotely sensed data are restricted by their computational and man-hour requirements and limited range of applicability.

METHODOLOGY

A new feature based approach has been developed to automatically estimate advective velocity fields from a temporal sequence of images. The approach involves two stages -- detection of features in all images in the sequence, followed by the matching of features across the images. To increase efficiency, the algorithm is executed hierarchically, initially at a reduced resolution to define the geographic search area and to bracket the transformation area of interest and then at full sensor resolution to refine the estimates obtained from the first stage. A simple land/water/cloud classification scheme is used to restrict feature detection to the surface-type of interest and to ignore artificial edges induced by clouds. Each component of the procedure is discussed in greater detail in the following sections.

Feature Detection The first step in identifying features is to detect boundary pixels -- points in an image lying on the border between regions (such as land and water or two bodies of water with different temperatures.) Several image processing edge detection techniques are appropriate for this task; the research reported in this paper uses the Sobel operator [Ballard and Brown, 1982]. This operator convolves two 3x3 templates with the image to compute

smoothed gradients, any pixels with significantly large gradients (exceeding a user-supplied threshold) are passed on to the second step. Besides smoothing the calculated gradients, the smoothing process inappropriately broadens the detected edges. Directional nonmaximal suppression is used to thin the lines to an acceptable single pixel width before moving to the next step of the algorithm.

Connected boundary pixels are then combined into linear features by means of linear regression. The algorithm examines each pixel in the gradient image for locations which exceed the threshold. When one is found, its neighboring pixels are compared with the threshold. This approach recursively builds a sequence of connected boundary pixels. When no additional high gradient pixels can be added to the list, the sequence of pixels is partitioned into approximately linear sections and represented by a segment with a location, length, and orientation (obtained through linear regression.) This step of the algorithm continues until all such lines in the image are found. Figure 2 demonstrates the results of this step of the approach as applied to the thermal features observable in the AVHRR thermal imagery in Figure 1. Note that the number of 'pieces of information' have been reduced from hundreds of thousands (262K/1.048M pixels in a 512x512 and 1024x1024 image, respectively) to a much more manageable few hundred 'significant' features.

The algorithm then consolidates information further by combining lines pairwise to produce corners. A corner is a more stable feature for estimating location than a line because line endpoints often vary, causing the center point to similarly wander. The location of the intersection or extrapolated intersection of two lines is known to be much more consistent.

Temporal Feature Matching Once linear and corner features are detected for each image in the temporal sequence, the system matches the features across images using an artificial intelligence evidence-accumulation technique known as hypothesis integration (HI). HI is an efficient matching technique related to the Hough transform and based on clustering [Hwang et al, 1986; Hwang, 1987]. Each matching between a model and image feature generates a candidate hypothesis -- the transformation required to map the model feature into the image location and orientation. All such pairwise matchings, one feature from each image, define the total sample transform space. The method assumes mismatches randomly fill the space, but matches between corresponding features tend to cluster. Consistent (similar) hypotheses are mutually supportive and thus are integrated. When sufficiently many mutually-supportive hypotheses have been clustered, the transformation is verified or refuted by comparing the transformed model with the image. This validation stage actually enhances robustness by testing transformed model features against the sensed image before accepting the transformation's validity. The transformation can theoretically include translation, rotation, and scaling dimensions, although a four dimensional clustering problem may well be computationally unmanageable.

Hypothesis integration is more efficient than other matching methods -- all possible pairwise matches need not be completed before a 'winning' candidate transform is elected. The difficulty is in deciding when the election is over -- when have enough transforms clustered? One of the extensions of the research reported in this paper is the definition of an appropriate statistical threshold for distinguishing clusters representing corresponding features from random clusters. In this study, transforms are weighted by the length of their features under the assumption that longer features are more likely to be detected in both images. This weighting, together with an assumption of uniformly distributed edge lengths, allows an expression for the probability density of the length contained within a discretized bin in the transform space to be determined, and thus a threshold. Since corresponding features will not always exist in multiple images due to sensor and edge definition noise, the expected reduction in the number of comparisons required cannot be stated with certainty, however, the savings over a technique that must match all possible pairs is substantial.

The method has been implemented in a hierarchical framework to further increase its efficiency. The entire problem is solved twice, once at lower resolution (nominally reduced from the original by a factor of four in each dimension), and a second time at full resolution, but in a restricted space. The knowledge obtained from the lower resolution solution allows the algorithm to concentrate the feature detection efforts in geographic regions of particular interest and to similarly focus the feature matching efforts in a narrow transform subspace.

For most applications, the densest cluster defines the best global transform mapping the model onto the sensed image. In this eddy tracking application of hypothesis integration, the noisier images, volatile thermal features, objects of uncertain and fluctuating shape, and erratic irregular background force the use of windows to determine the locally optimal transformation(s).

RESULTS

The methodology has been applied to images obtained from the AVHRR sensor flying on the NOAA-10 and NOAA-11 satellites. Data was acquired with the Seaspace capture system and calibrated and interactively registered with its Terascan software. Representative thermal images and their associated edges are shown in Figs 1 and 2, respectively. Both images are from NOAA-10, captured approximately 12 hours apart on March 11 (14:05 GMT) and March 12, 1989 (01:20 GMT). The scene is located in the eastern Gulf of Mexico with Florida, Cuba, and the Yucatan peninsula of Mexico clearly visible to the right, lower right, and lower left quadrants, respectively.

CONCLUSIONS

The methodology has been successfully applied to rectifying images by detecting and matching coastlines as determined by a vegetative index. The hypothesis integration methodology is being modified to handle the third (rotational) dimension and windowing necessary to track oceanic eddies. Results are forthcoming. In addition, other smoothing operators are being investigated in the feature detection stage of the algorithm.

Although this approach has been applied only to image rectification and eddy tracking, the development of automated procedures to detect and track phenomena through imagery is a general contribution with potential applications to a wide range of other remote sensing and fluid mechanics problems. Additionally, this research demonstrates that computer vision/artificial intelligence techniques can be fruitfully and efficiently applied in remote sensing applications.

ACKNOWLEDGEMENTS

The authors would like to thank Kenneth Chong, Ahmed Rasool, and Thomas Suniga for their assistance in processing the AVHRR images. Also, thanks go to the Dr. Vincent Hwang, formerly of the University of Texas Computer Science Department, for providing the initial version of his image processing system as the basis from which this software was derived, and to William Kastak for the use of his image rectification code. This research was supported by the Texas Advanced Technology Research Project (TATRP) under Grants 14-9700 and 14-9710.

REFERENCES

D. Ballard and C. Brown, *Computer Vision*, Prentice-Hall, Inc, 1982.

P. Cornillon, 1988, personal communication.

W. Emery, A. Thomas, M. Collings, W. Crawford, and D. Mackas, "An Objective Method for Computing Advective Surface Velocities from Sequential Infrared Satellite Images," *Journal Of Geophysical Research*, Vol 91, No C11, 1986, pp 12,865-12,878.

V. Hwang, L. Davis, and T. Matsuyama, "Hypothesis Integration in Image Understanding Systems," *Computer Vision, Graphics, and Image Processing*, 36, 1986, pp 321-371.

V. Hwang, "Recognition of Two Dimensional Objects Using Hypothesis Integration Techniques," Technical Report, University of Texas at Austin, AI TR87-57, 1987.

A. Vastano and S. Borders, "Sea Surface Motion over an Anticyclonic Eddy on the Oyashio Front," *Remote Sensing of the Environment*, Vol 16, 1984, pp 87-90.

A. Vastano and R. Reid, "Sea Surface Topography Estimation with Infrared Satellite Imagery," *Journal of Atmospheric and Oceanic Technology*, Vol 2, No 3, 1985, pp 393-400.

J. Vesecky, R. Samadani, M. Smith, J. Daida and R. Bracewell, "Observation of Sea-Ice Dynamics Using Synthetic Aperture Radar Images: Automated Analysis," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 26, No. 1, 1988, pp 38-48.

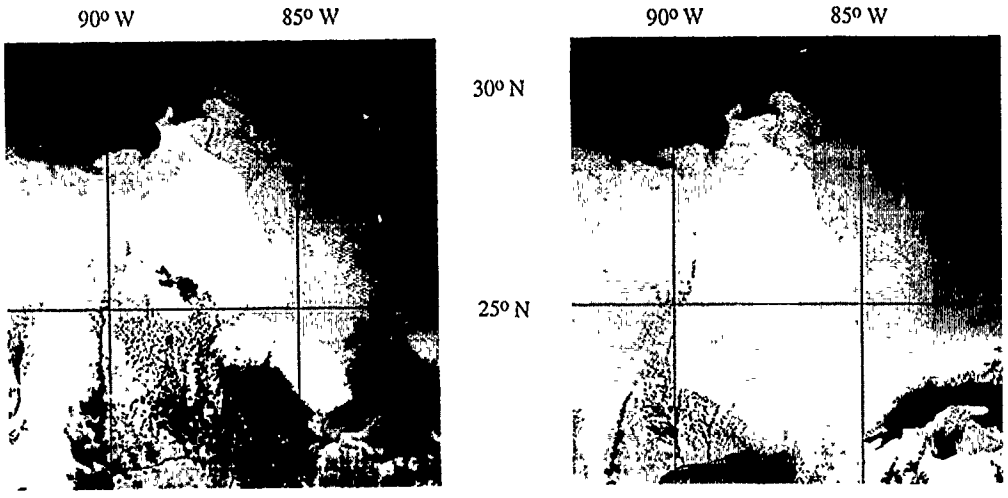


Fig 1 NOAA-10 AVHRR thermal images of the eastern Gulf of Mexico taken on March 11 and March 12, 1989 at 14:05 GMT and 01:20 GMT, respectively. Surface temperatures range from about 8.5° (dark) and 24° C (light). Clouds appear as black. The dramatic warm feature in the center of the images is the northern intrusion of the Loop Current.

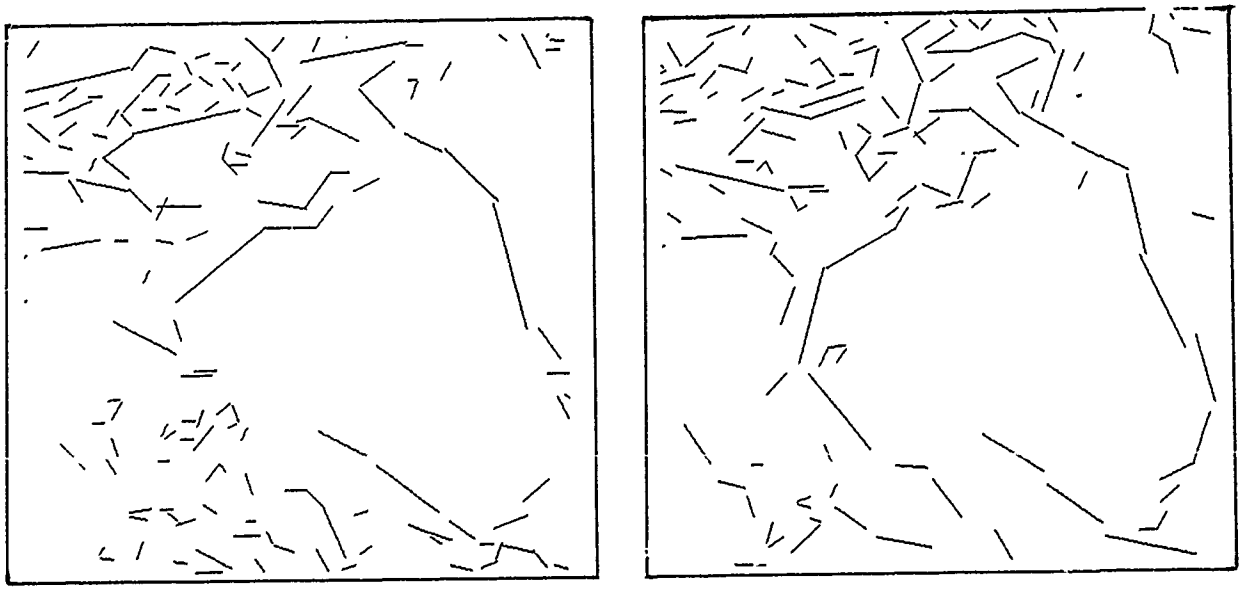


Fig 2 Edges detected in a subspace of the images in Fig 1. Edges were actually computed on an image coarser by a factor of 4 in each dimension, which was obtained from the original full resolution image by median sampling.

AVHRR observations of the horizontal structure of the surface layer of the ocean under low wind conditions

P. FLAMENT
Department of Oceanography
University of Hawaii at Manoa

Abstract. Diurnal warming layers that form in low wind display coherent horizontal structures in the form of streaks at least 50 km long, with a wavelength of 4-8 km and an amplitude of 0.5-1.5°C. These features have been observed in the California Current in three different occasions using a combination of satellite and in-situ measurements. It is hypothesized that their scale is set by planetary boundary layer circulations.

1. Introduction

Large diurnal warming of the surface layer of the ocean occurs in low wind and low cloud cover conditions (e.g. Bruce and Firing, 1974). The diurnal amplitude of surface temperature can then reach several °C, over regions that have been shown to correspond closely to anticyclonic ridges (Stramma *et al.*, 1986; Cornillon and Stramma, 1985). In contrast, when a well-developed wind-driven mixed layer is present, the diurnal amplitude seldom exceeds a few tenths of °C.

We present here 1-km resolution AVHRR infrared images and in-situ CTD profiles that suggest the existence of coherent horizontal structures in these warming layers, at scales of 3 to 10 km.

2. Satellite images

The area studied is shown in Fig.1. An anticyclone was centered at 135°W 40°N during July 1985, resulting in low winds off the coast of California. Three AVHRR images were acquired on 8 and 9 July, at about 14:00, 19:00 and 08:00 local (UT-8) time. Band 4 (10.8 μm) will be used for sea surface temperature.

Surface temperature on 8 July at 19:00 is shown in Fig. 2. Cold water due to coastal upwelling and offshore advection of coastal water is seen in the east of the image. Over most of the image, the temperature field consists of large nearly isothermal patches, separated by fronts a few km wide. For example, near [x=200,y=160] km, surface temperature does not vary much over a radius of ~30 km. This structure is typical of infrared images of the California Current.

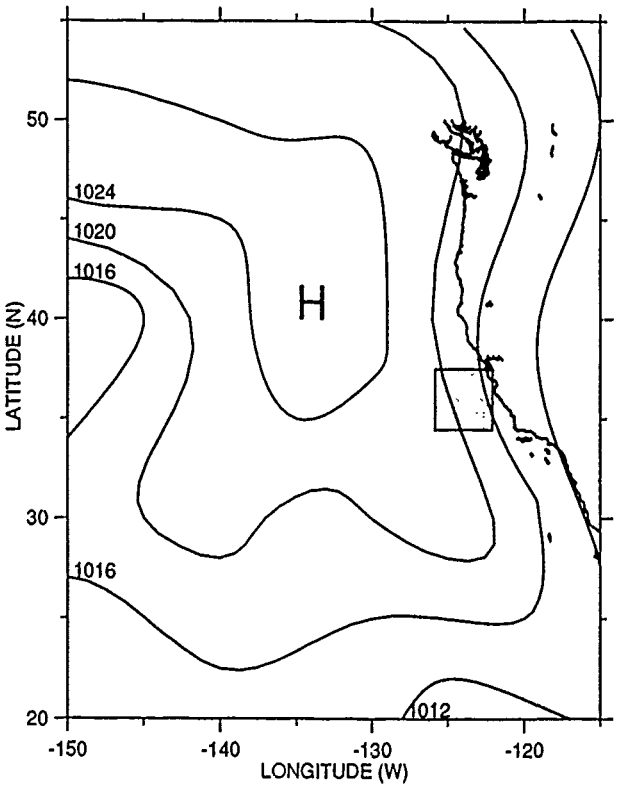


Fig. 1. Map of the area studied, with the surface pressure field on 11 July at 16:00 (UT-8). Note the anticyclonic ridge extending southeastward from 135°W 40°N. The position of the image shown in Figs. 2 and 3 is outlined.

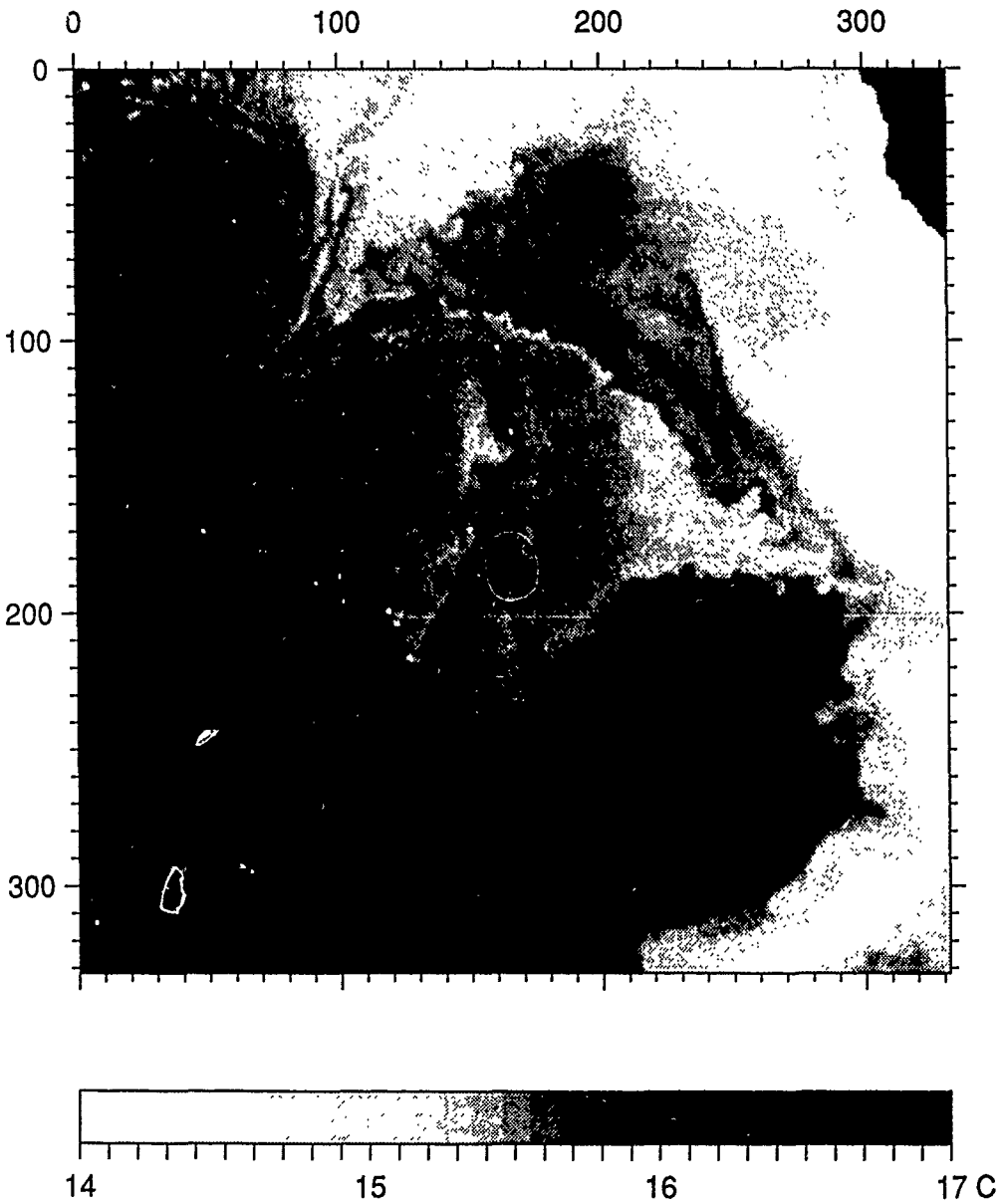


Fig. 2. Thermal infrared AVHRR image on 8 July 1985 at 19:00 (UT-8). Cold water is coded in white and warm water in black; the temperature scale is labeled in °C. The grid is labeled in km. The origin is at 125°48 37°30N.

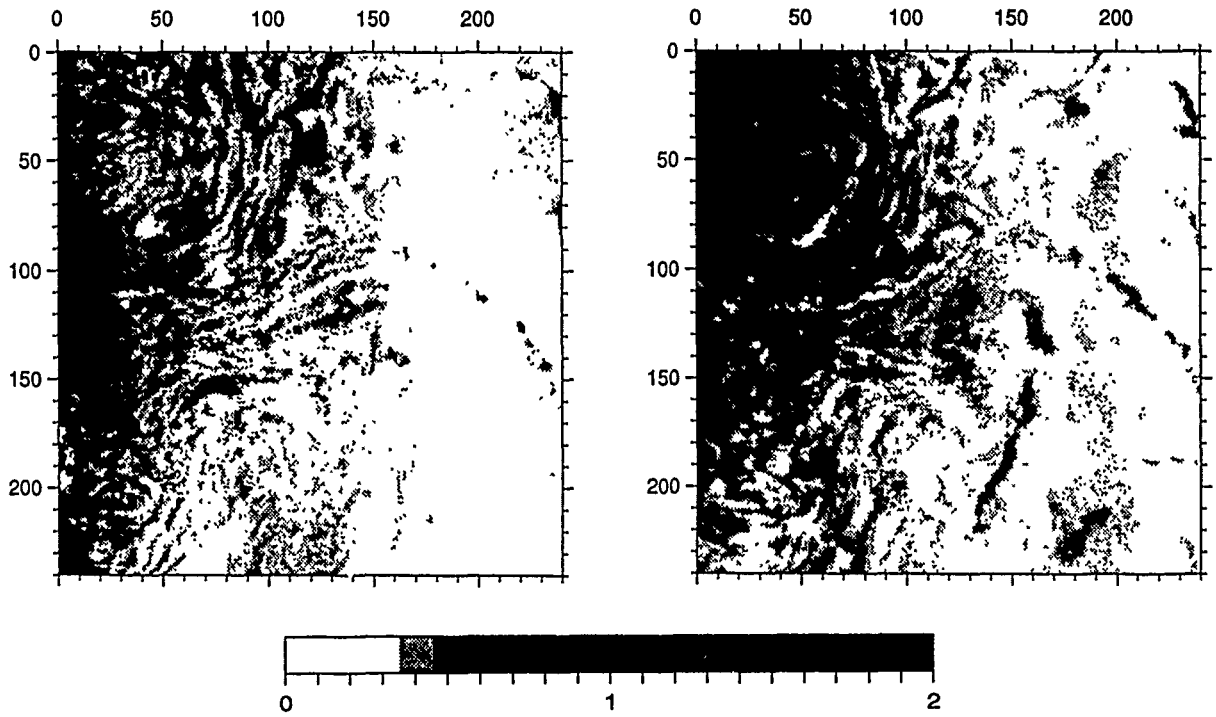


Fig. 3. (a) difference between the images on 8 July at 14:00 and at 19:00 (UT-8); (b) difference between the images on 8 July at 19:00 and on 9 July at 08:00 (UT-8). Differences smaller than 0.3°C are white; larger differences are shown in gray.

However, a very different structure is seen in the northwest of the image. In that area, surface temperature varies rapidly, and displays numerous streaks ~50 km long with a typical wavelength of 5-6 km and amplitude of 0.5 to 1°C. They are most clearly seen near [x=80,y=90] km. They seem to align preferentially along the streamlines of the mesoscale flow. These streaks are also seen in the preceding and following images, but seem most intense in the image shown. Although they are visually reminiscent of clouds, they are not clouds: the difference between brightness temperatures at 3.7 μm and 10.8 μm was negligible, unlike clouds which have a smaller emissivity (0.7) at 3.7 μm.

Fig. 3 shows the differences between the 14:00 and 19:00 images, and between the 19:00 and 08:00 images. Over the center and east of the images, the differences are very small. They are less than 0.3°C, except near fronts where advection dominates. Over the area of the streaks, however, the signature of a large diurnal warming event is clearly seen, with differences of the order of 0.5 to 1.3°C between pairs of images and a decrease of temperature from the afternoon image to the morning image.

3. In-situ measurements

Two weeks after these images, we had the opportunity to sample diurnal warming layers from ship. The wind was less than 1 m/s and the sea surface was glossy. On 25 July at 23:00, our attention was caught by periodic variations of intake temperature while steaming at about 2.5 m/s. A mixed-layer drifter drogued at 5 m was deployed and a 3-hour survey was conducted along the track shown in Fig. 4. The position of this survey is [x=260,y=200] km in Fig. 2. The motion of the drifter corresponded to a mean flow less than 10 cm/s, indicating that the area was not in a jet or mesoscale front.

Fig. 5 shows surface temperature and salinity at 50 cm depth, measured at an intake on the side of the ship. Salinity was virtually constant through the survey, but temperature varied between 14.9 and 15.7°C, with features identifiable at scales of 2 to 8 km. There is a clear symmetry at the southeast turning point of the survey, suggesting that the temperature variations were the expression of streaks running northeast to southwest.

A CTD was yow-yoed between the surface and 40 m depth during the survey. Using a winch

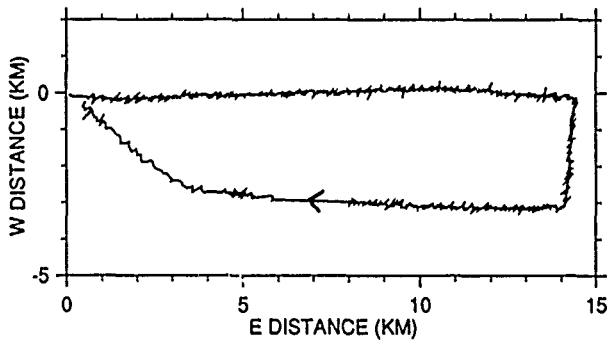


Fig. 4. Ship track over ground on 25 July at 23:00 (UT-8). The origin is at 122°20W 35°45N.

rate of about 40 cm/s, 80 vertical profiles were obtained. The profiles at km 1 and km 5 are shown in Fig. 6; the other profiles were in all aspects similar. Both profiles reveal a "fossil" mixed-layer about 29 m deep, capped by a near-surface restratification. The amplitude of the restratification varied between 0.3 and 1.3°C and its thickness between 3 and 8 m. Salinity was constant from the surface to 29 m (the small salinity anomaly near the surface is caused by inappropriate matching of the conductivity and temperature cell responses; it is not real).

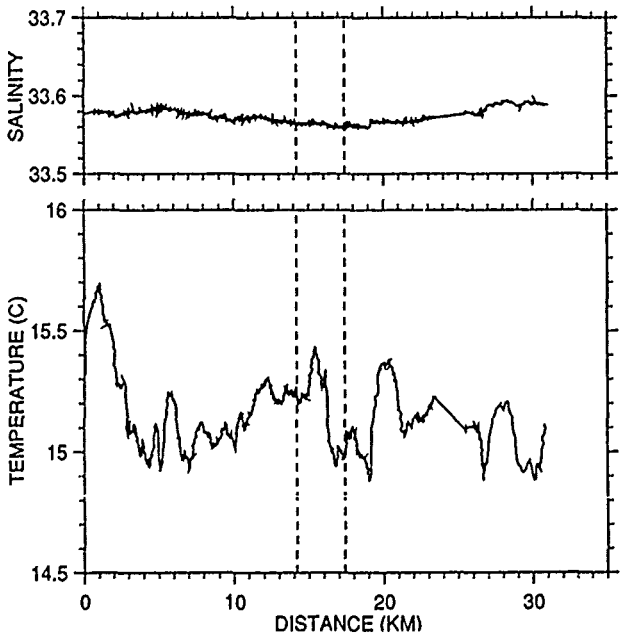


Fig. 5. Surface temperature and salinity sampled along the ship track, as a function of distance run. The turning points are shown by dashed lines.

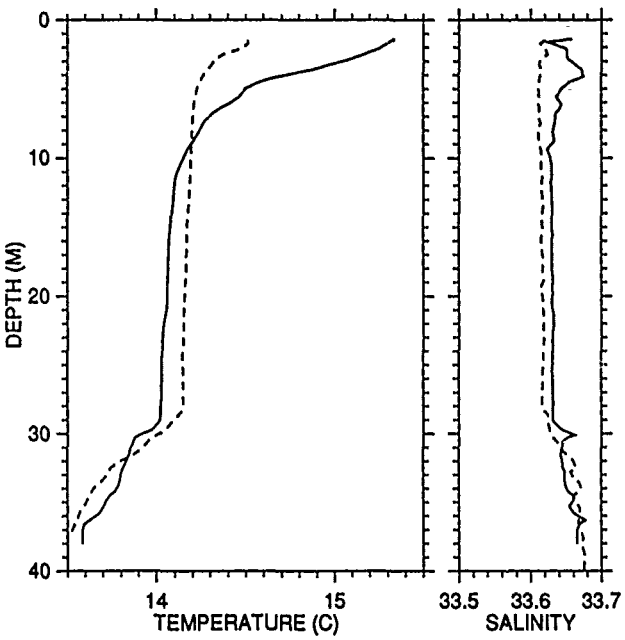


Fig. 6. Temperature and salinity profiles at approximately 1 km (solid) and 5 km (dashed) distance run.

4. Conclusion

We have shown, using a combination of satellite images and CTD profiles, that diurnal warming layers display coherent horizontal structures in the form of streaks at least 50 km long, with a wavelength of 4-8 km and an amplitude of 0.5-1.5°C. Although the survey was located 200 km to the southeast of the streaks seen in the images and was conducted 2 weeks later, we have all reasons to believe that the features observed were similar, given the steadiness of weather patterns during this period.

We have identified two similar events from satellite images and mooring data in April 1982 and in July 1988, which will be documented in a forthcoming paper. The coherent streaks appear quite common under low wind conditions in the California Current.

What processes may be responsible for these variations of surface temperature? The 4 to 8 km-wavelength scale of the streaks seems inconsistent with processes involving the upper ocean alone (i.e. Langmuir cells, interfacial waves, convective cells), which would scale with the mixed-layer thickness. Rather, this scale points to a coupling between the ocean and the planetary boundary layer, which may have a thickness of the order of a kilometer.

Helical circulation rolls occur frequently in planetary boundary layers (Brown, 1980). In general, they derive their energy from a combination of convective forcing and instability of the vertical shear. Their wavelength is typically 2 to 4 times the boundary layer thickness. Assuming pre-existing rolls in the atmosphere, the surface wind-stress is maximum underneath the axes of the rolls and minimum at the stagnation lines between adjacent rolls. Two scenarios are possible.

If surface temperature is principally governed by one-dimensional entrainment of a warm surface layer, cold anomalies will form underneath the axes of the atmospheric rolls and temperature streaks at one-half the wavelength of the rolls will appear. On the other hand, if surface temperature is principally governed by convergent and divergent advection of a thin warm surface layer, cold anomalies will form underneath the divergent stagnation lines and temperature streaks at the wavelength of the rolls will appear.

Since the cold (divergent) lines form underneath the downdrafts and the warm (convergent) lines form underneath the updrafts, the second scenario leads to an interesting feed-back mechanism further enhancing the atmospheric cells. Thus it is possible that the streaks are ultimately caused by a coupled ocean-atmosphere instability, without the need for pre-existing circulations. This conjecture, of course, will have to be tested by a purposely designed experiment.

References

Brown, R.A., "Longitudinal instabilities and secondary flows in the planetary boundary layer: a review," *Rev. Geophys. and Space Phys.*, vol. 18, pp. 683-697 (1980).

Bruce, J.G. and E. Firing, "Temperature measurements in the upper 10 m with modified expandable bathythermograph probes," *J. Geophys. Res.*, vol. 79, pp. 4110-4111 (1974).

Cornillon, P. and L. Stramma, "The distribution of diurnal warming events in the Western Sargasso Sea," *J. Geophys. Res.*, vol. 90, pp. 11811-11815 (1985).

Stramma, L., P. Cornillon, R.W. Weller, J.F. Price, and M.G. Briscoe, "Large diurnal sea surface temperature variability: satellite and in-situ measurements," *J. Phys. Oceanog.*, vol. 16, pp. 827-837 (1986).

Acknowledgements. This work has benefited from discussions with M. Abbott, C. Garrett, R. Garwood and J. Price. I thank D. Kelley for providing an outstanding plotting package. This work was supported by the Office of Naval Research through contracts N000014-80-C-0440 and N000014-87-K-007.

SATELLITE OBSERVED SPATIAL AND TEMPORAL VARIABILITY OF THE MIXED LAYER DEPTH IN THE SARGASSO SEA

Xiao-Hai Yan†, M. J. Bowman and J. R. Schubel

Marine Sciences Research Center
State University of New York
Stony Brook, New York 11794

ABSTRACT

Digital images representing the oceanic upper mixed layer depth in the vicinity of the Sargasso Sea have been generated from visible, near infrared, thermal infrared and microwave radiometers on board NOAA 7, 8 and Nimbus 7 satellites. Construction of such images requires a mixed layer remote sensing model, radiometric calibration and correction, geometric rectification, and registration of each image to a common grid. The mixed layer thermal inertia model, developed by Yan et al. (1988) and used in the construction of the mixed layer depth images, requires specification of the diurnal sea surface temperature range, sea surface wind speed and surface albedo. Our computed values of mixed layer depths are in closed agreement with the data from the LOTUS mooring deployed by Woods Hole Oceanogrphic Institution in the Sargasso Sea. The mixed layer depth images provide complementary information to conventional modeling and experimental observations for analyzing the spatial and temporal variability of mixed layer depths in the Sargasso Sea. This leads to a better understanding of the present state of development of mixed layer remote sensing models, and provides valuable input for the development of new remote sensing models of the oceanic mixed layer intended for climate studies and for describing the structure of the upper ocean.

1. Introduction

The ocean surface layer receives thermal energy from the absorption of solar and sky radiation and loses thermal energy by back radiation (longwave), evaporation, and sensible heat exchange at the air-sea interface. The net difference between these heat gain and heat loss terms is usually small compared to any one of the individual terms–that is, the heat budget for the near surface layer is, for most part, in near balance. However, this relatively small excess gain or loss of heat contributes significantly to the dynamics of circulation in the upper layers of the ocean and in the atmosphere. The net gain or loss of heat at a given location and a given time results in a rate of change in heat storage in the near surface layers, with a consequent change in the temperature and/or the thickness of the upper mixed layer. Heat is advected and diffused away from regions having a long term net gain in thermal energy through the air-sea interface toward regions having a long term net heat loss through this interface, and this process is a major contributor to the large scale thermo-haline circulation in the oceans, and also contributes a significant part of the transfer of heat from equatorial regions to the high latitude regions of the world oceans. The exchange of thermal energy between the upper mixed layer of the ocean and the atmosphere thus contributes significantly to the coupled ocean-atmosphere heat engine, and consequently in turn to large scale changes in weather patterns. Therefore, studies of the thermal structure of the surface layers of the oceans, the processes which occur in this boundary layer, and the exchange of heat and momentum between the ocean and the atmosphere are of considerable importance.

The response of the upper ocean to the vertical fluxes of heat and momentum at the surface has been studied by many authors (e. g. Clancy, et al., 1983, Davis, et al., 1981, Denman, 1973, Kraus, et al., 1976, Mellor, et al., 1974, Niller, et al., 1977, Price, et al., 1986) and has resulted in two distinct types of one dimensional mixed layer models. One type, known as an integral model, assumes that the actively mixing upper layer is homogeneous. Solutions are obtained from energy balance considerations in this uniformly well mixed layer. The second type, generally known as a turbulence closure model, uses a complete set of turbulence and thermal-dynamic equations to predict, over diurnal and longer time scales, the details of thermal and velocity structure in the upper ocean from oceanic boundary conditions. The disadvantage of such a model is the considerable time consumed in computation. The tests of these models are based on large numbers of conventional field measurements. For example, the traditional ocean thermal data base consists of approximately 1500 ship injection or bucket temperature measurements and 150 XBT measurements per day in the northern hemisphere (Clancy, et al., 1983).

The remote sensing of sea surface temperature from satellites offers climate researchers, oceanographers and atmospheric circulation modelers a very powerful source of data with which to monitor and study the global climate system and its two main components: the ocean and the atmosphere. A number of authors have investigated the various models of satellite-derived sea surface temperature. In the absence of cloud cover, the SST derived from remote sensing techniques now has attained an accuracy of about 0.1°K rms when compared to surface ship observations (McClain, et al., 1985). Satellite images, with their large aerial coverage at short sampling intervals, bridge the spatial and temporal gaps between hydrographic surveys and moored instruments. However, calibration problems, cloud contamination, the large quantity of data and difficulty in obtaining satellite data have frequently limited its use to only qualitative or semi-qualitative descriptions of physical properties. Although satellite sensors are routinely used to acquire information at all levels in the atmosphere, in the ocean, observations have been restricted to surface parameters except in some coastal areas. Therefore, the new challenge is to develop quantitative techniques for extrapolating satellite derived surface characteristics to infer subsurface oceanic conditions through a combination of these satellite derived observations and modeling.

The oceanic upper mixed layer depth (MLD) is an important physical property in the studies of upper ocean dynamics. There is as yet no method for the direct remote sensing assessment of the oceanic upper mixed layer depth and of the pattern of its evolution. However, the development of an upper ocean thermal inertia model designed to predict the thermal inertia of the mixed layer using satellite imagery may provide useful information for the estimation of the diurnal MLD. The aim of this study is to develop a mixed layer thermal inertia theory which will help in the development of a quantitative remote sensing approach (model) for prediction of subsurface information and for study of the upper ocean dynamics. Based on this remote sensing approach, diurnal mixed layer depth maps and thermal inertia maps are generated, allowing the study of the temporal evolution of the spatial patterns of the upper ocean diurnal mixing processes in the Northwest Sargasso Sea.

2.Mixed Layer Thermal Inertia Theoty and Model

Fundamentally, the principle of the mixed layer thermal inertia (MLTI) model depends on the fact that the diurnal sea surface temperature fluctuations (δSST) are primarily a function of the thermal inertia of the mixed layer and mixed layer thickness (diurnal average

† Presant Affiliation: Scripps Institution of Oceanography, Mail Code: A-030, University of California, San Diego, La Jolla, CA 92093

thickness), the local meteorological conditions (wind stress, mainly), heat flux, and physical characteristics of the sea surface (reflectance and albedo etc.). The value of δSST can be determined from satellite thermal infrared imagery, albedo can be calculated from satellite broad-band visible and near-infrared imagery, wind speed can be obtained from satellite microwave radiometer imagery and heat flux can be estimated using standard simplified method. Given these parameters, it is possible to determine the thermal inertia and diurnal mixed layer thickness.

a. Theory

Thermal inertia involves the temperature response of a body subject to a time-varying thermal energy flux at a surface (Price, 1977). Thus, if the surface heat flux is given by Q cos ωt (Q: the amplitude of heat flux, ω: frequency, t: time), then the difference in maximum and minimum values of the temperature at the surface is calculated to be (Price, 1977):

$$\delta T = 2\frac{Q}{(k\omega)^{1/2} C_p \rho} \tag{1}$$

where k is the diffusivity, ρ is the density, C_p is the heat capacity. We now define $(k)^{1/2} C_p \rho$ as thermal inertia:

$$(TI) \equiv (k)^{1/2} C_p \rho \tag{2}$$

Then (1) become:

$$\delta T = 2\frac{Q}{(\omega)^{1/2} (TI)} \tag{3}$$

It is evident from (3) that thermal inertia is a measure of the thermal response of a system to a time-varying thermal input at the surface.

The same concept can be applied to the ocean. As an analogy, we write:

$$(TI) = \zeta\frac{Q}{(\omega)^{1/2} \delta SST} \tag{4}$$

Where ζ is a dimensionless coefficient. The unit of thermal inertia is $Joule\ m^{-2} scc^{-1/2} K^{-1}$. The physical significance of this unit is that it is also approximate to a scale of heat flux per unit of temperature change. Thermal inertia (TI) is related to the depth of the influence of surface input, D as (Carslaw and Jaeger, 1956):

$$D \sim \left(\frac{k}{\omega}\right)^{1/2} \sim (TI) \tag{5}$$

This suggests that there should be a linear relationship between thermal inertia and the mixed layer depth (MLD).

In mechanics, inertia is a measure of the resistance of a substance to a change in its motion. Similarly, the thermal inertia of the mixed layer is a measure of the resistance of the water body to a change in its temperature resulting from heat gain and loss across the sea surface and from mixing processes. In other words, the thermal inertia of the upper ocean involves the temperature response of the water in the mixed layer subjected to a time-varying energy flux heat and momentum at the water surface. The concept of thermal inertia is very similar to that of heat capacity in solid bodies, although the diffusion processes in solids are molecular (much smaller compared with turbulent diffusion in the ocean). This means that if a substance has a large heat capacity, it usually has a large thermal inertia and a small range of temperature changes. In the ocean, since the spatial and temporal changes in heat capacity and density of sea water are relatively small, the thermal inertia is determined primarily by the spatial and temporal changes of the turbulent mixing (or eddy diffusion) coefficient, the k may vary largely in time and space along with the turbulent mixing processes. The square root of the turbulent mixing coefficient has length to the first power as one of its dimensions. This explains, in part, why, in areas of the ocean where, on a diurnal time scale, advective changes in temperature are small, thermal inertia has a high correlation with mixed layer depth.

The expression (1)shows that the thermal inertia of the mixed layer is proportional to the amount of solar energy absorbed by the surface layer, and inversely proportional to the diurnal sea surface temperature difference. The former term mentioned above in turn is sensitive to the sea water turbidity and surface albedo, and the latter term (δSST) is controlled mainly by the turbulent mixing processes (in mid ocean, wind mixing mainly). For example, if the heat flux at the surface (Q) increases while the sea surface temperature range (δSST) remains unchanged, this means that the thermal inertia of the water column is increased and that there must have been a strong surface wind stress which mixed the thermal energy down to a certain depth. When the heat flux is unchanged, the thermal inertia and the mixed layer depth are inversely proportional to the diurnal temperature range of the mixed layer. This is consistent with the definition of thermal inertia and with many observations (e.g. Price et al., 1986).

The reasons that we introduced the thermal inertia concept into the studies of oceanic upper mixed layer dynamics are two fold. First, we found that heat capacity does not provide an accurate description of the nature of temperature changes in the upper ocean as it does in the terrestrial cases, for sea water is to be considered as constant in the most of time, while in some ocean regions, diurnal temperature difference are large and in other regions small. The thermal inertia of the upper ocean such defined, however, can provide a better description of the nature of SST changes in a diurnal scale. Second, if we can map mixed layer thermal inertia using satellite data, there will be many important applications. The most obvious example is we can map diurnal mixed layer depth by remote sensing. In this study, mixed layer thermal inertia (MLTI) models have been developed to estimate the thermal inertia of the mixed layer using satellite data.

b. Model

The diffusion coefficient, heat capacity, and density of the sea water in the mixed layer are not measurable directly by remote sensing. Nor is the thermal inertia, but it can be inferred from measurements of the temperature change of sea surface water combined with a consideration of all the heating, cooling and atmospheric forcing processes which occur at the air-sea interface. The main purpose of the MLTI model is to express the thermal inertia and the mixed layer depth as functions of the quantities which can be measured directly or indirectly by available satellite, i.e., δSST, wind speed and surface albedo.

The simulation of the surface and near surface temperature structure is performed by analysis of the boundary conditions of an upper mixed layer of thickness approximately equal to the penetration depth of the diurnal temperature wave or diurnal trapping depth. Conduction and diffusion processes in the upper ocean which control the behavior of the temperature distribution of the upper ocean take the form of the conservation law:

$$\frac{\partial T}{\partial t} = k_t\frac{\partial^2 T}{\partial z^2} + \frac{\partial}{\partial z}\left(k_v\frac{\partial T}{\partial z}\right) - \frac{1}{C_p\rho}\frac{\partial I}{\partial z} - \frac{\partial}{\partial x}(u_a T) - \frac{\partial}{\partial y}(v_a T)$$

$$-\frac{\partial}{\partial z}(w_a T) + \frac{\partial}{\partial x}\left(k_h\frac{\partial T}{\partial x}\right) + \frac{\partial}{\partial y}\left(k_h\frac{\partial T}{\partial y}\right) \tag{6}$$

where T is the sea water temperature, k_t is the thermal diffusivity, k_v and k_h are vertical and horizontal turbulent diffusion coefficients, t is time, u_a, v_a and w_a are the x, y, z components of an advection current and I is the downward flux of solar radiation, I(z).

The studies of Camp and Elsberry (1978), Haney (1980) and others suggest that on shorter time scales (days to weeks), vertical mixing is the dominant process governing the response of the upper ocean. Since the Mixed Layer Thermal Inertia (MLTI) model, like all boundary layer models, is essentially one dimensional, and since we are concerned only with the upper ocean's relatively short time scale response (e.g. diurnal), and since the Sargasso sea is a relatively quiescent region, we can neglect the advection and horizontal diffusion of all quantities (i.e. all terms that involve k_h, u_a, v_a and w_a). This does not imply that lateral gradients are unimportant. It does mean, however, that diurnal variations in the thermal structure of the upper layer produced by horizontal advection and horizontal diffusion are assumed to be much smaller than those produced by vertical fluxes. The molecular thermal diffusion process is much smaller than turbulent mixing processes. Therefore, the term with could also be dropped.

Because the parameterized is dependent on Richardson number and turbulent kinetic energy, (i.e., velocity shear dependent), to complete the model it is necessary to include the momentum conservation equations.

It is well known that and do not have unique values and that they vary by many orders of magnitude in space and time. Hence, a suitable time and space dependent parameterization is necessary.

In our MLTI model the turbulent vertical diffusion coefficients are parameterized using the method suggested by Mellor and Yamada (1974) in their level 2 scheme. We chose Mellor and Yamada's parameterization to use in the MLTI model for two reasons. First, because it is a k theory scheme, which is required by the MLTI model. Most bulk or integral model schemes do not have a direct dependence on k, (e.g. Clancy, et al., 1983, Denman, et al., 1973, Kraus, et al., 1976, Niiler, et al., 1977, Price, et al., 1986). Second, because it has been widely used and successfully tested, and it is a part of the operational forecasting system TOPS (Thermodynamic Ocean Prediction System [see Clancy et al., 1983]). The main problem associated with using such a parameterization in remote sensing models is that it is time consuming. Although the Mellor and Yamada's parameterization is not entirely satisfactory, the possible replacement for it would be either a more complex and time consuming higher order turbulent closure model or a procedure based on Price's (1986) scheme.

To solve equation (6) for the surface temperature, which can be measured by remote sensing, we require a form of the solution appropriate to the diffusion equation and the boundary conditions. The major modeling problem is in parameterization of turbulent mixing and in defining the terms in the upper boundary conditions the wind stress and heat flux across the air-sea interface. We can not simply use in situ (LOTUS) measured Q in our remote sensing model, because one image covers about 20 degrees of latitude and because Q changes greatly from northern upper end to southern lower end of the images. Therefore, it is necessary to define each term carefully in order to estimate heat flux from place to place for remote sensing modeling. In this version of our model we have tried to use simple formulations for the various terms, based on various theoretical and empirical expressions available in the literature. The lower boundary condition is that the temperature is constant at some depth, z = d, where T (d,t) = T_d. The upper (surface) boundary condition is based on an assumption that a heat balance condition nearly exists at the surface. The incoming heat fluxes - the reflection corrected incident solar radiation and sky radiation must nearly balance the outgoing fluxes - the sensible heat flux in the air-sea interface, the latent heat flux in the air-sea interface and the long wave radiation emitted in the sea surface.

For the surface boundary conditions of momentum equations, the wind stress should be stipulated. For the simulation of the near surface and upper layer meteorological and oceanographic processes during LOTUS, the initial conditions of the temperature and velocity fields were obtained from a linear interpolation of the LOTUS mooring data at depth of 0.0, 5, 10, 15, 25, 50, 75, 100 m on May 13, 1982 with a constant wind stress equal to that at the starting point.

The non-linear form of the boundary condition prevents a direct analytical solution of the governing equation. An exact numerical solution can be calculated. However, the difficulty of illustrating the physical significance of terms and the computer CPU time costs all provide incentives for investigating less accurate but acceptable algorithms. The procedure for generating (or running) a MLTI model is to develop a simulation procedure which predicts the value of thermal inertia and mixed layer depth as a function of albedo, diurnal SST fluctuation (δSST) the meteorological condition (wind stress and heat flux). The two methods of simulation we used are Fourier series and finite difference algorithms.

The Fourier series method of handling the non-linear boundary condition is to linearize it and to express the solution to this condition in an exact form as an infinite Fourier series. The linearization is achieved by performing a Taylor series expansion of SST around the sky temperature and discarding quadratic and higher order terms. A major advantage of the Fourier series method is that it provides an analytical solution which can be computed rapidly compared with finite difference algorithms. However, it is limited since it is necessary to linearize all the upper boundary conditions in terms of SST in order to obtain a solution for the temperature and also since it is necessary to make several additional assumptions to simplify the problem. The details of this method and its applications have been discussed by Yan et al (1989).

The finite difference procedures are more flexible in handling a wider range of boundary conditions and complex subsurface thermal properties, but are inherently extravagant in computing time. This method requires solving the heat flow equation by differential equations and then numerically integrating forward in time. The result gives the temperature of sea water as a function of time and depth. In practice, a two-time step finite difference technique is used. At t=0 , an initial temperature profile is specified. A vertical grid of one meter increments from the surface to the depth, d, is used. For each subsequent time step, an iteration can be performed to determine the coefficient k. The temperature for the next time step, t=1, is then computed for each step in z from the finite difference approximation of equation(6). The complete set of solutions was obtained, i. e., the temperature measurements were determined as a function of depth and time. Based on these solutions, a table was generated providing a calibration measurement with a known value of thermal inertia, mixed layer depth and a given set of environmental factors (wind stress, surface albedo etc.). The procedure to construct a reference table is straightforward. The diurnal average mixed layer depth and diurnal temperature difference can be retrieved from these profiles. Model simulation condition is given and the thermal inertia is known (from the model). The inverse procedure is then used to express the thermal inertia and mixed layer depth (diurnal average depth) as functions of the diurnal temperature difference, wind stress and surface albedo. To complete the table, it is required to run the model over a complete set of all possible ranges of the boundary conditions. The initial conditions used by the model which generates a look up table for image processing is a specified profile with a generally exponentially decreasing temperatures. It should be mentioned here that although the look-up table only includes three variables (δSST, wind stress and albedo), it does not mean that other variables are not important. It does mean that δSST, wind stress and albedo were determined from satellite data, while other variables such as solar radiation, air temperature and humidity etc., were estimated in other ways by Yan et al. (1988). It may be possible in the future to use remotely sensed , and other terms in the heat flux equation in the model whenever they are available.

3. Image Processing

Application of the model results to the data from the meteorological satellite NOAA -AVHRR (visible, near -infrared and thermal infrared) and Nimbus-7-SMMR (microwave) permits the construction of mixed layer depth image maps and thermal inertia image maps. Briefly, mixed layer depth images and thermal inertia images are produced by the following procedures (Yan, et al., 1988). Day-time and night time temperature images are made from thermal infrared radiance data acquired by NOAA7 and NOAA8 AVHRR sensors. The data are rectified to account for such systematic effects as instrument optics, viewing geometry, and platform motion.

A similar procedure is followed to produce wind speed and albedo images. The images are registered to the standard grid. Next, a day-night SST difference, δSST is made by subtracting the night-time image values from the daytime values, on a pixel by pixel basis. The MLTI model is run repeatedly to generate a reference table of thermal inertia values as a function of sea surface wind, diurnal SST difference and surface albedo, and surface meteorological conditions.

To produce a thermal inertia image, each pixel is assigned a thermal inertia value from the model based on the δSST, albedo and wind speed for that pixel. Finally, the thermal inertia image is calibrated into the mixed layer depth image through a linear calibration equation. Details have been discussed by Yan, et al.(1989).

4. Model Results

The model was run for the period, May 14 to August 4, 1982, using both LOTUS buoy measured surface forcing data and satellite data, and based on the methods discussed in previous sections. Figure 1 shows the model/data comparisons, where the in situ diurnal mixing depth formed from the LOTUS buoy measurements. (Yan, et al., 1988)

A number of important conclusions can be drawn from these results. It can be inferred that the basic assumptions in running the MLTI model are correct and that vertical mixing is the dominant physical process that controls the distribution of thermal inertia in mid-ocean, far from major current systems. Although advection affects the thermal structure in many coastal waters, its role in determining thermal structure in most mid-ocean regions is relatively unimportant. Clearly, the one-dimensional MLTI model will be most useful in regions where horizontal advection is negligible.

There are some close relationships among mid-ocean diurnal surface warming, thermal inertia of the upper ocean, and surface forcing. Sea surface diurnal warming is an indicator of the thermal inertia in the upper ocean (inverse relationship). The thermal inertia in the upper mixed layer is controlled mainly by the surface energy flux- the heat flux and wind forcing, and proportional to the mixed layer depth.

During periods of relatively strong wind forcing and strong surface cooling (e.g., days 36 (June 18, 1982), 40 (June 22, 1982), 48 (June 30, 1982) and a period of days 35-55), the thermal inertia tends to increase and the mixed layer tends to deepen because water is entrained into the layer from below by turbulent mixing. Relatively strong wind stirring prevented solar radiation from significantly raising the sea surface temperature. On these days, the wind stress was larger than 0.11 pascals and net heat flux was less than 800 $W m^{-2}$.

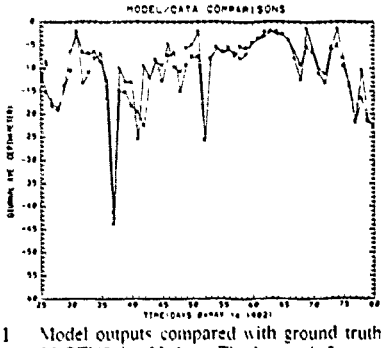


FIGURE 1 Model outputs compared with ground truth diurnal mixing depth of LOTUS for 55 days The line with * is model outputs The line with + is ground truth data

The resulting mixed layer depths were larger than 15 m and diurnal δSST ranged from 0.2°C to 0.3°C. If the mixed layer is shallow initially, the entrainment heat flux at the base of the mixed layer can be almost an order of magnitude larger than the surface heat flux. During periods of relatively weak wind forcing and strong surface heating (e.g., days 31 (June 13, 1982), 63 (July 15, 1982), 64 (July 16, 1982) and a period of days 60-65), however, the source of turbulent kinetic energy may have become too weak to maintain active entrainment at the base of the mixed layer, causing the thermal inertia to decrease and the mixed layer to retreat to a shallower depth. A combination of calm winds and clear skies permitted the appearance of rather pronounced stratification and diurnal warming. On the days 31, 63 and 64, the wind stress was less than 0.01 pascals and the net heat flux was larger than 900 Wm^{-2}. The resulting mixed layer depth was less than 5 m and diurnal δSST ranged from 1.8°C to 3°C. The thermal inertia and thermal structure of mixed layer are modified substantially on time scales of a few days by the passage of atmospheric disturbances, as evidenced by the data presented by Price et al. (1986), Clancy et al. (1981), the remote sensing results presented here and many others.

A daily variation in the temperature of the surface layer of the ocean is known to be produced by diurnal heating due to absorbed solar radiation. The amplitude of the daily temperature variation is usually small because of turbulent mixing which usually dominates the molecular thermal diffusivity. A solar radiation of 1000 Wm^{-2}, when absorbed in a mixed layer with thermal inertia of 15000 TIU (mixed layer depth about 10 meter) would give a heating rate of only 0.1 °C/hour and a daily variation of less than 0.5°C. If the turbulent mixing is reduced and a mixed layer with thermal inertia of less than 3000 TIU (mixed layer depth is about 1-2 meter thick) occurs, a heating rate of 1°C/hour may be expected and daily variations of several °C should be observed. With the exception of very shallow waters, large diurnal surface temperature variations in the open ocean (e.g., the Sargasso Sea) correspond to low wind speed because turbulence in the surface layer is mainly induced by the surface wind stress.

From May 14, 1982 to August 14, 1982, 25 synoptic thermal inertia and mixed layer depth images were mapped, of which 15 sets of measurements exhibited small thermal inertia (typically less than 5000 TIU) and very shallow mixed layer depth (typically less than 10 meters) for particular areas of approximately 1400 km long and up to 350 km wide, with an area of approximately 350,000 km^2. These results are very close to the observation and estimation of area of diurnal warming (exceeded 1°C) in the north-west Sargasso Sea by Stramma et al. (1986). The major feature demonstrated in the mixed layer depth images and thermal inertia images is the east-west oriented mixed layer depth and thermal inertia patterns (e.g. Yan, et al.,1988). These patterns usually last for several days in the northwest Sargasso Sea in summer time and coincident with the same east-west patterns of large diurnal sea surface warming events and wind patterns.

5.Significance

In this study, a unique Mixed Layer Thermal Inertia (MLTI) model has been developed to estimate the thermal inertia of the mixed layer using satellite data. Temporal variations in the spatial distribution of the thermal inertia of the mixed layer provide information applicable to studies of variations in the upper ocean diurnal mixed layer depth in regions of the ocean, where diurnal variations in the thermal structure of the upper layer due to horizontal advection and horizontal diffusion are small, and where the effects of salinity structure on variations in the diurnal mixed layer structure are small. The mixed layer depth and the thermal inertia of the mixed layer estimated by the MLTI model using remote sensing data opens up many potential areas of application in physical, biological and geological oceanography as well as in atmospheric investigations. Although these methods must still be regarded as experimental, and much work remains to be carried out in the investigation, interpretation, and refinement of the models and of image processing, the thermal inertia imagery and mixed layer depth imagery offer oceanographers a new quantitative remote sensing approach for the study of oceanic upper mixed layer dynamics, for the prediction of subsurface information and which gives a new view of the upper ocean.

Midday solar heating of the upper ocean, especially under conditions of light wind and clear skies, tends to decrease the thickness of the mixed layer and to increase stratification of the upper layer. When the stratification of the upper layer becomes greater than about 0.001°C/m, the surface acoustic duct is destroyed. This produces the "afternoon effect" which is important in some acoustic surveillance applications (Clancy, et al., 1983). The digital thermal inertia images and mixed depth images produced by this study may have applications in predicting the region of the "afternoon effect".

Because the thermal inertia technique provides information on the relationships among heat flux, wind speed and SST changes within the diurnal cycle, there are possibilities that with the same formulation one can make an estimate of wind speed or heat flux by inverting the other two using an iterative methods. Application of the simplified scaling thermal inertia model to the data from NOAA-7 and 8 satellites also permits the construction of surface wind speed images of the North west Atlantic Ocean. Yan et al., (1988) These images were compared with SMMR wind images. The thermal inertia wind image was shown to have a reasonablely good fit with the SMMR wind image and all other comparison groups show the same close similarity of wind speed patterns between the SMMR wind and thermal inertia wind images. The primary importance of this study shows that a successful combination of research on model studies with remote sensing digital image processing techniques provides new information for the study of upper ocean dynamics and air-sea interactions. That will, in turn, improve the value of remote sensing of the oceans. In this study, the models and image processing results demonstrate a close relationship among the mid-ocean diurnal surface warming, the thermal inertia of the upper ocean, the diurnal mixed layer depth and the surface forcing. We believe that this remote sensing model is a useful tool to study the upper ocean dynamics and might be applied to operational upper ocean forecasting.

This work has been partially supported by NASA through grant No. NASA NAGW-1271 (through Scripps Institution of Oceanography).

REFERENCES

Camp, N. T., and Elsberry, R. L. (1978). Oceanic thermal response to strong atmospheric forcing, the role of one-dimensional processes, J. Phys. Oceangr. 8:215-224

Carslaw, H.S., and J.C. Jaeger, (1959). Conduction of heat in solids, Oxford, London, 510P.

Clancy, R.M. and Pollark, K.D. (1983). A real-time synoptic ocean thermal analysis/forecast system, Prog. Oceangr. 12:383-424

McClain, E.P. (1981). Multichannel improvements to satellite-derived global sea surface temperature. Adv. Space Res. 2:43-47

Mellor,G.L., and Yamada, (1974). A hierarchy of turbulent closure model for planetary boundary layers J. Atmos. Sci., 31.1791-1806

Niler, P.P., and Kraus, E.B., (1977). One-dimensional models of the upper ocean, in Modelling and Predictions of the Upper Layers of the Ocean, edited by Kraus, E.B., PP 143-172 Pergamon, Elmsford N.Y., 1977.

Price, J.C. (1977). Thermal inertia mapping: a new view of the earth. J. Geophys. Res. 82:2582-2590

Price, J.F. and Weller, R. A. (1986). Diurnal cycling: observations and models of the upper ocean response to diurnal heating, cooling and wind mixing J. Geophys. Res. 91:8411-8427

Stramma, L, and Cornillon, P. (1986). Large diurnal sea surface temperature variability: satelite and in situ measurements J. Phys. Oceangr., 16:827-837

Yan, X-H., Schubel, J.R., and Pritchard, D. W. (1988). Oceanic upper mixed layer depth determined from satellite data J. Geophys. Res. (submitted)

Yan, X-H., Okubo, A., Schubel, J.R., and Pritchard, D. W. (1989). An analytical mixed layer remote sensing model J. Geophys. Res. (submitted)

SPECTRAL ANALYSIS OF THE AVHRR SEA SURFACE TEMPERATURE VARIABILITY OFF THE WEST COAST OF VANCOUVER ISLAND

Ralf Burgert and William W. Hsieh

Department of Oceanography, University of British Columbia, Vancouver, B.C., Canada V6T 1W5

Abstract:

From 16 AVHRR infrared satellite images off the west coast of Vancouver Island, British Columbia, collected during the five summers of 1984-1988, four alongshore temperature transects were sampled. Upon Fourier transforming the transect data, we found that the energy spectra of the temperature variance in alongshore wavenumber space in general followed a -2.1 power law, which agreed with previous observations from other parts of the world.

Summer images may be divided into 2 types: Upwelling dominated and non-upwelling dominated. When a strong upwelling-induced alongshore cold front was observed, the regimes shoreward and seaward of the front had distinctly different spectra. Cross spectral analysis of transect data between images taken a day apart in the presence of strong upwelling events revealed significant coherence at the low wavenumber regime (wavelength 300 km and above, corresponding to the large eddies) and often at the high wavenumber regime (wavelength 30 km or below, corresponding to the fine structures of the eddies). The coherence dropped for images taken 2 or more days apart, suggesting a decorrelation time scale of about 2 days.

Keywords: Spectral analysis, sea surface temperature

1. Introduction:

The spatial scales of sea surface temperature (SST) variability can be derived readily by remote sensing. McLeish (1970), Saunders (1972a), and Holladay and O'Brien (1975) took transects of SST using a radiometer from an airplane. Lutjeharmes (1981) used satellite images for this purpose, but measured the sizes of features on the image directly.

Advances in computer image processing now allow transects to be taken directly from satellite images. Being one-dimensional, such transects are easily analyzed, and unlike ship transects, long transects by satellite can be made at essentially one instant. Deschamps et al. (1981) was the first to apply this technique to the problem of SST variability by calculating the spatial structure functions of transect data. The spatial structure function has been the preferred technique in the past as it shows less variation from noise than the Fourier spectrum. However, the quality of satellite images available today is much higher, due not only to technical and processing advances, but also to the larger collection of image data. It is now practical to find the Fourier spectrum of satellite data if only very clear images are selected.

Hence this paper differs from previous work in two main ways: Fourier spectra were calculated rather than spatial structure functions, and high quality images were analyzed over a much longer period. Using an approach similar to Ikeda et al. (1984), we sampled transects of length 357 km parallel to Vancouver Island, at various distances offshore, from a set of 16 images over five summers and studied their spectral properties.

The summer surface flow on the continental shelf off Vancouver Island is predominantly southerly (Freeland and Denman, 1982), except close to the coast where the flow is less well defined as outflow from Juan de Fuca Strait tends to flow northward as a coastal current (Weaver and Hsieh, 1987). The local wind conditions generally favour upwelling, with a large upwelling eddy frequently observed at the mouth of the Juan de Fuca Strait. Fluctuations in the wind leads to individual strong upwelling events which usually last a few days.

2. Data Acquisition and Processing:

Our satellite image data for the five summers from 1984 to 1988 (taken by the NOAA 6, 7, 8, and 9 satellites) were obtained from the Satellite Oceanography Laboratory archive at the University of British Columbia. Only images having two or more cloud-free days within a week were chosen, so that the evolution of the SST patterns could be monitored. Selected images (July 14,15,16,17,18, 1984; August 13,15,16,17, 1985; August 6,14, 1986; August 21,22, 1987; August 21,22,23, 1988) were sampled along lines shown in Fig. 1, which also shows the coastline and gross features of the bottom topography. All images have been previously processed at different scalings after the radiation to temperature conversion. However, these differences, as well as deviations from bucket temperatures in the satellite data, are acceptable as we are studying only the variability of the temperature, not their absolute values.

The transect SST data were linearly detrended and Fourier transformed. Spectra from several images taken within a week were assembled and averaged (except where used in analyses of shorter period variability in Section 4), and were then further smoothed by a moving-average filter to increase statistical confidence.

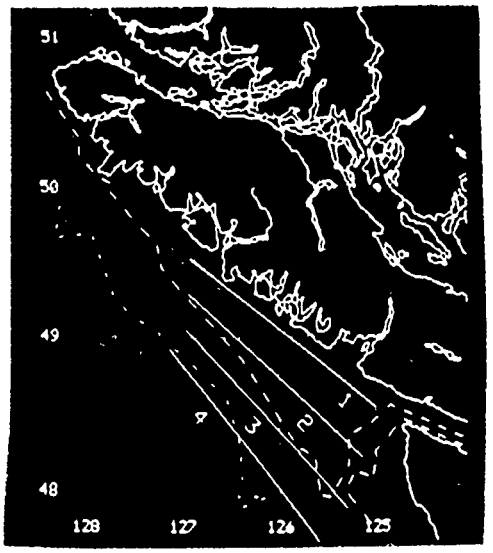


Fig. 1. The West coast of Vancouver Island: The transects (solid lines) and some gross features of the bathymetry. Inshore dotted line is the approximate 180 m isobath, and offshore dotted line, the 1800 m isobath.

3. Spectral Analysis:

Fig. 2 shows the energy spectra averaged over all available images on a log-log scale. The mean of the least-squares fits to all the transects is also shown. The slope of this fit implies that the one-dimensional energy spectrum E(k) follows a power law of k^n with n = -2.1 ± 0.2. This agrees well with the results of Saunders (1972a) who obtained a value of -2.2 ± 0.1 for the same measurement in the Mediterranean, and those of Deschamps et al. (1981 and 1984), who obtained a value of n = -1.8 ± 0.1 for the same part of the world, and a value of approximately -2 in 1984 off the coast of France. It agrees less well with Holladay and O'Brien (1975) who reported a value of n = -3 for the upwelling region off Oregon. From LANDSAT multispectral scanner band 4 data (corresponding roughly to green light), Gower et al. (1980) also reported a value of n = -2.92 south of Iceland. Saunders mentions that the value -2 does not agree with any theoretical values of two-dimensional turbulence, citing values of -1, -3, and -5 predicted by various studies (Saunders, 1972b). The closest theoretical values are those of Kraichnan (1971), who predicts values of -3 for enstrophy transfer and -5/3 for energy transfer. It is possible that the area of this study, and that of Saunders' and Deschamps', are not entirely dominated by isotropic turbulence. For instance, coastal topography and currents in our region will modify the

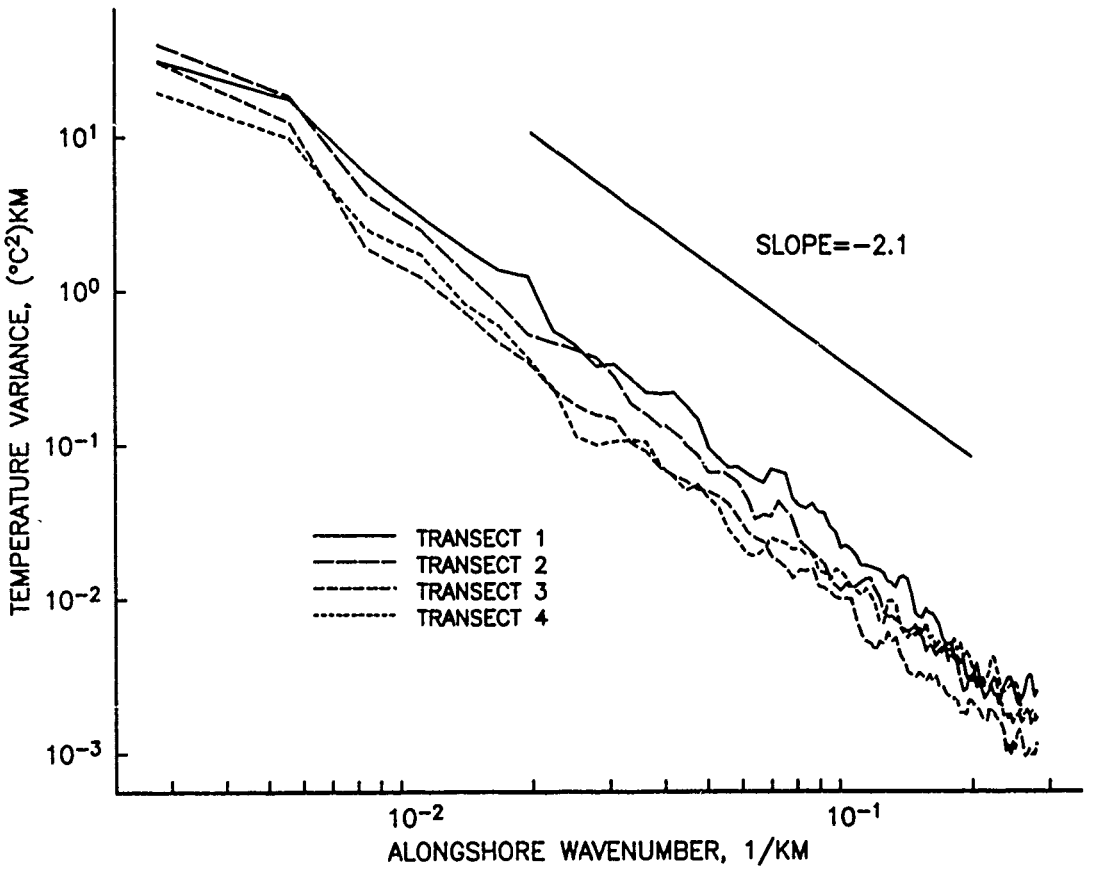


Fig. 2. The 5-summer mean SST energy spectra for the four transects with best-fit slope of -2.1 shown as a straight line.

turbulence from simple two-dimensional isotropic turbulence. However, since the circulation patterns of the three regions are vastly different, it seems unlikely that such good agreement in the observed values could be due to any common feature of the regions except the presence of two-dimensional turbulence.

4. Discussion:

Cross-spectra of transects from different images were also computed. Fig.3 shows the coherence and phase spectra between two transect 4 temperature series from images received on Aug. 21 and 22, 1988, which reveal typical summer upwelling event. The coherence spectrum reveals significant coherence at the very low wavenumber regime (below about 0.02 km^{-1}) and at the high wavenumber regime (above roughly 0.2 km^{-1}). The intermediate regime between the two reveals generally insignificant coherence for images only a day apart.

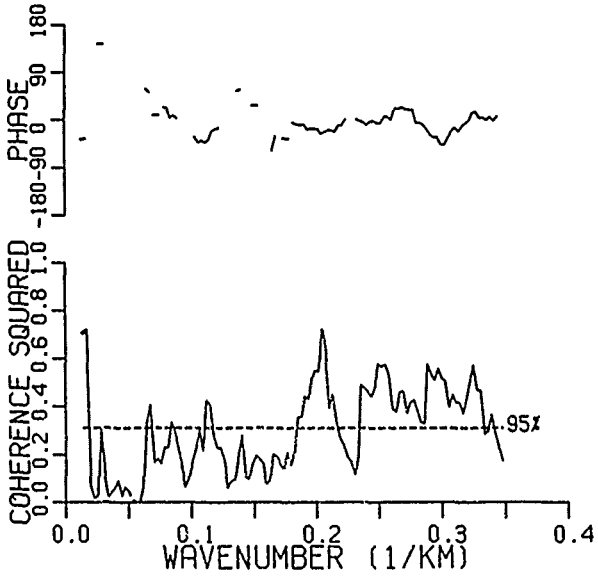


Fig. 3. Cross-spectrum between transect 4 temperature data from Aug. 21 and 22, 1988. The 95% significance level is shown as a dashed line. The phase spectrum is plotted only when the coherence is above the 80% significance level.

To understand the coherence, we look at the original transect 4 data shown in Fig.4 for Aug.21, 22 and 23, 1988, which shows the gradual warming of a cold upwelling event over the 3 days. The dominant wavelength as measured from the two troughs is about 250 km, which generally agrees with the coherence peak at just under 0.02 km^{-1} in Fig.3, (corresponding to a wavelength of about 350 km). There is also southward propagation seen from Fig.4 and from the phase spectrum of Fig.3, where the phase lag of 37 degrees gives a southward phase speed of 36 km day^{-1} or 40 cm s^{-1}, a speed roughly comparable to the prevalent summer current flowing southward off the shelf (Thomson, 1981, Fig. 13.19). Closer examination of Fig.4 reveals that some of the smaller scale structures, e.g. at the 30-km scale, remain distinguishable throughout the period, thereby contributing to the rise of coherence for wavenumber greater than about 0.2 km^{-1} (corresponding to the 30-km scale) in Fig 3.

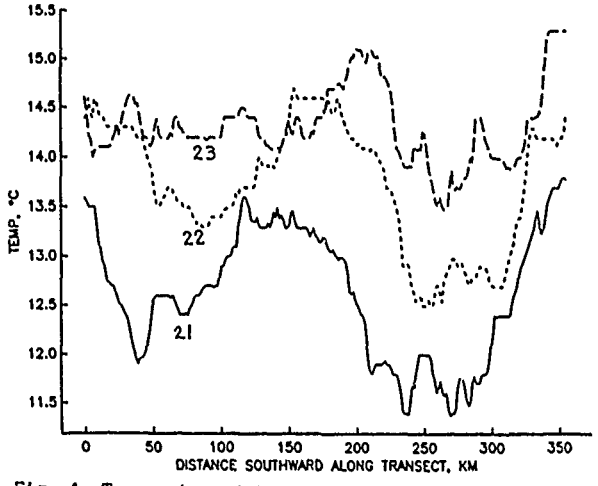


Fig. 4. Temperature data along transect 4 for Aug. 21 (solid curve), Aug. 22 (dotted) and Aug. 23 (dashed), 1988.

Fig. 5 shows transect 4 data from Aug.17, 1984 and from the same day one year later. As both days were characterized by the absence of strong upwelling events, the resulting data are in sharp contrast to those in Fig.4. If we conceptually divide each transect temperature into three parts: the overall mean, the linear trend and the eddy fluctuations, we notice the absence of strong eddy component in Fig.5 data as compared to Fig.4 data. Curiously, the two transect data in Fig.5 show that despite having a similar alongshore linear trend (a rise of 1.5 °C over 350 km southward), the overall means are about 2.5 °C apart. Thus, for summer transect data from different times, in the absence of strong upwelling events, the general linear trend tends to persist despite a possible major shift in the mean.

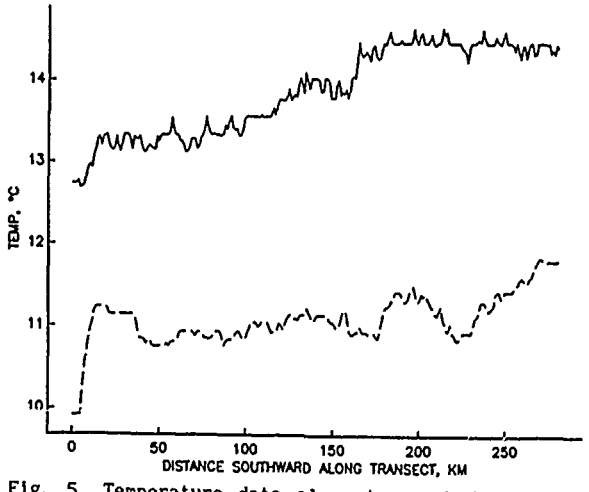


Fig. 5. Temperature data along transect 4 for Aug. 17, 1984 (solid curve) and Aug. 17, 1985 (dashed curve).

5. Conclusions:

From alongshore SST transects taken from AVHRR images spanning five summers off the west coast of Vancouver Island, we have confirmation of Saunders' (1972a and b) and Deschamps' et al. (1981 and 1984) measurement of the energy spectra power law exponent n. The good agreement with these values, obtained from very different parts of the world, suggests that the value obtained does indeed represent the presence of two-dimensional turbulence, despite our poor theoretical understanding of such processes in a non-isotropic environment (e.g. presence of coastal topography and currents).

Summer images may be divided into 2 types: Upwelling dominated and non-upwelling dominated. When a strong upwelling-induced alongshore cold front was observed, the regimes shoreward and seaward of the front had distinctly different spectra (Burgert and Hsieh, 1989). Cross spectral analysis of transect data between images taken a day apart in the presence of strong upwelling events revealed significant coherence at the low wavenumber regime (wavelength 300km and above, corresponding to the large eddies) and often at the high wavenumber regime (wavelength 30 km or below, corresponding to the fine structures of the eddies). Between these two regimes, the coherence is generally insignificant except for occasional minor peaks. The coherence drops for images taken 2 or more days apart, suggesting a decorrelation time scale of about 2 days. In the absence of strong upwelling and associated eddies, summer transect temperature data from different years often revealed a similar alongshore linear trend in addition to possible large differences in the mean temperature.

Acknowledgment:

We would like to thank Andy Thomas for the use of some of his images and software, and Denis Laplante for his help. This research is supported by the Natural Sciences and Engineering Research Council of Canada and the Department of Fisheries and Oceans.

References:

1. McLeish, W. 1970. Spatial Spectra of Ocean Surface Temperature. Journal of Geophysical Research. 19: 6872-6877.

2. Saunders, P.M. 1972a. Space and Time Variability of Temperature in the Upper Oceans. Deep Sea Research. 19: 467-480.

3. Holladay, C.G. and J.J. O'Brien. 1975. Mesoscale Variability of Sea-Surface Temperature. Journal of Physical Oceanography. 5: 761-772.

4. Lutjeharmes, J.R. 1981. Spatial Scales and Intensities of Circulation in the Ocean Areas Adjacent to South Africa. Deep Sea Research. 28: 1289-1302.

5. Deschamps, P.Y., R. Frouin and L. Wald. 1981. Satellite Determination of the Mesoscale Variability of Sea Surface Temperature. Journal of Physical Oceanography. 11: 864-870.

6. Ikeda, M., L.A. Mysak and W.J. Emery. 1984. Observations and Modelling of Satellite Sensed Meanders off Vancouver Island, Journal of Physical Oceanography. 14: 3-7.

7. Freeland, H.J. and K.L. Denman, 1982. A topographically controlled upwelling center off southern Vancouver Island. J. Mar. Res. 40: 1069-1093.

8. Weaver, A.J. and W.W. Hsieh. 1987. The influence of buoyancy flux from estuaries on continental shelf circulation. J. Phys. Oceanogr. 17: 2127-2140.

9. Deschamps, P.Y.; R. Frouin and M. Crepon. 1984. Sea Surface Temperatures of the Coastal Zones of France Observed by the HCMM Satellite, Journal of Geophysical Research. 89: 8123-8149.

10 Gower, J.F.R., K.L. Denman and R.J. Holyer. 1980. Phytoplankton patchiness indicates the fluctuation spectrum of mesoscale oceanic structure. Nature. 288: 157-159.

11 Saunders, P.M. 1972b. Comments on Wave Number-Frequency Spectra of Temperature in the Free Atmosphere. Journal of Atmospheric Science. 29: 197-199.

12 Kraichnan, R.H. 1971. Inertial Range Transfer in Two and Three Dimensional Turbulence. Journal of Fluid Mechanics. 47: 525-535.

13 Thomson, R.E. 1981. Oceanography of the British Columbia coast. Can. Spec. Publ. Fish. Aquat. Sci. 56: 291p.

14 Burgert, R. and W.W. Hsieh, 1989. Spectral analysis of the AVHRR sea surface temperature off the west coast of Vancouver Island. Atmos.-Ocean (in press).

SEA SURFACE THERMAL SIGNATURE AND ESTUARINE FLUSHING EVENTS IN THE STRAITS OF GEORGIA AND JUAN DE FUCA.

P.H.LeBlond, D.A.Griffin and D.Laplante

Department of Oceanography, University of British Columbia, Vancouver, B.C. Canada V6T 1W5

Abstract

Significant amounts of fresh water are exported seawards from the Strait of Georgia through the Strait of Juan de Fuca when northwesterly winds coincide with periods of reduced tidal mixing. These events are accompanied by a recognizable sea surface temperature signal and usually occur during clear-sky conditions: they may thus be examined through thermal infrared satellite imagery. NOAA AVHRR images of the region will be used to discuss and examine the flushing events.

Introduction

Freshwater runoff has an important influence on coastal oceanographic conditions. The lighter fresh water floats upon and gradually mixes with sea-water, affecting vertical stability and mixing,and hence the location of fronts between stratified and homogenized regions. The estuarine circulation caused by runoff is characterized by seaward flow in a relatively thin (5-20 m) surface layer, partly compensated with a slower return flow at depth. This type of circulation in fjords and other estuaries is already well documented (Dyer, 1973). Estuarine influences often prevail well onto the continental shelf and can give rise to significant buoyancy-driven nearshore currents, as on the coasts of Norway, Scotland, and western North America (Skreslet, 1986).

Wind forcing and tidal mixing are the main factors influencing the seaward export of fresh water. The wind pushes surface waters roughly in the direction in which it blows, with some degree of deflection to the right (in the northern hemisphere) due to the Coriolis force, the exact angle depending on the wind's duration, the presence of lateral constraints and the vertical stratification. The main effect of tidal currents is to contribute to vertical mixing. In shallow areas where tidal currents are sufficiently strong, turbulent mixing may destroy the vertical density stratification. Fronts occur at the transition between stratified and mixed regions. Tidal currents ebb and flow with semi-diurnal and diurnal periodicity as well as over longer periods: the vertical stratification and the position of fronts may thus be expected to show a similar variability.

Observations in channels connecting sources of runoff with the ocean (Geyer and Cannon, 1982; Nunes and Lennon, 1987) have shown that the variation of tidal currents which occurs on a fortnightly cycle of neap and spring tides is particularly important in modulating the estuarine circulation and water exchange between coastal basins and the ocean.

This paper describes the influence of wind forces and tidal mixing on the export of freshwater from the Strait of Georgia, through the narrow and shallow tidally mixed Boundary Passage, to Juan de Fuca Strait and the northeast Pacific Ocean, and discusses the application of AVHRR infrared imagery to the detection and interpretation of the phenomenon. The geographical area is shown in Fig.1,wherein place names referred to in the text are indicated.

Surface Salinity in Juan de Fuca Strait

The tidal mixing rate is proportional to that of energy loss through bottom friction, and hence to the cube of the flow speed. A comparison of the cube-root-mean-cubed (crmc) tidal velocity in Boundary Passage to fluctuations of surface salinity at Race Rocks shows a clear relationship (Fig.2). A time series synthesized from the long-period tidal components MSf (fortnightly) and Mm (monthly) of Race Rock surface salinity variations (middle line, Fig.2) exhibits most clearly the relationship with mixing in Boundary Passage, with minimum salinity occuring two to four days after times of minimum mixing (neap tides), indicated by dashed lines.

The combined effects of discharge, tidal mixing and wind on surface salinity in the

southern Strait of Georgia and eastern Juan de Fuca Strait are shown in Fig.3. The effect of the wind on the salinity at Race Rocks is felt clearly when a northwesterly wind (from 310°T) pushes the surface waters of the Strait of Georgia seawards during a period of weak tidal mixing. The combined effects of wind and tidal mixing are represented by a Froude number, low values of this number correspond to strong stratifications during low current regimes, when the surface brackish layer may readily slip seawards (Griffin and LeBlond,1989). The lowest curves compare actual surface salinity at Race Rocks with their low frequency tidal variation. As seen already in Fig 2, the low frequency tidal contribution explains the actual salinity variations most of the time, but clearly NOT when a northwesterly wind coincides with neap tides during the high discharge period. These events are well predicted by the Froude number and occur when the latter falls below a value of approx. 0.8.

Remote sensing

The mechanisms described above account for surface salinity variations throughout a wide area of coastal waters from Boundary Passage to the Vancouver Island shelf (Hickey et al, 1989). A good description of the spatial variability of the process requires a denser network of salinity monitoring stations. Satellite remote sensing would seem an ideal method of obtaining the desired aereal coverage but

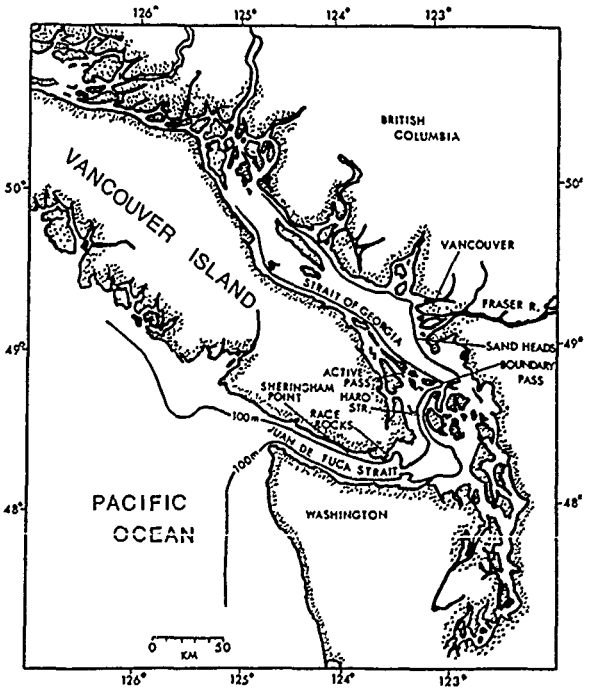


Fig.1 Area map, showing the main source of freshwater runoff (the Fraser River),and the channels through which it flows seawards.

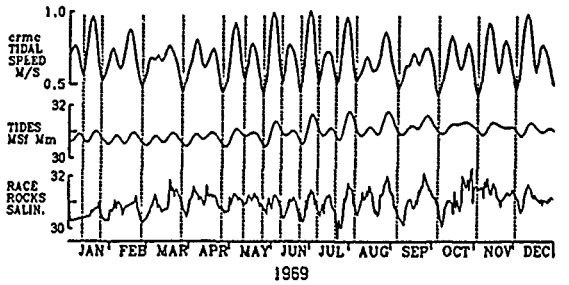


Fig.2 Time series of cube-root-mean-cubed (crmc) tidal currents in Boundary Passage compared with variations of surface salinity at Race Rocks (bottom) and the latter's low frequency tidal part (middle curve).

for the fact that there is currently no method of extracting salinity directly from emission or reflection spectra of the sea surface. During the summer however inshore waters are thermally stratified, with warm and fresher surface water overlying deeper, colder and saltier water. Temperature may then be used as a proxy variable for salinity: strong mixing will be associated with cold as well as salty water at the surface; stratified waters will be warm as well as brackish at the surface. Thermal imagery of the sea surface can then reveal the evolution of fresh water flushing processes.

NOAA-7 AVHRR thermal imagery with resolution of 1.1 km was obtained from the University of British Columbia's Satellite Oceanography Laboratory. Northwesterly winds over the coastal area usually coincide with the presence of a high pressure system so that clear skies prevail during potential flushing conditions. Oceanographic conditions observed during the summer of 1984 are shown in Fig 4. A period of coincidence of neap tides with a northwesterly wind over the southern British Columbia coast is seen to occur in late July. It is clearly associated with an increase in temperature and a decrease in salinity (not shown) at Race Rocks. A sequence of surface temperature maps were obtained overlapping the period of interest. Interpretations, redrawn from calibrated and color-coded infrared imagery to show the main thermal features of the sea surface,are shown in Fig.5.

The first image (July 14, 1984) shows the surface temperature field soon after spring tides and on the first day of a reversal in wind direction following nearly a week of southeasterlies. Cold (< 15°C) surface waters are seen throughout Juan de Fuca Strait and Boundary Passage; the southern Strait of Georgia is warm (between 15 and 20°C) and the northern end even warmer (>20°C).

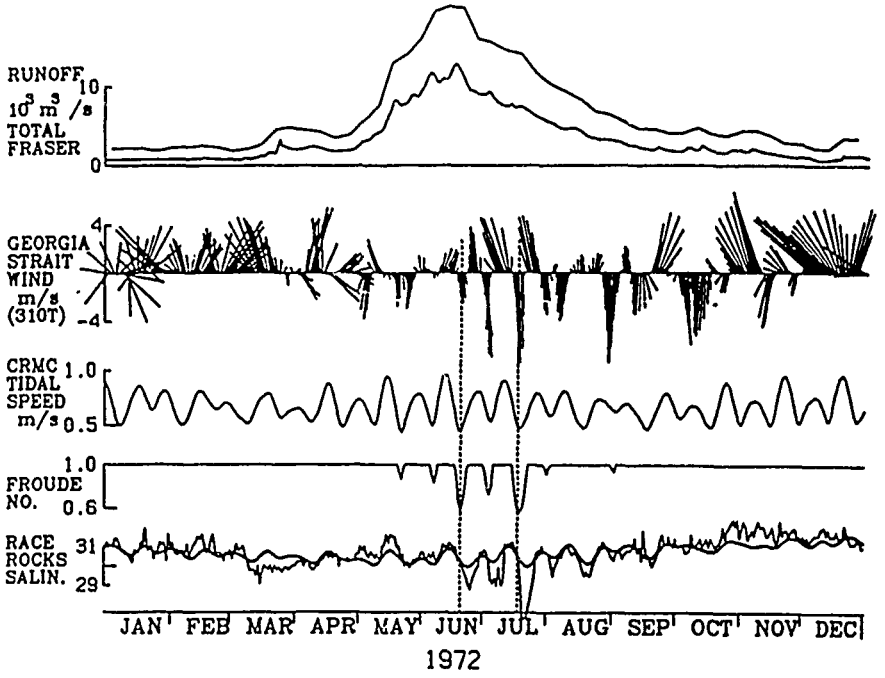


Fig.3 Discharge, wind and tidal mixing effects on surface salinities at Race Rocks. From the top: freshwater runoff into the Strait of Georgia (total and Fraser River contribution); wind vectors at Sand Heads (rotated: up is 310°T, along axis of Strait of Georgia); crmc tidal speeds in Boundary Passage; Froude number characterising mixing in Boundary Passage; surface salinity at Race Rocks (smooth curve is the low-frequency tidal part).

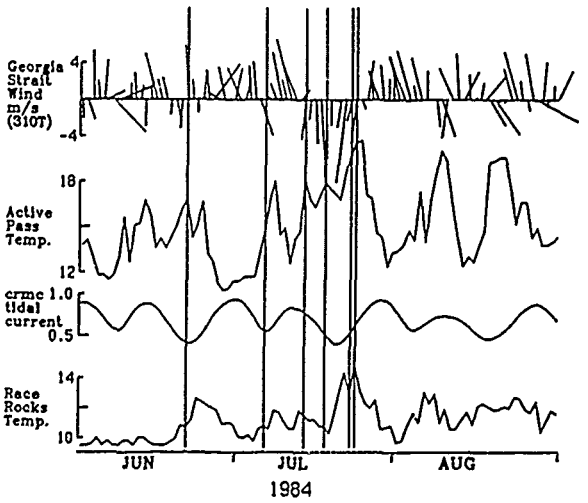


Fig.4 Oceanographic and meteorological conditions pertinent to freshwater export from the Strait of Georgia during the summer of 19894. Vertical lines indicate times of satellite images. The images seen in Fig.5 are taken on the last four days shown.

After four days of strong northwesterlies (July 18), pronounced upwelling has brought cold water to the surface on the west coast of Vancouver Island (temperatures below 10°C); upwelling has also taken place in the northeastern part of the Strait of Georgia, where patches of surface water of temperatures below 15°C are now seen. Tidal mixing is weaker and warm water is beginning to be pushed through Boundary Passage without being completely mixed with the colder water beneath.

Five days later (July 23), warming under sunny skies has brought the temperature of the surface waters of the Strait of Georgia back to 20°C and above. Evidence of the lack of mixing during neap tides is seen in the warm surface waters of Boundary Passage and eastern Juan de Fuca Strait. Mixing quickly dominates again: the next day (July 24), surface temperatures in that same area have already decreased below 15°C. A pulse of warm surface water is seen to progress seawards in Juan de Fuca Strait.

Conclusions

We have used sea surface temperature as a proxy variable for salinity to describe an important feature of the export of freshwater seaward from the Strait of Georgia. Satellite imagery provides wide-area coverage of the process and opens the door to remote monitoring and short-term prediction of oceanographic conditions in that area.

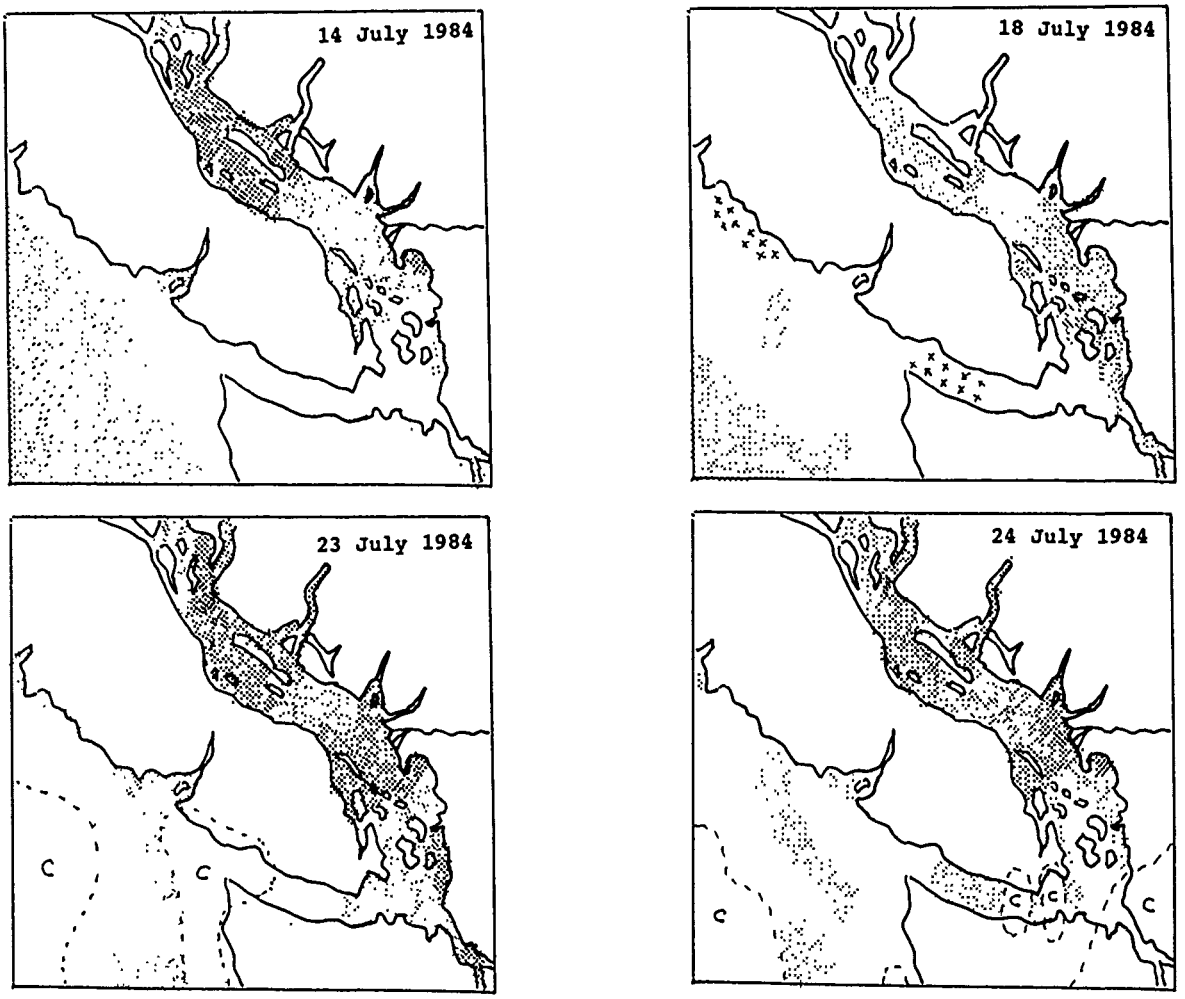


Fig 5. Sketches of sea surface temperature distribution in southern British Columbia waters in July 1984. Areas labelled C are cloudy. In order of increasing temperatures, areas hatched with crosses are coldest, below 10°C, those left blank are below 15°C, those lightly stipled, below 20°C and heavily stipled areas are above 20°C.

Bibliography

Dyer,K.R. Estuaries. J.Wiley, New York, 140 pp. 198

Geyer,W.R. and G.A.Cannon." Sill processes related to deep-water renewal in a fjord", J.Geophys.Res.,87 pp 7985-7996,1982.

Griffin,D.A. and P.H.LeBlond."Estuary/ocean exchange controlled by spring-neap tidal mixing", Estuarine,Coastal and Shelf Science, (in press).

Hickey,B.M.,R.E.Thomson,H.Yih and P.H.LeBlond."Wind-and buoyancy-driven fluctuations in the Vancouver Island coastal current." J.Geophys. Res. (submitted).

Nunes,R.A.and G.W.Lennon."Episodic stratification and gravity currents in a marine environment of modulated turbulence", J.Geophys.Res.,92,pp 5465-5480,1987.

Skreslet,S.,(Editor) The Role of Freshwater Outflow in Coastal Marine Ecosystems. Springer, 453 pp, 1986.

WATER QUALITY MAPPING OF AUGUSTA BAY, ITALY FROM LANDSAT-TM DATA

Siamak Khorram and Heather Cheshire

Computer Graphics Center
North Carolina State University
Raleigh, U.S.A.

Alberto L. Geraci and Guido La Rosa

Istituto di Macchine, Facolta' di Ingegneria
Universita' di Catania, Catania, Italy

ABSTRACT

Landsat TM digital data were used to map the distributions and concentrations of selected water quality indicators in and around August Bay, Sicily. The general approach involved near simultaneous acquisition of TM data and water quality samples from 42 sites, laboratory analysis of samples, extraction of sample site DNs from the TM data, development and validation of regression models based on sample data, application of models to the entire study area, and generation of color coded output maps. Results were good for modeling temperature, turbidity, Secchi disk depth and chlorophyll a, and indicate that remotely-sensed data may be applicable to monitoring water quality in this geographic area.

Key Words: Landsat TM, water quality, remotely-sensed data, Sicily.

INTRODUCTION

Recently, concern has grown over the quality of water in and around Sicily, Italy. As population and industrial activities have increased, more pressure is put on the water resources of the area. As part of a cooperative project to investigate the usefulness of remotely-sensed data to the analysis of resource related problems, Landsat Thematic Mapper (TM) data were used to assess water quality on Augusta Bay, Sicily. Upwelling radiance from the water column contains information about constituents in water and this radiance is detected by the TM.

Information on the distribution of water quality parameters in estuaries and bays provides greater understanding of ecology, biology, and dynamics in these systems. The distribution of certain parameters, such as chlorophyll a, can be an indication of biological activity and the productivity level in estuaries and bays. Other parameters such as turbidity and Secchi depth provide a better understanding of water circulation and currents, organic and inorganic matter distributions and hydrodynamic conditions in these systems. The water quality parameters discussed in this paper have been studied by conventional survey techniques in many similar bodies of water throughout the world. These surveys can be time consuming and expensive, and they frequently fail to represent the distributions of parameters of interest in heterogenous and patchy areas. Researchers have had to rely on sampling methods and interpolation and extrapolation between sample points. On the other hand, remotely-sensed data acquired from aircrafts and satellites provide a synoptic view not attainable from conventional methods and have produced a viable alternative for water quality modeling, mapping, and management.

Several studies have been done on other bodies of water utilizing remotely-sensed data with successful results (Khorram et al., 1987; Lathrop and Lillesand, 1986; Lin et al., 1984; Khorram, 1981a and 1981b; Johnson and Harriss, 1980; Uno et al., 1980). The modeling of water quality variables using Landsat TM data is based on physical principles utilizing statistical methods (Tassan, 1987; Tassan and Storm, 1986; Catts et al., 1985; Khorram and Cheshire, 1985; Farmer et al., 1983). Theoretical algorithms have been developed and tested for cholorophyll and sediment mapping in coastal waters of Venice and Ancona by Tassan and Storm (1986). Similar to our study, Tassan (87) utilized Landsat TM data for water quality mapping in the coastal waters of the Adriatic Sea with successful results.

Augusta Bay is located on the eastern coast of Sicily. It is a natural harbor enclosed by a man-made breakwater with two major inlets to the bay. The surface area of Augusta Bay is approximately 150 hectares. The harbor is heavily used for shipping by industries located along the coast of Sicily. The study area encompassed all of Augusta Bay and much of the sea surrounding the bay.

MATERIALS AND METHODS

General Approach

The general approach involved acquisition of water quality samples from boats near simultaneous with Landsat overpass; location of sample sites on the Landsat TM scene; extraction of digital numbers from all TM bands; development and verification of regression models relating concentrations of selected water quality variables to the spectral data; application of models to the entire study area; and the production of color coded resultant images each depicting the distribution of a selected water quality variable.

Collection of Ground Truth Data

Water quality samples were collected at 42 sample sites in and around Augusta Bay, Sicily, on June 7, 1988. Three rafts and one larger vessel were used in the data collection which was completed within one hour (11:00 a.m. to 12:00 p.m. local Daylight Savings Time).

Satellite overpass occurred around 11:05 local time. The sampling network was designed such that samples would represent the full range of conditions expected in this area. Secchi disk depth and temperature were measured in situ and samples were iced in the field and taken to the water quality lab in Messina, Sicily for analyses for turbidity, salinity, water density, suspended sediments (seston), and chlorophyll a. Secchi depth was not recorded at three of the sites and surface temperature was not recorded at two sites. Each boat carried navigational instrumentation and sample site locations were recorded in latitude and longitude. Sample site locations were also plotted on nautical charts.

Processing of Landsat Digital Data

Image analysis was performed at the Computer Graphics Center, North Carolina State University, USA, using the Land Analysis System (LAS) image processing software developed by Goddard Space Flight Center, National Aeronautics and Space Administration (NASA). Landsat 4 TM geometrically corrected digital data were acquired from EOSAT. A window encompassing the study area was extracted from quadrants three and four of the full TM scene (World Reference System Path 188, Row 34, June 7, 1988). Sample site geographic coordinates (latitude and longitude) were transformed to image coordinates (window X,Y) using the Space Oblique Mercator (SOM) projection system as an intermediate step. Locations of sample sites were visually confirmed by referring to their locations plotted on the nautical charts. No discrepencies were found in locating the sample sites.

Digital numbers (DNs) for the six reflective bands and the one thermal band were extracted for the single pixel and for the nine pixel block encompassing the center pixel corresponding to each sample site. Within band averages of the DNs from the nine pixels were used in the analyses. Single pixel values and kernel (block) averages were compared to verify that kernels represented spectrally uniform areas. One sample site was found to have a large deviation between the single pixel DN values and the kernel averages. That site was found to be next to a rock jetty. Kernel averages included water and rock surfaces, therefore, location of the site was moved by one pixel and one line on the image to avoid the jetty. New kernel averages for the 7 bands were uniform and representative of spectral values for the original single pixel. The land/water interface was determined from clustering and land values were radiometrically masked to zero prior to applying the models.

Statistical Analysis

All statistical analyses were done using the Statistical Analysis System (SAS) Software Package. Water quality variables were treated as dependent variables and Landsat TM kernel band averages were independent variables in the regression analyses. Prior to regression analysis, a correlation matrix was produced for all water quality variables and all kernel averages for the 42 sites. Also, plots of each water quality variable versus each band were produced. Examination of these plots and preliminary models indicated that a natural log transformation would be appropriate for all variables except the thermal data. All subsequent analyses were done on natural log transformed data. Ten sample sites were selected to be withheld from the regression analysis to be used to verify the models. Due to the small range of values represented by the water quality measurements, care was taken that sites with the maximum and minimum concentrations or values for each water quality parameter were retained in the data set used in model development.

Data plots and residuals for preliminary models indicated two sites which consistently appeared as outliers except with respect to the temperature versus Band 6 (thermal band) TM data. Reference to the original field (laboratory) data and the nautical charts indicated that Secchi depths at these two sites were at least as deep as bottom depth. Spectral values were probably higher because they included bottom reflectance, not just reflectance from the water column. A comparison was made between Secchi depth and bottom depth at all sites. Three sites (10, 21, and 23) where bottom reflectance was thought to be a potential problem were completely dropped from the analyses of all variables except temperature. Temperature DNs represent surface conditions only, not conditions in the water column. This was confirmed by the plots of the original data. Two of the sites dropped were in the data set used in model development. One of the sites was in the data set used in model verification. Final selection of models was based on R^2, probability of a greater F-value for the overall model, and significance of the t-test for regression parameters. A significance level of 0.05 was used for all models and regression parameters.

RESULTS

Results of the analyses include models predicting the concentrations of chlorophyll a and turbidity, a model predicting Secchi disk depth and a calibration model for determining surface temperatures from TM Band 6 thermal data. Results also include statistical summaries and color coded maps showing the distributions of the selected water parameters. No relationships have yet been found between the TM spectral data and suspended sediments or salinity. This may be because of the very low range of values or because of their low concentrations. Results of the water quality measurements are summarized below for all sites used in model development and model validation.

Variable	n	Mean	Maximum	Minimum
Secchi (m)	36	10.25	20.00	3.50
Turbidity (NTU)	39	1.00	5.00	0.00
Chlorophyll (ug/l)	39	1.26	5.54	0.00
Temperature (C)	40	20.76	23.	19.70

The following model was found to be the best predictor of the natural log of Secchi depth:

$$LNSECCHI = -44.00 + 20.42*\ln(B1) - 14.07*\ln(B2)$$

$$n=27 \text{ and } R^2=0.83$$

This model had an R^2 of 0.87 when applied to the verification data set (n=9). In addition to the three sites dropped from the analysis, three sites had missing data values.

For predicting the log of turbidity concentrations, the best model was:

$$LNTURB = -17.633 + 5.23*\ln(B2)$$

$$n=30 \text{ and } R^2=0.52$$

The verification R^2 was 0.57 (n=9).

For chlorophyll a, the following model was selected:

$$LNCHLA = 81.38 - 39.38*\ln(B1) + 29.08*\ln(B2)$$

$$n=30 \text{ and } R^2=0.84$$

The R^2 for the verification data set was 0.75 (n=9).

The natural log transformation was not used for calibration of the temperature model and no sites for which temperature measurements were available were dropped from the analysis. The model is:

$$TEMP = -19.57 + 0.33*B6$$

$$n=30 \text{ and } R^2=0.55$$

The validation R^2 was 0.50 (n=10).

All four models were applied to the entire study area, the resultant values were grouped into meaningful ranges and the maps were color coded (Figures 1 and 2; chlorophyll and turbidity maps are not shown).

DISCUSSION

Augusta Bay was relatively clear on the date this study was conducted as evidenced by the low concentrations of the water quality variables. However, we were successful in developing models for four of the variables of interest. Many of the field variables were highly correlated. In particular, Secchi depth and chlorophyll a had a correlation of -0.81. Previous studies have indicated that Landsat TM data are not sensitive to very low concentrations of chlorophyll a. In this area, we were successful in predicting chlorophyll concentrations that were all less than 6 ug/l. The similarity between the chlorophyll model and the model for Secchi depth indicates that light penetration depth may actually be the variable predicted by the chlorophyll model and this, in turn, is highly correlated with chlorophyll concentrations. Alternatively, there is the possibility that the clarity of the water (indicated by very low turbidity and suspended sediments and deep Secchi depths) meant that the spectral response was dominated by chlorophyll or chlorophyll-related products. Further studies are needed to determine if chlorophyll concentrations can be consistently predicted from TM spectral data in this geographic area.

The effects of bottom reflectance were most apparent when analyzing the preliminary models for chlorophyll. When two outliers were dropped from the model development data set due to the influence of bottom reflectance on the TM values, model R^2 changed from 0.70 to 0.84. It is important that the spectral data used in the analysis are a function of light upwelling from the water column and are influenced by what is in the water column rather than by bottom reflectance or reflectance from surrounding objects (e.g. jetties).

Model R^2 for turbidity was relatively low but not inconsistent with results which can be expected for such low concentrations (range of values 0 - 5 NTU). The same is true for the temperature calibration model. Less than 4°C of difference existed between surface temperatures at the sample sites. One interesting result from applying the temperature model was the appearance of two circular areas of relatively low temperatures off the coast. Such areas may be indicative of cooler water being upwelled from greater depths and may be a result of water circulation patterns in this area influenced by the inlet to Augusta Bay.

In general, the distributions of water quality variables depicted in the color coded maps follow the expected distributions of these variables and Landsat-TM data proved useful in modeling of selected water quality parameters.

REFERENCES

1. Catts, G.P., Khorram, S., Cloern, J.E., Knight, A.W., and DeGloria, S.D., "Remote sensing of tidal chlorophyll a variations in estuaries", Int. J. Remote Sensing, Vol. 6, No. 11, pp. 1685-1706, 1985.

2. Farmer, F.H., Jarret, O. Jr., and Brown, C.A. Jr., "Visible absorbance spectra: a basis for in situ and passive remote sensing of phytoplankton concentration and community composition", NASA, Washington, D.C., Tech. Paper TP-1-32, 1983.

3. Johnson, R.W., and Harriss, R.C., "Remote sensing for water quality and biological measurements in coastal waters", Photogrammetric Engineering and Remote Sensing, Vol. 46, No. 1, pp. 77-85, 1980.

4. Lathrop and Lillesand, "Use of thematic mapper data to assess water quality in Green Bay and central Lake Michigan", Photogrammetric Engineering and Remote Sensing, Vol. 52, No. 5, 1986.

5. Lin, G.A., Borstad, and Gower, J.F.R., "Remote Sensing of chlorophyll in red spectral region"; in Remote Sensing of Shelf Sea Hydrodynamics, J.C.J. Nihoul, Ed. New York: Elsevier Oceanography Series, Vol. 38, pp. 317-337, 1984.

6. Khorram, S., "Use of Ocean Color Scanner data in water quality mapping", Photogrammetric Engineering and Remote Sensing, Vol. 47, No. 5, pp. 667-676, 1981a.

7. Khorram, S., "Water quality mapping from Landsat digital data", International Journal of Remote Sensing, Vol. 2, No. 2, pp. 145-153, 1981b.

8. Khorram, S. and Cheshire, H.M., "Remote sensing of water quality in the Neuse River Estuary, North Carolina", Photogrammetric Engineering and Remote Sensing, Vol. 51, No. 3, pp. 329-341, 1985.

9. Khorram, S., Catts, G.P., Cloern, J.E., and Knight, A.W., "Modeling of estuarine chlorophyll a from an airborne scanner", IEEE Transactions on Geoscience and Remote Sensing, Vol. GE25, No. 6, pp. 662-670, 1987.

10. Tassan, S., "Evaluation of the potential of the Thematic Mapper for marine application", International Journal of Remote Sensing, Vol. 8, No. 10, pp. 1455-1478, 1987.

11. Tassan, S. and Storm, B., "An algorithm for the retrieval of sediment content in turbid coastal waters from CZCS data", International Journal of Remote Sensing, Vol. 7, No. 5, pp. 643-655, 1986.

12. Uno, S., Sugahara, Y. and Hayahawa, S., "Remote sensing of chlorophyll found in bodies of water", Proceedings, Fourteenth International Symposium on Remote Sensing of Environment, Ann Arbor, MI, pp. 1147-1157, 1980.

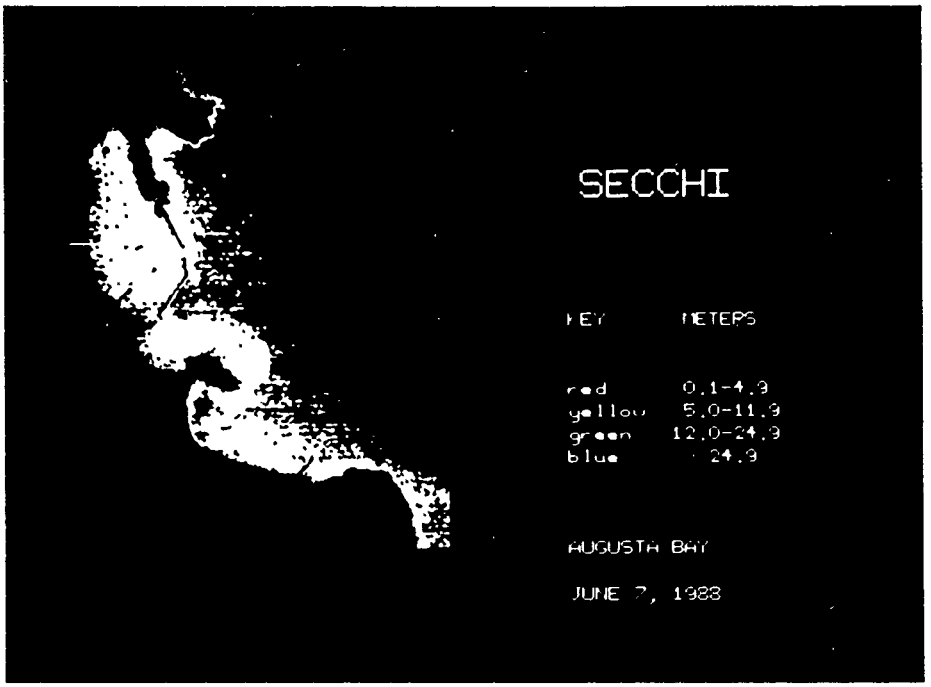


Figure 1. Landsat TM-based image representing Secchi depth in and around Augusta Bay, Sicily.

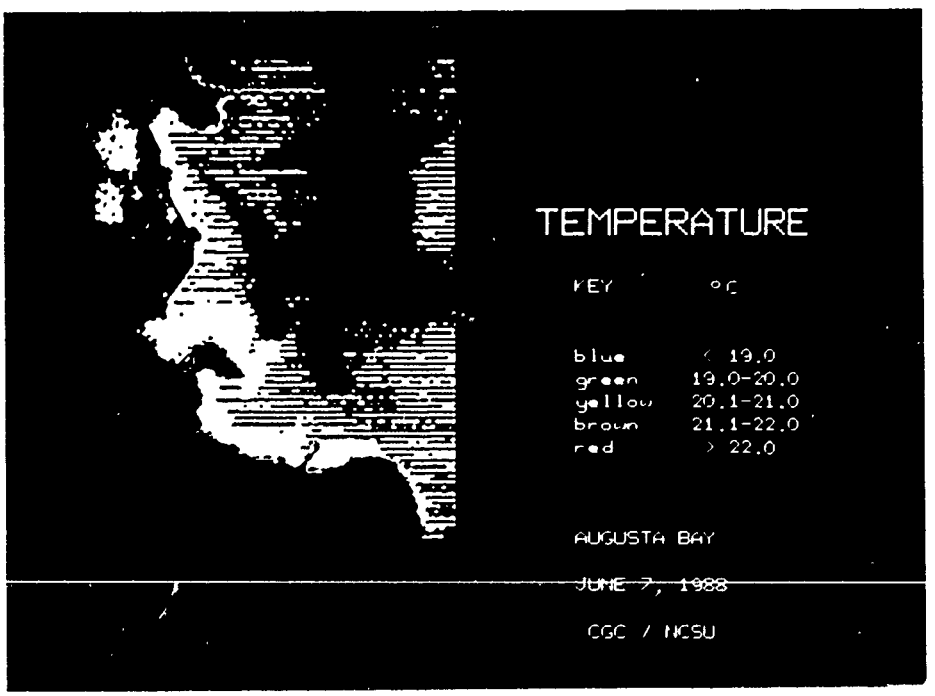


Figure 2. Temperature distribution in and around Augusta Bay, Sicily, as derived from Band 6 of Landsat TM data.

RIVERS IN THE SEA: CAN WE QUANTIFY PIGMENTS IN THE AMAZON AND THE ORINOCO RIVER PLUMES FROM SPACE?

Frank E. Müller-Karger[1], John J. Walsh[1],
Kendall L. Carder[1], and Rod G. Zika[2]

1: University of South Florida, Department of Marine Science,
140 Seventh Ave. South, St. Petersburg, Florida 33701
Phone: (813) 893-9186, Telemail: F.MULLER.KARGER, FAX: (813) 893-9610
2: University of Miami, RSMAS, 4600 Rickenbacker Cswy, Miami, Florida 33149
Phone: (305) 361-4715

ABSTRACT

Coastal Zone Color Scanner (CZCS) images of the western tropical Atlantic (1979-1982) were combined into monthly mean surface pigment fields. These suggested that early in the year Amazon River water flows along northeastern South America directly toward the Caribbean Sea. However, after June, the North Brazil Current is shunted eastward, carrying a large fraction of Amazon water into the North Equatorial Countercurrent (NECC). This eastward flow causes diminished flow through the Caribbean, which permits northwestward dispersal of Orinoco River water due to local Ekman forcing. The Orinoco plume crosses the Caribbean, leading to seasonal variation in surface salinity near Puerto Rico. At least 50% of the pigment concentration estimated in these plumes seems due to viable phytoplankton.

KEYWORDS: CZCS, Orinoco, Amazon, river

INTRODUCTION

Close to 25% of the primary production of the oceans occurs near continents (Walsh, 1988). This carbon is frequently in surplus relative to demand by the food web, and may be lost to deeper waters or sediments. Thus, continental margins may be a sink for excess carbon dioxide released to the atmosphere by human activity. In particular, rivers are important point sources of nutrients (and pollutants), with the Amazon and Orinoco rivers accounting for nearly 20% of the annual fresh water discharge into the world's oceans. But little is known about where this water and nutrients go. We have to understand the dispersal and productivity of this discharge to make sound management decisions about natural resources in the tropical Atlantic.

Typically, it is hard to detect river plumes offshore with remote sensors due to low instrument sensitivity in the visible or lack of contrast in the infrared. But plumes can be readily detected with sensitive instruments like the Coastal Zone Color Scanner (CZCS, aboard the Nimbus-7 satellite; see Hovis et al., 1980). Most bio-optical studies have focused on waters where optical constituents covary with pigment concentration, yielding algorithms to estimate phytoplankton biomass to within 40% of in situ concentrations (Gordon et al., 1982; Gordon and Morel, 1983; GORDON et al. 1983a). However, along continental margins waters are frequently Case II (Morel and Prieur, 1977) and identification of constituents is impossible with the poor spectral resolution of the CZCS (Fisher et al., 1986; Carder et al., 1988).

Below we examine the dispersal of the Amazon and the Orinoco plumes using the CZCS. We build a case for the presence of phytoplankton in the plumes using a mass balance for nutrients and historical observations of plankton. Finally, we examine CZCS water-leaving radiance data from these plumes and compare it with radiances from other blooms.

METHODS

We obtained water-leaving radiances (mW cm^{-2} μm^{-1} sr^{-1}) and pigment concentrations (mg m^{-3}) from 159 CZCS images collected over the Caribbean Sea and NE coast of Brazil between November 1978 and December 1982. We also used images processed by R. H. Evans (RSMAS, University of Miami) and G. Feldman (NASA Goddard Space Flight Center) showing normalized CZCS water-leaving radiances and pigment over the entire North Atlantic in 1979. Atmospheric corrections were those of Gordon et al. 1983a and Gordon et al. 1983b. Pigment concentrations were obtained using Gordon et al. (1983a). Products were mapped to standard projections and navigated to match a standard coastline.

RESULTS and DISCUSSION

There was good agreement between historical pigment concentration data and CZCS data in the greater Caribbean region. Concentrations were high in known areas of upwelling and also in the plumes of rivers. An unknown amount of dissolved organic material is present in the plumes, but since the rivers discharge high concentrations of nutrients (>10 μM nitrate-nitrogen and >30 μM total nitrogen), some of the color was probably due to viable algae. Direct observations of high concentrations of phytoplankton have been made in the Amazon plume by Wood (1966), Hulburt and Corwin (1969), Demaster et al. (1983), Borstad

(1982), and Dustan (unpublished data collected in 1982 and 1983). Similar communities may grow in the Orinoco's plume, as suggested by high concentrations of pigment measured in the Gulf of Paria in 1986 (0.7-2.3 mg m^{-3}) and south of Puerto Rico during surface salinity minima in October-November 1980 (>0.2 mg m^{-3}; Yoshioka et al., 1985).

We compared pigment estimates from full-resolution CZCS images (1 km x 1 km pixels) with a limited number of concurrent ship-based observations (n=27; Muller-Karger et al., in press). The CZCS underestimated the sum of surface in situ chlorophyll a and phaeopigments off Puerto Rico by a factor of about 0.61 in 1980 (n=6). Comparisons in the southern Caribbean in 1986 showed more scatter, but CZCS data were on average within a factor of 0.92 of surface in situ pigment concentration (n=21). The CZCS overestimated two observations of surface pigment concentration in the Orinoco river plume within 100 km of the delta (20 April 1986 in the Gulf of Paria). It was not possible to evaluate sub-pixel variability.

Dramatic seasonal changes were observed in the spatial arrangement of pigments in the western tropical Atlantic and Caribbean Sea. During June-January, Amazon water is preferentially dispersed offshore into the tropical Atlantic via the retroflection of the North Brazil Current and the North Equatorial Counter Current (Figure 1; Müller-Karger et al., 1988). A fraction of the discharge may reach Africa. In contrast, during February-May, the Amazon plume flows to the northwest along the coast, directly toward the Caribbean Sea.

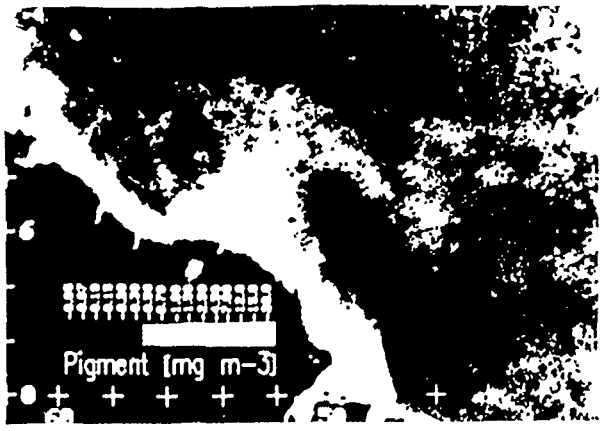


Figure 1. Composite of 15 CZCS images of the western tropical Atlantic (21 July 1979 - 9 January 1980), showing the offshore dispersal of Amazon water (the color bar is applicable only to this image).

In the Caribbean, concentrations > 0.5 mg m^{-3} were seen almost exclusively south of 14°N during January-May, in particular near capes and headlands off Venezuela and eastern Colombia. Since such areas receive little river input, and since sea surface temperature here is always low relative to the central Caribbean, it is likely that upwelling of nutrient-rich waters occurred near the coast (see Müller-Karger et al., in press). The plumes of the Amazon and Orinoco could be seen entering the Caribbean, extending westward along the coast.

In contrast, between June and December, a patch with concentrations > 0.5 mg m^{-3} covered an area >$3x10^5$ km^2 in the eastern Caribbean (Figure 2). Over scales larger than about 500 km, features were elongated to the west or northwest, reflecting the dominant direction of water movement. At smaller scales, southward and eastward motions were inferred from consecutive scenes (Müller-Karger, 1988). The largest eddy motions were about 250 km in diameter.

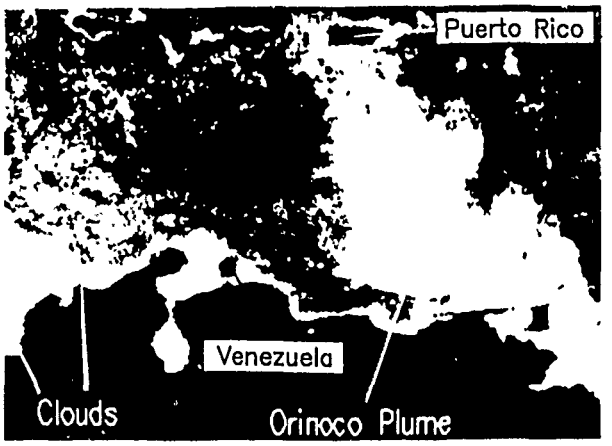


Figure 2. Composite of 9 CZCS overviews of the Caribbean Sea in October 1979.

The CZCS data show that the origin of the large pigment patch in the eastern Caribbean during the second half of the year is the Orinoco River rather than the Amazon (Figure 3). Since at this time winds are weak, sea surface temperature is high, and the seasonal thermocline is well developed, it is unlikely that this increase in chlorophyll is due to upward mixing of nutrient-rich deep water.

Nutrient loss processes in the Orinoco's plume were examined with transects along the plume's axis and with a simple trophic model based on histograms of the distribution of CZCS-derived pigments.

Transects were approximately 1000-1400 km long, the typical length of the plume in September-November. We converted distance to time assuming that particles moved at 0.3 m s^{-1} along the plume's axis. Invariably, pigments decreased along the plume. For example, the least squares curve that best fit data collected on 9 October 1979 was: $Pigment=5.34*e^{(-0.061*t)}$, (n=163, $r^2=0.51$), with pigments in [mg m^{-3}] and time (t) in [days]. This suggested that pigments decayed exponentially with a half life of ca. 11 days. Other dates also showed pigments disappearing faster than by conservative dilution in a one-dimensional model. This implied that the plume contributed to surface food webs and offshore sedimentation.

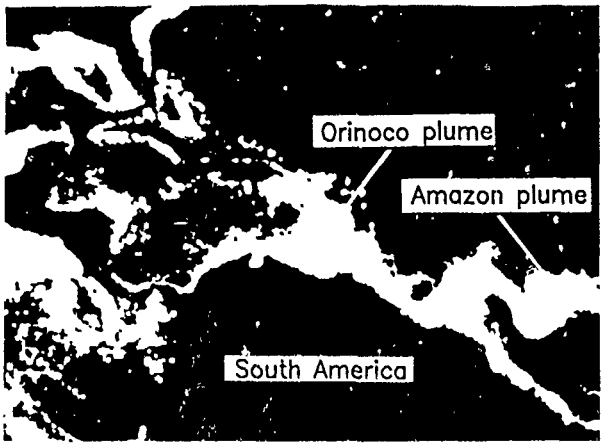


Figure 3. Composite of all CZCS data collected in October 1979 over the North Atlantic, clearly showing the separation of the Amazon and Orinoco plumes (light color).

The trophic model was based on a series of nominal conversions between CZCS-derived pigment and nitrogen concentrations, and between pigment concentration and productivity. The reader is referred to Müller-Karger et al. (in press) for detailed explanations. Here we focus on some of the results. All calculations requiring input of viable chlorophyll used an arbitrary value of 50% of the CZCS-derived pigment level, a necessary correction since phaeopigments are a byproduct of grazing and senescence of phytoplankton (Jeffrey, 1980; Margalef, 1965). Also, the presence of Gelbstoffe may lead to an overestimate of phytoplankton concentration (see Baker and Smith, 1982; Carder et al., 1988).

The model estimated average primary productivity (cf. Eppley et al., 1985) and nitrogen uptake in the plume, Dugdale and Goering's (1967) "f" ratio, turnover times of freshwater, particulate carbon and nitrogen concentrations, and the average number of times that a nitrogen atom is recycled before export. The Orinoco may contribute 2-12% of the daily nitrogen requirements of the phytoplankton growing in the plume, and lead to the fixation of 7-29x10^5 tons carbon per year. The rest of the demand appears to be met by nitrogen cycling, which means nitrogen must be recycled > 20 times before export.

Assuming that the turnover rate for carbon is similar to the nitrogen cycling time, the time needed to fill the estimated volume of fresh water in the plume was roughly equivalent to the number of cycles undergone by nitrogen. Also, the time needed to fill this equivalent volume of fresh water was similar to the time needed to balance the surrogate particulate nitrogen with N discharged by the river (we used the Redfield ratio; Redfield et al., 1963). Furthermore, if we assume that nitrogen disappeared from the plume at the decay rate obtained from our transects, nitrogen input by the Orinoco decreased to trace levels (i.e. 1% of the average input) in roughly the time needed for the plume to cover the area observed in the images, namely a little over 1 month. The above estimates suggest that at least 50% of the pigment observed with the CZCS in the plume was viable phytoplankton.

Automated differentiation between river plumes and other blooms is not easy. Figure 4 shows CZCS water-leaving radiances (440 nm) from the Orinoco plume and from an upwelling plume off Margarita Island (Venezuela). Nearshore, sediments are responsible for some backscatter. Over 100 km offshore, radiances in the river and upwelling plumes behave similarly. Clearly, plume radiances are affected by Gelbstoff and pigment concentration, but unfortunately there is no data on the concentration of Gelbstoffe in Caribbean waters that could be used to model the optical characteristics of such plumes. It is important that we build a bio-optical data base of these regions in order to effectively use future remote sensing data.

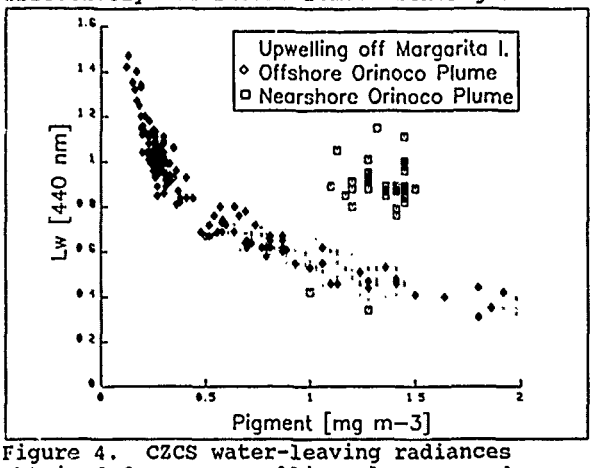


Figure 4. CZCS water-leaving radiances obtained from an upwelling plume around Margarita (periods), offshore Orinoco plume (diamonds, >200 km offshore) and nearshore Orinoco plume (squares, <50 km from Dragon's Mouth) on December 16, 1978 (CZCS orbit 738).

CONCLUSION

With more detailed measurements of the vertical and horizontal constituents and optical properties of the plumes, we may be able to separate Case I from Case II waters. Preliminary observations suggest that chlorophyll may be overestimated by >200% with the "standard" CZCS pigment algorithm in waters with a high Gelbstoff to chlorophyll ratio (Carder et al., submitted). These errors may be brought to under +-50% if additional spectral information is available. Carder et al. (submitted) propose quantifying chlorophyll-like pigments separate from Gelbstoffe and chlorophyll degradation products using additional water-leaving radiances measured at 412 nm. The promise of this new method can only be tested with an adequate data set obtained during combined ship-board and airborne experiments.

ACKNOWLEDGEMENTS

We thank Charles R. McClain (NASA, Goddard Space Flight Center) and Thomas R. Fisher (HPEL, University of Maryland), who guided the original processing and interpretation of the CZCS imagery. John Sissala, Mike Doline

and Richard Sipes (Nimbus Operations office, General Electric) helped select the imagery. Judy Chen, Mike Darzi and Jim Firestone (General Sciences Co.) implemented the image analysis software (SEAPAK). Robert Evans, Otis Brown, and Jim Brown (University of Miami, RSMAS) provided additional images and implemented the image analysis software (dsp). This work was supported by the Ocean Processes Branch at NASA Headquarters and by the NASA Graduate Student Researcher's Program at the Goddard Space Flight Center (Grant No. NGT 21-002-822).

REFERENCES

Baker, K. S., and R. C. Smith, "Bio-optical classification and model of natural waters", Limnology and Oceanography, 27, 500-509, 1982.

Borstad, G. A., "The influence of the meandering Guiana Current on surface conditions near Barbados - temporal variations of Trichodesmium (Cyanophyta) and other plankton", Journal of Marine Research, 40, 435-452, 1982

Carder, K. L., R. G. Steward, G. R. Harvey, and P. B. Ortner, "Marine Humic and fulvic acids: Theur effects on remote sensing of ocean chlorophyll", Limnology and Oceanography, 34(1), 68-81, 1989.

Carder, K. L, S. Hawes, K. A. Baker, R. C. Smith, and R. G. Steward, "Remote Sensing algorithms for discriminating marine humus from chlorophyll", submitted, Limnology and Oceanography (special issue on ocean optics, 1989).

Demaster, D. J., G. B. Knapp, and C. A. Nittrouer, "Biological uptake and accumulation of silica on the Amazon continental shelf", Geochimica et Cosmochimica Acta, 47, 1713-1723, 1983.

Deuser, W. G., F. E. Müller-Karger and Ch. Hemleben, "Temporal variations of particle fluxes in the deep subtropical and tropical North Atlantic: Eulerian versus Lagrangian effects", Journal of Geophys. Research, 93 C6, 6857-6862, 1988.

Dugdale, R. C., and J. J. Goering, "Uptake of new and regenerated forms of nitrogen in primary productivity", Limnology and Oceanography, 12, 685-695, 1967.

Eppley, R. W., E. Stewart, M. R. Abbott, and U. Heyman, "Estimating ocean primary production from satellite chlorophyll. Introduction to regional differences and statistics for the Southern California Bight", Journal of Plankton Research, 7, 57-70, 1985.

Fisher, J., R. Doerffer, and H. Grassl, "Factor analysis of multispectral radiances over coastal and open ocean water based on radiative transfer calculations", Applied optics, 25, 446-456, 1986.

Gordon, H. R., D. K. Clark, J. W. Brown, O. B. Brown, and R. H. Evans, "Satellite measurement of the phytoplankton pigment concentration in the surface waters of a warm core Gulf Stream ring", Journal of Marine Research, 40, 491-502, 1982.

Gordon, H. R., D. K. Clark, J. W. Brown, O. B. Brown, R. H. Evans and W. W. Broenkow, "Phytoplankton pigment concentrations in the Middle Atlantic Bight: Comparison of ship determinations and CZCS estimates", Applied Optics, 22, 20-35, 1983a.

Gordon, H. R., J. W. Brown, O. B. Brown, R. H. Evans and D. K. Clark, "Nimbus 7 CZCS: reduction of its radiometric sensitivity with time", Applied Optics, 22, 3929-3931, 1983b.

Gordon, H.R., and A.Y. Morel, "Remote assessment of ocean color for interpretation of satellitevisible imagery: A review", Lecture Notes on Coastal and Estuarine Studies, No. 4. New York: Springer-Verlag, 114 p, 1983.

Hovis, W. A., D. K. Clark, F. Anderson, R.W. Austin; W.H. Wilson, E. T. Baker, D. Ball, H. R. Gordon, J. L. Mueller, S. Z. El-Sayed, B. Sturm, R. C. Wrigley, and C. S. Yentsch, "Nimbus-7 Coastal Zone Color Scanner: system description and initial imagery", Science 210:60-63, 1980.

Hulburt, E. M., and N. Corwin, "Influence of the Amazon River outflow on the ecology of the western tropical Atlantic. III. The planktonic flora between the Amazon River and the Windward Islands", Journal of Marine Research, 27, 55-72, 1969.

Jeffrey, S. W., "Algal pigment systems", In: Primary Productivity in the Sea, P. G. Falkowski, editor, Plenum Press, 33-58, 1980.

Margalef, R., "Composicion y distribucion del fitoplancton", Memoria, Sociedad de Ciencias Naturales La Salle, XXV, 141-205, 1965.

Müller-Karger, F. E., "Pigment variability in the Caribbean Sea: A study using the Coastal Zone Color Scanner", Ph.D. Dissertation, Horn Point Environmental Laboratories, University of Maryland, 218 pp, 1988.

Müller-Karger, F. E., C. R. McClain, T. R. Fisher, W. E. Esaias, and R. Varela, "Pigment distribution in the Caribbean Sea: Observations from Space", Progress in Oceanography, In press, 1989.

Müller-Karger, F. E., C. R. McCLAIN, and P. L. RICHARDSON, "The dispersal of the Amazon's water", Nature, 333, 56-59, 1988.

Redfield, A. C., B. H. Ketchum, and F. A. Richards, "The influence of organisms on the composition of sea water", In: The Sea, M. N. Hill, editor, Chapter II, Interscience, New York, 26-77, 1963.

Walsh, J. J., "On the nature of continental shelves", Academic Press, 508 p, 1988.

Wood, E. J. F., "A phytoplankton study of the Amazon region", Bulletin of Marine Science, 16, 102-123, 1966.

Yoshioka, P. M., G. P. Owen, and D. Pesante, "Spatial and temporal variations in Caribbean zooplankton near Puerto Rico", Journal of Plankton Research, 7, 733-751, 1985.

The Radar Image of the Turbulent Wake Generated by a Moving Ship

James K.E. Tunaley, John R. Dubois and J. Brian A. Mitchell

London Research and Development,
755, Queens Avenue, London, Ontario, Canada

ABSTRACT

A wake behind a surface ship can frequently be seen in synthetic aperture radar imagery at both L and X-bands. One of the features is often the "turbulent wake", which appears sometimes as a reduction in scattering cross-section and sometimes as a "bright" streak (or streaks) extending some kilometres behind the vessel. A model for the scattering of em waves from the turbulent wake is presented. The model is based on the propagation of surface waves through the steady component of the flow associated with the wake together with the Bragg mechanism for scattering. Cross-sections across the wake are calculated and some of them exhibit the type of behaviour that has been observed. A comparison of the theory with SEASAT imagery is underway.

KEY WORDS: RADAR, WAKE, TURBULENCE

I. INTRODUCTION

A library of about 50 SEASAT and some airborne SAR wake images has been compiled at the Defence Research Establishment Ottawa. Providing the spatial resolution is sufficiently fine in both azimuth and range (such as that obtainable in SAR), radar images of the wakes from surface ships seem to fall principally into four categories. These are the Kelvin wake, the turbulent wake, wakes from internal waves and narrow-V wakes. Not all types are seen at a given frequency: for example narrow-V wakes have been seen with L-band radars but not at X-band. Work is underway to develop a fairly comprehensive simulation of the SAR image of a wake. The aim of this is to study the production of the various types of wake image under different conditions. However, since there will generally be a lack of knowledge of the conditions under which an image is produced, a very accurate solution to the problem is not required at this stage. In this paper attention is confined to the turbulent wake. Though not always visible in radar imagery, the turbulent wake is most often seen as a dark streak behind a surface ship, stretching back for several kilometres.

The effect would seem to indicate a reduction in the amplitudes of the surface waves responsible for the scattering of the em waves in the turbulent wake. However it must be noted that sometimes the dark streak may exhibit bright edges and perhaps other structure. The Bragg mechanism of scattering, which will be adopted here, is capable of explaining most of the phenomena associated with ocean scatter, except at small radar depression angles and higher radar frequencies. An improvement has been made by Holliday et al, 1986, 1987, which is roughly equivalent to the inclusion of the effect of varying the angle of incidence because of the wave slopes. It appears that this results in an enhanced scattering cross-section at X-band over that predicted by the simple Bragg scatter mechanism because longer waves make a significant contribution.

When a vessel is underway, a "turbulent" hydrodynamic wake stretches out behind it (Schlichting, 1979; Townsend, 1976)). This type of wake consists not only of vortices with some energy spectrum but it also has a steady component concentrated in a narrow region along the ship's track. Whilst the theory of the production of the turbulent wake is not sufficiently advanced to be able to make accurate predictions for a given ship, an order of magnitude for the various quantities can be obtained readily. It can be shown that the non-steady component of vortices is not likely to be important to the radar scattering problem: a treatment of only the "dc" component is necessary (Tunaley et al, 1986).

Once the profile of the mean flow has been found, the propagation of a wave, with a given wavevector, outside the wake can be studied as it impinges upon the mean flow. The theory is similar to that involved in the propagation of acoustic waves in the atmosphere where there is a wind. The theory is well established (Lighthill, 1978). It turns out that, because the phase velocity of surface gravity waves is quite small at wavelengths corresponding to L or X-bands, the effect of a change in the flow velocity of a few centimetres per second over a distance of a few metres can have a dramatic effect. Thus a wave travelling from outside the wake will generally change its wavelength, phase and group velocities and its direction significantly as it enters the wake. Indeed, for some positions and orientations within the wake, it may be impossible for a wave to propagate from the outside: this clearly suggests the possibility that serious reductions in scattering cross-section can occur.

In the present study a simplified approach is adopted. It is assumed that, for a small area about each position within the wake, scattering arises from a Bragg wave. The ray path is then traced back to the outside of the wake so that the amplitude of the wave can be found. A simplified power law spectrum (Phillips, 1977) is chosen; it is assumed that this

spectrum of waves has been generated by the wind at some distance from the wake and that this spectrum has a simple dependence on the wind direction. Whilst the wind will produce new waves which will decay, as described by Hughes, 1978, it is further assumed that the wind cannot produce significant new waves over the extent of the wake itself and that the associated damping of the waves is negligible. These latter assumptions are reasonable for L but not for X-band.

II. THE TURBULENT WAKE IN THE FLUID

The turbulent wake from a towed vessel occurs because water has a small but finite viscosity and because it wets the submerged surface of the ship. This causes a frictional drag which imparts momentum to the fluid. The rate of transverse diffusion of this momentum by the random velocity field of vortices tends to be small. Even if the wake were to start in a regular manner, it would become random downstream (Schlichting, 1979). Thus eddies of a wide range of scales will be produced as well as a mean flow component directly astern of the ship along the ship track: the wake can be regarded as an extension of the boundary layer.

Mainly two types of shear flow are treated in the literature. These are the two-dimensional flow past a long cylinder oriented transverse to the principal flow direction and the axisymmetric three-dimensional wake behind a body of revolution. Information concerning the mean flow velocity distributions and the development of the shape of the wake can be obtained using the mixing length theory of Prandtl (Prandtl and Tietjens, 1934). The theory is based on the idea that the appropriate dynamical quantity can diffuse throughout the flow in an analogous manner to molecules diffusing in a gas. In a sense, the "mixing length" is similar to the mean free path but the diffusing elements are macroscopic "lumps" of fluid rather than microscopic objects.

With Prandtl's assumptions on the geometrical and mechanical similarity of the flow in different sections of the wake, the differential equations can be solved for the turbulent wake region away from the object creating the disturbance. Thus it is to be expected that the solution cannot depend too greatly on the nature of the object itself and that the results will be more or less generally applicable. The solutions for a sphere have been discussed by Schlichting. The result is not restricted to spherical bodies as long as the appropriate drag coefficient is inserted to represent the correct rate of momentum production: it can be employed to find the order of magnitude of the wake parameters astern of a ship.

To estimate the wake parameters for a ship, a drag coefficient representing the net rate of production of linear momentum by a self-propelled ship is needed. At moderate speeds the drag due to this production of momentum by the Kelvin wake tends to be of the same order as that due to viscous drag. Since the results of the theory for the wake width and the velocity of the flow depend on the drag coefficient to the power of 1/3, estimates based on the skin drag for towed body can be utilised. Thus the production of momentum by the screws will tend to cancel that caused by skin drag. However, there will typically be a net production of fluid momentum in the opposite direction to the ship velocity because of wavemaking resistance and wind drag forces. Figures 1 and 2 show graphs of the wake half width and the velocity at the centre of the wake as a function of distance behind a frigate sized ship with a drag coefficient of 0.1. The notable feature is the persistence of the wake for several kilometres, owing to the slow rate of spreading.

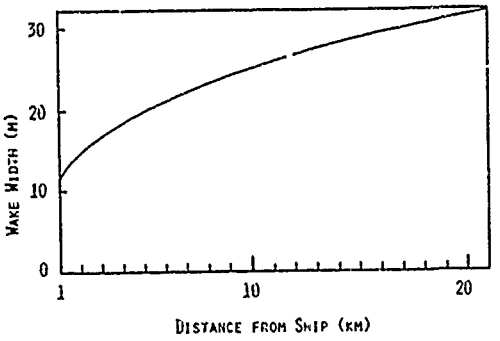


Figure 1. The wake half width as a function of distance behind a frigate travelling at 20 kt.

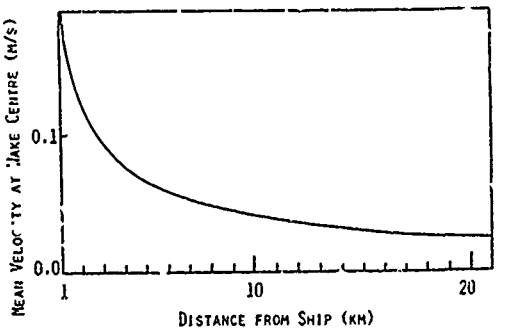


Figure 2. The mean velocity of the wake at its centre as a function of distance behind a frigate moving at 20 kt.

III SURFACE WAVE PROPAGATION

This section is concerned with waves on a surface upon which the velocity changes smoothly from point to point. It is assumed that the velocity changes over a wavelength are very small. The propagation of waves on or within a moving medium has been treated by Peregrine, 1976, Lighthill, 1978, and Phillips, 1977. The frequency of a wave measured at any point in an inertial reference frame is the same. However the frequency measured in a frame which is stationary with respect to the local water surface will generally be Doppler shifted in frequency. If ω is the angular frequency outside the wake in a frame in which the water is stationary and σ is the angular frequency measured by an observer moving with the fluid, the usual treatment of the Doppler effect yields

$$\sigma = \omega - \underline{k}.\underline{U}$$

where $\underline{U}$ is the local fluid velocity. The notion of a local frequency is useful because the wavelength on the surface is directly related to it. For example a wave impinging on a wake flow which has a component of velocity in the same direction as the wavevector will have a reduced local frequency and its wavelength will be increased. For dispersive waves, the phase and group velocities will also change.

For radar wavelengths at L-band, the Bragg waves will typically be in the gravity wave regime where the propagation is determined by the competition

between inertial and gravitational forces. Because the changes in the wake velocity are small over a wavelength, ray theory can be applied. It is also assumed that the changes in the mean wake velocity and the wake breadth are small compared to the wake breadth so that the water surface over the wake can be treated as a two-dimensional medium stratified in a direction parallel to the ship's track. Refraction of rays can be analysed by noting that the phase of the wave must be continuous across a line of stratification. This leads to Snell's law. Lighthill, 1978, has shown that the component of the wave action flux density perpendicular to the lines of stratification (in this case lines of constant fluid velocity) is a constant along a ray path. This vector is defined by

$$\underline{b} = \rho g A^2 (\underline{U} + \underline{C}_g)/(2\sigma)$$

Here A is the wave amplitude, $\underline{C}_g$ is the local group velocity and ρ is the density of the fluid. The use of this conservation principle allows the wave amplitude of a ray at various points across the wake to be calculated. Thus, as well as dealing with the "stretching" or "contraction" of a wave as it is Doppler shifted, it takes into account the changing cross-section of a "ray tube", defined as the region enclosed by two neighbouring rays, as it passes through the medium. These relations must be augmented by the dispersion relations and the expressions for the phase and group velocities. A solution of the equations shows that some waves may be totally reflected from the wake, even for small flow velocities.

The relationships allow all the parameters of a gravity wave propagating into a wake to be found after the wave outside the wake has been specified. For the backscattering of radio waves, the water waves must have the Bragg wavelength and be oriented parallel to the projection of the radio wavevector on the mean water surface. Therefore we have the inverse problem: it is necessary to specify the waves inside the wake and trace rays back to the outside. For waves travelling against the flow, it may happen that, at some positions in the wake, a solution to the equations does not exist. In this case these points in the wake will not support a Bragg wave and the radar cross-section there will be zero.

Generally, the waves outside the wake can be described by a spectral density, such as the Phillip's spectrum but including a factor to take account of the orientation relative to the principal wind direction. Because of wind gusting, it may be expected that the dependence on orientation may not be all that strong and, in the present study, we illustrate the radar cross-sections with two extreme models. In the first, we introduce a spectral density for the waves:

$$F_0(k,\theta) = k^{-4} \cos^2(\theta - \alpha),$$

for $-\pi/2 < (\theta - \alpha) < \pi/2$ and $F_0 = 0$ otherwise. The angle α is the mean direction of the wind relative to the radar, k is the wavenumber and θ is the wavevector angle.

The second model simply represents an isotropic wave spectrum.

Because the wavelength (or k) changes as a wave propagates through the wake, the region in wavevector space also changes. A correction must be made for this transformation which involves a Jacobian.

Programs have been written to generate plots of the relative radar cross-section, which is proportional to the spectral density of the Bragg waves. The algorithm is equivalent to the tracing of rays from a given point in the wake to the outside. Rays travelling both towards and away from the radar are examined. For each point on the wake section, tests are made to determine if a contribution to the wave amplitudes can arise from a reflection process in the wake and to establish whether it is possible for a ray to propagate from the outside. In some circumstances there appear to be abrupt changes in the radar cross-section and these are associated with (i) the onset of reflection as the observation point moves into the wake, (ii) changes in the wave spectrum as a function of orientation.

Plots of relative scattering amplitude for positions across the wake are shown in figures 3-6 for a large merchant ship at a distance of 1 km downstream. The Bragg wavelength has been set at 0.3m. Different values of radar and principal wind directions with the "anisotropic" spectrum are chosen to illustrate the variety of radar signatures. The turbulent wake can be dark, bright or dark with bright edges. It can also exhibit more complicated

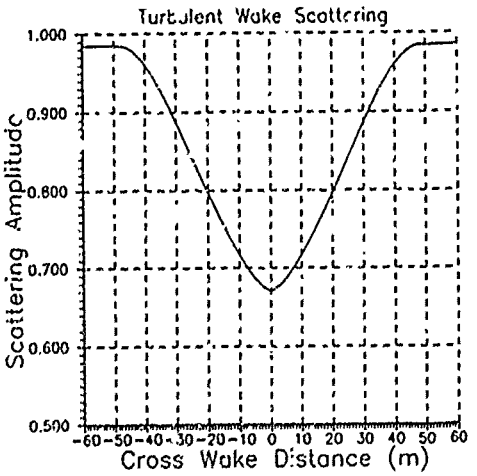


Figure 3. The radar scattering across a wake: wind waves asymmetric.

structure. In the latter case it may be asymmetric. This is due to the fact that on the windward side of the wake, there are Bragg waves reflected from the wake as well as ambient waves.

The case of an isotropic wind generated wave spectrum is also interesting. Clearly the wind direction is irrelevant and it turns out that the radar cross-sections are independent of the sign of the angle of incidence. There are three basic signatures. All are symmetric. These are the bright wake, the dark wake and the dark wake with bright edges. An example of the latter is shown in figure 7.

V. CONCLUSIONS

The model that has been developed leads to results that seem to be consistent with the experimental findings. It certainly predicts the persistence of the wake for many kilometres behind a frigate sized or larger ship. In making comparisons with the existing data it must be appreciated that some confusion may exist between the narrow-V wakes and bright arms of the turbulent wake. Furthermore a specification of the wind direction may be insufficient to determine the type of wake - the angular dependence of the wind generated wave

spectrum is important.

VI. ACKNOWLEDGEMENT

This work has been conducted for the Canadian Department of National Defence in connection with a contract between London Research and Development and Supply and Services Canada. The authors wish to thank K.H. Wu and M. Rey of the Defence Establishment Ottawa for acting as the scientific authorities for this project and for proposing, facilitating and encouraging the work.

REFERENCES

D. Holliday, G. St-Cyr and N.E. Woods. "A Radar Ocean Imaging Model for Small to Moderate Incidence Angles", Int. J. Remote Sensing, vol 7(12), pp 1809-1834, 1986.

D. Holliday, G. St-Cyr and N.E. Woods. "Comparison of a New Radar Imaging Model with SARSEX Internal Wave Image Data", Int. J. Remote Sensing, vol 8(9), pp 1423-1430, 1987.

B.A. Hughes. "The Effect of Internal Waves on Surface Wind Waves 2. Theoretical Analysis". J. Geophysical Research, vol 83 (C1), pp 455-465, 1978.

J. Lighthill. "Waves in Fluids", Cambridge University Press, 1978.

D.H. Peregrine. "Interaction of water waves and currents". Advances in Applied Mechanics, vol 16, pp 9-117, 1976.

O.M. Phillips. "The dynamics of the upper ocean", Cambridge University Press, 1977.

O.M. Phillips. "Spectral and Statistical Properties of the Equilibrium Range in Wind-Generated Gravity Waves", J. Fluid Mech., vol 156, pp 505- 531, 1985.

L. Prandtl and O.T. Tietjens. Fundamentals of Hydro- and Aeromechanics, Dover, New York, 1934.

H. Schlichting. "Boundary-Layer Theory", McGraw-Hill Book Co., 1979.

A.A. Townsend. "The Structure of Turbulent Shear Flow", Cambridge University Press, 1976.

J.K.E. Tunaley, P.A. Rochefort and J.B.A. Mitchell. "The modelling and classification of SAR and ISAR ship images including their wakes: Part 2". London Research and Development Interim Report under Supply and Services Contract D.S.S. 13SV.97714-4-2819, March, 1986.

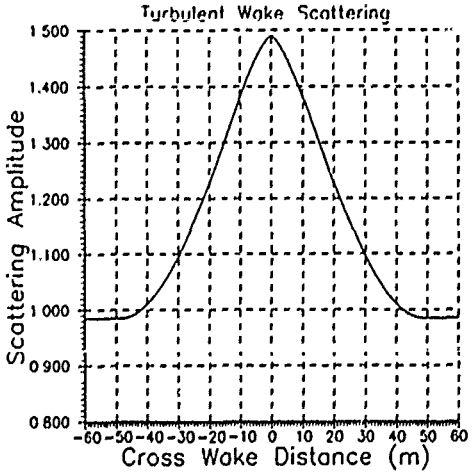


Figure 4. The radar scattering across a wake: wind waves asymmetric.

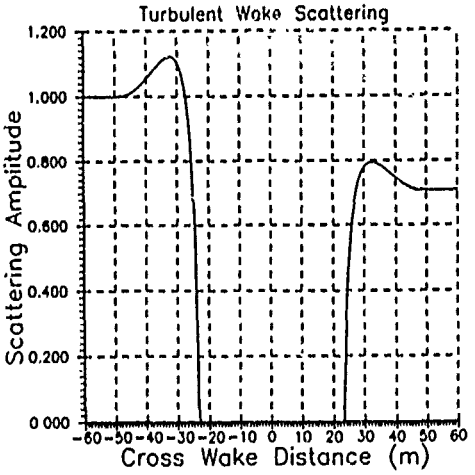


Figure 5. The radar scattering across a wake: wind waves asymmetric.

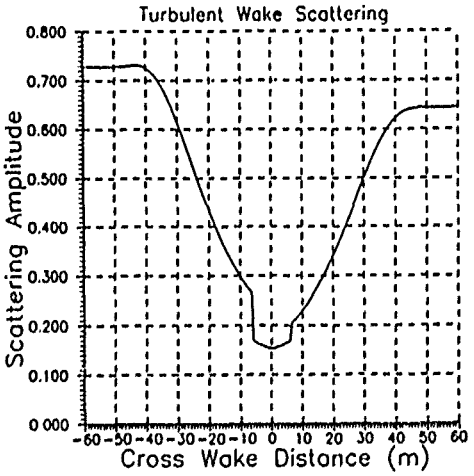


Figure 6. Radar scattering across a wake: wind waves asymmetric

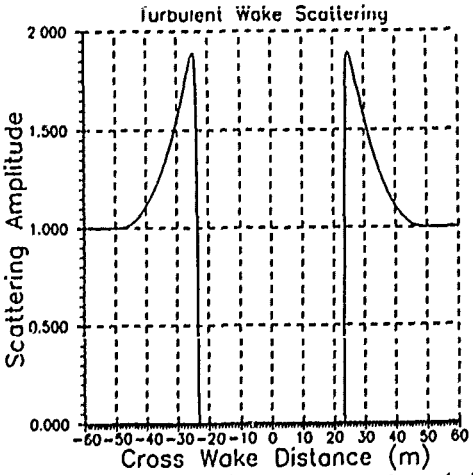


Figure 7. Radar scattering across a wake: wind waves symmetric.

DETECTION OF LOW CONTRAST FEATURES IN SAR IMAGES

Ken Nicolas

Code 5382
Naval Research Laboratory
Washington, D.C. 20375-5000
(202) 767-2003
FAX: (202) 767-5599

We investigate the implementation of processing algorithms to retain the highest probability of detection for ship wakes in SAR images. We also examine their ability to determine ship direction and speed.

The sea clutter background and the speckle nature of a SAR image both create a complex environment which makes the extraction of low contrast features, such as ship wakes, difficult. A further complication is the variable morphology of ship wakes. The bright arms of the "V" wake are sometimes not present, or appear as single or multiple linear features extending aft of the ship. The appearance of the centerline wake as a negative contrast linear feature is dependent on the sea clutter background. The linearity of these features is in turn dependent on the constancy of the ship heading, and the surface currents through which the ship passes.

We discuss the effectiveness of the algorithms, and their dependent on the assumptions made about the appearance of the wake, and its relationship with the background. To evaluate the processing algorithms, we procede with the following steps.

The first step includes comparisons between several pre-detection filters which smooth and edge enhance both one-look and four-look SAR images. We examine averaging algorithms such as the median filter and adaptive filters which use local statistics clustered near the target pixel.

Next we apply several wake detection algorithms to the smoothed images. A sector search algorithm uses knowledge of the ship location, by centering on the brightest pixels in the image. Other methods such as the Hough transform and matched filtering technique are also investigated.

The resulting noise levels and feature contrasts for each combination of processing steps are then compared. The probability of detection, for the range of observed ship wake parameters, is then compared for each detection algorithm.

EFFECT OF THE KELVIN WAKE TRANSVERSE WAVE COMPONENT ON SAR IMAGERY OF SURFACE WAVE WAKES

B.A. HUGHES

DEFENCE RESEARCH ESTABLISHMENT PACIFIC
FMO VICTORIA, B.C. VOS 1BO CANADA

Abstract

A theoretical investigation is given of the effect of a long wavelength corrugation on the SAR image formed by first-order Bragg scattering from a deterministic sea-wave pattern such as a Kelvin wake. A standard SAR processor is assumed. It is shown that the corrugating wave can horizontally rotate the effective Bragg acceptance angle by up to ~10° for typical Kelvin wake parameters.

Key Words

SAR, Kelvin wake, Narrow-V, Bragg scattering, curvature

1. Introduction

One of the significant features in SAR images of Kelvin wake systems at low wind speed is the appearance of a narrow V near the centre of the wake (Hammond, et.al., 1985; Hughes and Dawson, 1988). Velocity bunching effects are known to be of possible importance in the occurrence of the V, but, so far, surface shape effects per se have not been studied. For steady Kelvin wakes, particle velocity (or acceleration) and curvature of the longer waves are interdependent, but they produce different results in the SAR processing and so they are not directly interchangeable.

In the present paper a deterministic problem is solved in which the Bragg-scattering field (the diverging wake wave-field) exists in the presence of a long "corrugating" wave (the transverse wake wave-field). The Kelvin wake is assumed to be steady, and so the Bragg-scatterers are phase-locked to the corrugating wave. They both have the same component of phase velocity along the ship's track, namely, the ship's speed. The corrugating field is not particularized to the detailed Kelvin wake structure but is left as an infinite sinusoid. A spectral description is used for the Bragg scatterers rather than a correlation description. This offers a simplified interpretation based on narrow-band processes, although the Bragg scatterers are also left unspecified in detail. The problem is solved in terms of a filter function which, when applied to the Bragg scattering divergent wave field, provides the unsquared complex SAR image.

A mismatch in aircraft speed is allowed for in the processing, and the beam and pulse shapes are assumed to be Gaussian.

In the processor simulation, phase terms in the defining integral are expanded to second order in distance and time (relative to the centre of the given pixel and the pulse centre-time). Only zeroth, first-order and some non-negligible second-order terms in the other pertinent small quantities are kept in the integrand phase. In this case the small quantities are proportional to corrugation wave-height, ship-speed and aircraft speed mismatch terms.

The theoretical analysis defining the approximating expansions is provided in the next section, followed by the results and a discussion. The main results are contained in a numerical example for the condition pertinent to narrow-V wakes in Dabob Bay, WN (Hughes and Dawson, 1988).

2. Theoretical Analysis

It can be shown straightforwardly (Hughes, 1981, Wright, 1968) that the undetected image $W(\vec{x}_o)$, for 1st-order Bragg scattering with matched filter processing, is given by

$$W(\vec{x}_o) = \iint_{-\infty}^{\infty} \varsigma(\vec{x}',t')I(\vec{x}',\vec{x}_o,t')d\vec{x}'dt' \qquad (1)$$

Here $\varsigma(\vec{x}'t')$ is the height of the sea surface above its ambient mean level, and $I(\vec{x}',\vec{x}_o,t')$ is the equivalent impulse response function of the entire process, given by

$$I(\vec{x}',\vec{x}_o,t') = -\frac{z^2\omega_c^2}{4\pi^2c^2R^4\Delta}\exp\left[\frac{2i\omega_c}{c}\frac{x_o^2}{\sqrt{x_o^2+z_o^2}}\right.$$

$$+\frac{i\omega_c}{c}(r_1+r_2-2r_o) - \left\{\frac{(y'-V_1t')^2+(y'-Vt')^2}{Y_g^2}\right\}$$

$$\left.-\frac{(r_1+r_2-2r_o)^2}{2c^2T_g^2}(\omega'^2T_g^4-1)\right] \qquad (2)$$

where

$$r_1 \simeq \sqrt{x'^2+(y'-Vt')^2+z^2}\ , \qquad (3)$$

$$r_2 \simeq r_1\ , \qquad (4)$$

$$r_o \simeq \sqrt{x_o^2+(y_o-V_1t')^2+z_o^2}\ . \qquad (5)$$

In the coordinate frame used here, $\vec{x}'(=x',y')$ is a horizontal vector fixed to the ambient ocean surface, z and z_o are the height of the aircraft above the true and ambient sea surface respectively, V and V_1 are the true aircraft speed and a (possibly) mismatched version, both directed along the azimuth (y), ω_c and ω' are the central radar frequency and FM chirp rate, T_g and Y_g are pulse lengths in time (range) and beamwidth (azimuth) to the e^{-1}-points for a Gaussian beam, R is the slant range from radar receiver to the scene centre, Δ is the time interval between pulses and c is the radar signal propagation speed. The horizontal coordinate vector $\vec{x}_o$ refers to the processed scene in ground coordinates.

A Fourier transform description for ς will now be used and the r's of Equations (3) to (5) will be binomially expanded about a common position and terms up to and including second order in deviations from this position will be kept. The altitude z in Equation (3) will also be allowed to be a function of position in order to describe the effect of the height of the corrugating wave.

The entire wake wave-field including the corrugating portion is reducible to a steady-state by a simple transformation to coordinate axes moving with the source ship. Thus, with $\vec{c}_s$ as the ship's velocity, and with distance measured from the ship being given by $\vec{\eta}$, the following pertain:

$$\vec{\eta} = \vec{x}' - \vec{c}_s t' \quad , \tag{6}$$

$$\varsigma(\vec{x}',t') = \varsigma(\vec{\eta}) \quad ,$$

and, from Equation (1)

$$W(\vec{x}_o) = \iint_{-\infty}^{\infty} \varsigma(\vec{\eta}) \int_{-\infty}^{\infty} I(\vec{\eta}+\vec{c}_s t', \vec{x}_o, t') dt' d\vec{\eta} \quad . \tag{7}$$

The divergent wave-field ς can now be expressed as a two-dimensional Fourier transform,

$$\varsigma(\vec{\eta}) = \iint_{-\infty}^{\infty} p(\vec{k}) e^{i\vec{k}\cdot\vec{\eta}} d\vec{k} \tag{8}$$

so that

$$p(\vec{k}) = \frac{1}{4\pi^2} \iint_{-\infty}^{\infty} \varsigma(\vec{\eta}) e^{-i\vec{k}\cdot\vec{\eta}} d\vec{\eta} \quad , \tag{9}$$

and $$W(\vec{x}_o) = \iint_{-\infty}^{\infty} p(\vec{k}) e^{i\vec{k}\cdot\vec{x}_o}$$

$$\times \left(\iiint_{-\infty}^{\infty} e^{i\vec{k}\cdot(\eta - x_o)} I(\vec{\eta}+\vec{c}_s t', \vec{x}_o, t') dt' d\vec{\eta} \right) d\vec{k} \tag{10}$$

The term $e^{i\vec{k}\cdot\vec{x}_o}$ is taken out explicitly in the integrand of this last expression in order that $W(\vec{x}_o)$ may be seen to be simply a filtered version of ς by comparing Equations (10) and (8). The filter function, denoted Q_2, is given by the term in the braces in Equation (10), i.e.,

$$Q_2(\vec{k};\vec{x}_o) = \iiint_{-\infty}^{\infty} e^{i\vec{k}\cdot(\eta - x_o)} I(\vec{\eta}+\vec{c}_s t', \vec{x}_o, t') dt' d\vec{\eta} \tag{11}$$

and it contains all of the frequency-shaping, phase-shifting and demodulation that the SAR process applies to the ς-field in the Fourier domain. In this steady-state formulation, all of the fluid-dynamically interactive effects of the corrugating field and its currents on the ς-field are contained in $p(\vec{k})$ and need not be expressed explicitly.

To progress further with the Kelvin wake problem it is necessary to return to Equations (2) to (5) and expand the expressions for r_1, r_2 and r_o using the definitions

$$x' = x_o + \mu \quad , \qquad y' = Vt' + y'' \quad , \tag{12), (13}$$

$$y_o = V_1 t' + q'' \quad , \qquad z = z + \varsigma_t(\vec{\eta}) \quad , \tag{14), (15}$$

$$R_o = \sqrt{x_o^2 + z_o^2} \quad . \tag{16}$$

Here ς_t represents only the transverse (corrugating) part of the Kelvin wake.

With terms expressed up to and including second order in μ, y", q" and ς_t,

$$r_1 = r_2 = R_o + \frac{\mu x_o + \varsigma_t z_o}{R_o} + \frac{u^2 + y''^2 + \varsigma_t^2}{2R_o} - \frac{x_o^2 \mu^2}{2R_o^3}$$

$$- \frac{\mu \varsigma_t x_o z_o}{R_o^3} - \frac{z_o^2 \varsigma_t^2}{2R_o^3} + \ldots . \tag{17}$$

$$r_o = R_o + \frac{q''^2}{2R_o} + \ldots . \tag{18}$$

and, from Equations (2) and (12) to (14)

$$I(\vec{x}', x_o, t') = I(\mu, y'', \vec{x}_o, q'') \tag{19}$$

$$= - \frac{z_o^2 \omega_c^2}{4\pi^2 \omega_c^2 R_o^4 \Delta} \exp\left[\frac{2i\omega_c}{c} \frac{x_o^2}{R_o} + \frac{i\omega_c}{c} \Phi(\mu, y'', q'') \right.$$

$$- \frac{y''^2 + [y'' + (V-V_1)/V_1 (y_o - q'')]^2}{Y_g^2}$$

$$\left. - (\mu x_o + \varsigma_t z_o)^2 (\omega'^2 T_g^4 - 1)/2c^2 T_g^2 R_o^2 \right] \tag{20}$$

where

$$\Phi(\mu, y'', q'') = r_1 + r_2 - 2r_o = 2\frac{\mu x_o + \varsigma_t z_o}{R_o} + \frac{\mu^2 z_o^2}{R_o^3} + \frac{y''^2 - q''^2}{R_o}$$

$$+ \frac{\varsigma_t^2 x_o^2}{R_o^3} - 2\frac{\mu \varsigma_t x_o z_o}{R_o^3} \tag{21}$$

and where, by Equations (6), (12) and (14)

$$\mu = \eta_x - x_o + c_{sx}(y_o - q'')/V_1 \quad , \tag{22}$$

$$y'' = \eta_y + (c_{sy} - V)(y_o - q'')/V_1 \quad . \tag{23}$$

Here, η_x and η_y are the components of $\vec{\eta}$, and c_{sx} and c_{sy} are components of $\vec{c}_s$.

The integration in t' in Equation (11) can be performed after converting t' to q" with the use of Equation (14), and the result is

$$Q_2(\vec{k};\vec{x}_o) = \iint_{-\infty}^{\infty} e^{ik\cdot(\eta-x_o)} K(\vec{\eta},\vec{x}_o)\,d\vec{\eta} \qquad (24)$$

where

$$K(\vec{\eta},\vec{x}_o) = -\frac{z_o^2\omega_c^2}{4\pi^2c^2R_o^4\Delta V_1}\frac{\pi T_g}{\sqrt{2|c_1|}}e^{A_1+B_1^2/4C_1} \qquad (25)$$

To obtain this expression, the argument of the exponential in Equation (20) has been written as

$$\arg = A_1+B_1q"-C_1q"^2 \qquad (26)$$

and, by substitution of Equations (22) and (23) into (21) and (20) and subsequent inspection, the coefficients A_1, B_1 and C_1 can be readily obtained. They will not be expressed explicitly here, but, because of the original quadratic-form expansions of r_1, r_2 and r_o, it can be noted that arg itself is, at most, quadratic in products of $\vec{\eta}$ and q", and so, A_1 is quadratic in $\vec{\eta}$, B_1 is linear in $\vec{\eta}$ and C_1 is constant. The same orders exist for ς_t. A further simplification will also be made to these coefficients, namely that terms quadratic in the small quantities $|\vec{c}_s|/V$ and ς_t/R_o will be ignored. For the usual range of SAR parameters, these quantities are of order 0 1 or less and the ignored quadratic terms are always dominated by linear counterparts.

3. Results

Without specifying the functional form of ς_t on $\vec{\eta}$, the Fourier transform represented by Equation (24) cannot be carried out. However, insight can be gained into possible effects of ς_t in the resulting image, by examining the special case in which

$$c_{sx} = 0 \qquad (27)$$

i e., the source ship travelling parallel to the SAR aircraft track, and

$$\varsigma_t = \varsigma_o + \eta_y\left(\frac{\partial\varsigma_t}{\partial y}\right)_o + \frac{1}{2}\eta_y^2\left(\frac{\partial^2\varsigma_t}{\partial y^2}\right)_o \quad . \qquad (28)$$

(For simplicity, the processor mismatch is also set equal to zero, i.e., $V=V_1$.)

The surface underlying the Bragg-wave is thus given a tilt and a curvature in the azimuthal direction (only). The argument of the exponent in Equation (25) can be put in a quadratic normal form

$$\arg = A - [(\eta_x-x_s)/Res_x]^2\ln 16 - [(\eta_y-y_s)/Res_y]^2\ln 16 \qquad (29)$$

where Res_x and Res_y represent full-widths at half-maximum and x_s and y_s represent the position in $\vec{\eta}$ at which the maximum occurs. For the case presented by Equations (27) and (28),

$$Res_x = \frac{\frac{cR_o}{\omega' x_o T_g}\sqrt{\ln 4}}{(1+1/\omega'^2T_g^4)^{1/2}} \qquad (30)$$

$$Res_y = \frac{\frac{2cR_o}{\omega_c T_g}\sqrt{\ln 4}}{(1-4(\frac{\omega' T_g}{\omega_c})^2(\frac{x_o}{Y_g})^2 z_o(\frac{\partial^2\varsigma_t}{\partial y^2})_o)^{1/2}}(1+\frac{c_{sy}}{V}) \qquad (31)$$

The expression for Res_x is unmodified by the presence of ς_t (to this order), but it can be seen that Res_y can be significantly altered by ς_t's curvature. This gives rise to the possibility of brightened spots appearing in a wake image at positions where there is a coincidence of maximum (positive) $\partial^2\varsigma_t/\partial y^2$, and Bragg scatterers in the diverging wake components.

For stronger conclusions the form of ς_t will be specified more fully and numerical evaluations will be performed.

The specific values chosen for numerical treatment pertain to the Canada Centre for Remote Sensing L-band system (originally constructed by the Environmental Research Institute of Michigan) as configured in July 1983 and used in overflights of Dabob Bay, WN (Hughes and Dawson, 1988). For JOWIP 8/2

$$\omega_c = 2\pi\ (1.185)10^9 \text{ rad/s} \qquad (32a)$$
$$\omega' = 2\pi\ (33.3)10^{12} \text{ rad/s} \qquad (32b)$$
$$x_o = 3033.5 \text{ m} \qquad (32c)$$
$$z_o = 7010.0 \text{ m} \qquad (32d)$$
$$y_o = 0 \text{ m} \qquad (32e)$$
$$Y_g = 162.1 \text{ m} \qquad (32f)$$
$$T_g = 0.951\ 10^{-6} \text{ s} \qquad (32g)$$

where the last two values are Gaussian equivalents which result in unperturbed resolution widths of $\sqrt{20}$ m in both directions. For JOWIP 8/2,

$$V = 134 \text{ m/s} \qquad (32h)$$
$$c_s = 8 \text{ m/s } -13^\circ \text{ off azimuth} \qquad (32j)$$

and the amplitude of the corrugating wave had a value of

$$\varsigma_o \approx 0.1 \text{ m} \qquad (32k)$$

at a distance of ~2.5 km behind the source ship (USS QUAPAW) (Hammond, et.al., 1985). For this ship speed, the corrugating transverse wake component wavelength is

$$\lambda_s = 41 \text{ m} \qquad (32m)$$

calculated from the gravity wave formula $2\pi c_s^2/g$.

The corrugating wavefield ς_t is modelled as a simple sinusoid. This idealization is used because variations in actual ς_t's perpendicular to the wake axis are minimal particularly near the axis where the L-band Bragg waves are located in the diverging field. Thus,

$$\varsigma_t = \varsigma_o\cos(\frac{2\pi}{\lambda_s}[\eta_x\sin\theta+\eta_y\cos\theta]+\phi) \qquad (33)$$

where θ is the angle of the ship's track with respect to the aircraft track (13° for the present example), and ϕ is a parameter that allows the full range of phase of ς_t to be investigated numerically.

In the absence of a corrugating perturbation, Q_2 (Equation (24)) is a general two-dimensional Gaussian form in $\vec{k}$ and it can be described by its maximum height, the wavenumber position at maximum, and the widths along the $\vec{k}$-axes to the half-maximum points. This is illustrated in contour form in Figure 1. If

the peak is also rotated, the angle of rotation is necessary to complete the description. In the presence of the perturbation, these parameters change, and the effect of the perturbation can be described fully by the amounts of the changes. Figure 2 shows the changes in form of the half-maximum contour of Q_2 for different phases, i.e. different values of ϕ in Equation (33), along the corrugating wave. Figures 1 and 2 both use the parameters specified in Equations (32a) to (32m) except for Equation (32k). Figure 1 uses ς_o=0 and Figure 2 uses ς_o=0.5 m.

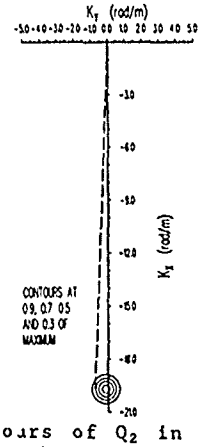


Figure 1. Contours of Q_2 in the absence of a corrugating wave (i.e. ς_o=0) but with the source travelling along θ=13°. The 1st-Bragg wavenumber for this case is (-19.712..,-0.13238..) rad/m. The dotted line indicates the extreme in propagation direction (2.20°) for Q_2 at its half-maximum.

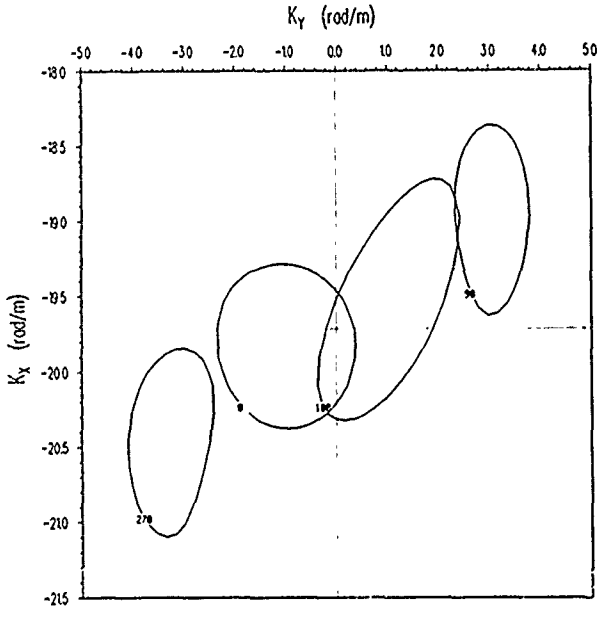


Figure 2. The half-maximum contour of Q_2 for various phases (degrees) along a corrugating wave of amplitude ς_o=0.5 m propagating at θ=13° to the azimuthal direction. Note different scales from Figure 1.

Without a corrugating wave, it can be seen from Figure 1 that surface waves with a wavenumber centered at -19.712..,-0.132 rad/m and within a circular band of width ±0.6240...rad/m are accepted by the processor with a gain of between 1 and 0.5. This bandwidth implies that all waves propagating at an angle within ±1.81° centered at 0.38° off this range direction and with a k_x-value of -19.712 rad/m, are accepted by the processor with a gain reduction of no more than a factor of 2. (The slight offset of the peak from k_y=0 is due to the non-zero value of θ, i.e. the non-zero value of c_{sx}. This range-component of motion in the Bragg scattering wavefield coupled with the aircraft motion and time interval between successive radar pulses results in a slight skewing of the Bragg-wave field in the SAR pattern.)

With a corrugating wave as used in Figure 2, it can be seen that the surface wave wavenumbers accepted by the processor vary over a much more considerable range, allowing waves propagating at up to 11.5° to the range direction to be "seen" (at ϕ=90°), with a gain reduction of no more than a factor of 2 from the peak value. This variation is depicted more fully in Hughes (1989).

The extreme k_y-values are largest for ϕ≈110° (and smallest at 290°). Using the central k_x-value of -19.712 rad/m, the extreme values represent surface waves propagating at angles with respect to the range direction as shown in Figure 3. Here the propagation angles are given as a function of ς_o, and propagation directions that pertain to the center of the processor peak are also given. The displayed data shows strongly linear dependence with ς_o, and very little variation for θ-values over the range θ=0° to 13°.

The peak values of Q_2 for ς_o = 0.5 metres vary only over the range 3.35 to 4.90, only 1.6 dB, and so the direct variation of Q_2 is a relatively minor amount. For the Kelvin wake problem, the periodic variation of the filter window in k_y-space is potentially much more significant.

For experimental comparisons, there are three images of clear narrow-V wakes that were taken during the JOWIP measurements, those designated JOWIP 8/2, 8/3 and 9/2. They are shown in Figure 4. From these images the orientation of the most steeply inclined

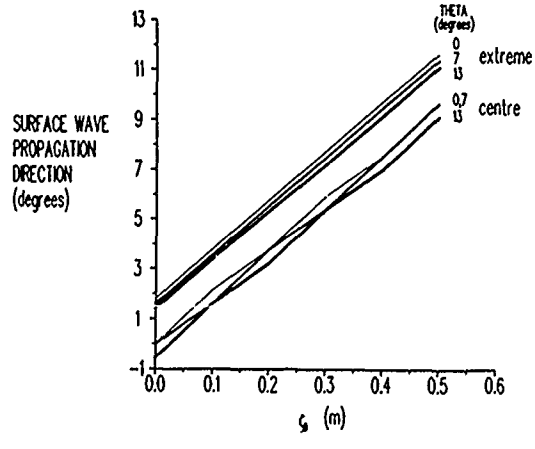


Figure 3. Surface wave propagation direction (as measured from the range coordinate) versus amplitude ς_o of the corrugating wave.

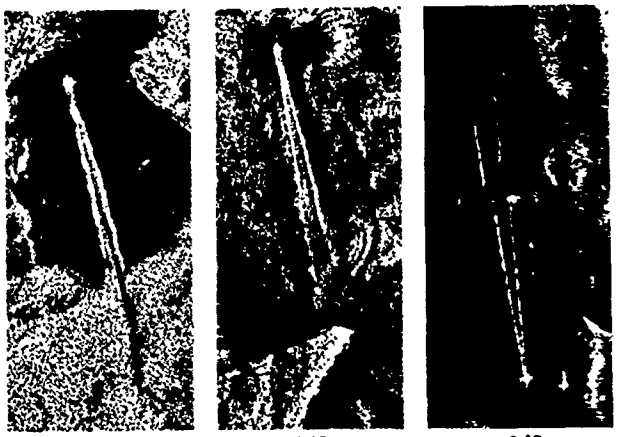


Figure 4. SAR images of narrow-V wakes.

narrow-V arms for each pass have been measured, with respect to the SAR aircraft track, and the values are given in Table 1. Ship track angles as determined from the positioning equipment are also included for each of these runs (Hughes and Dawson, 1985).

Table 1. Ship Track and Steepest Wake Arm Inclination to SAR Track

SAR Pass	Wake Arm Angle	Ship Track Angle
8/2	18°	13°
8/3	15°	11°
9/2	11°	9°

It can be seen from the Table that the wake arm inclinations are of the same order or are larger than the processor acceptance angles (at half-maximum) for ς_o=0.5 m. However, the variational effects just described will contribute strongly toward enhanced (or decreased) outputs in the SAR image even for these cases.

4. Discussion

The SAR images of Kelvin wake narrow-V's in the SAR imagery are considered to be partly due to some randomization and intermittency in the wake structure itself, as is inferred from experimental evidence (Wyatt and Hall, 1988), however, the peculiarities of the SAR processor described above do provide an alternative mechanism for enhancement of local regions of the wake. The fact that these are predicted to occur once per cycle of the transverse wave, and for the enhancements in each of the two arms of the narrow-V to occur at the same transverse wave phase angle, also lends credence to the validity of the effect, because these are in accordance with experimental data. No attempt has been made to model fully the Kelvin wake field, transverse and diverging, because it is known that accurate results for the short wavelength components are very difficult to achieve, requiring complicated source descriptions at the vicinity of the source ship, and also because of the intermittency just referred to. Indeed, the predicted SAR processor variations are not large enough by themselves, but combined with the intermittency and randomization of the direction of the diverging wavefield, they could possibly provide the mechanism underlying the appearance of the brightened V-arms.

The curvature effect examined in Section 3, Equations (28) to (31), can be much larger than traditional velocity bunching effects. The latter are known to produce radar cross section aberrations that are given by

$$|\sigma_{VB}| = z_o\left|\frac{\partial^2 \varsigma_t}{\partial y^2}\right| \frac{c_{sv}}{V_1} \tag{34}$$

where the pattern is considered to be moving in the y-direction as a solid body (Hammond, et.al., 1985). The comparable term from the curvature effect is

$$|\sigma_{cur}| = z_o\left|\frac{\partial^2 \varsigma_t}{\partial y^2}\right| 2\left(\frac{\omega' T_g x_o}{\omega_c Y_g}\right)^2 \tag{35}$$

which is obtained by integrating $K(\vec{\eta},\vec{x}_o)$ over all $\vec{\eta}$, as if the Bragg-scatterers covered the entire pixel area uniformly, and using Equations (29), (30) and (31). In Equations (34) and (35) only the perturbations in σ are given. The denominator in Res_y from Equation (31) is also expanded binomially and the first non-unity term kept.

For the numerical values of Equations (32a-j), the relative magnitudes of σ_{VB} and σ_{cur} are 0.06 and 0.5, each of these values coming from the last terms in Equations (34) and (35) respectively. This evaluation is only for a very specific kind of surface (parabolic within the pixel area) and so is only indicative of possibilities. But it does show that curvature effects are not ignorable, a priori.

The perturbation in the SAR processor output due to the curvature of the corrugating wave can be understood from a simple geometrical picture of the radar wave at the sea surface. In the absence of the corrugation, a field of plane radar waves incident on the sea surface at an angle θ_{inc} intersect the surface in a series of straight lines as shown in Figure 5(a). In the presence of a corrugation propagating perpendicularly to the radar waves (in the horizontal), the intersection of the radar waves at the surface also becomes corrugated, Figure 5(b). The corrugation appears not only in the vertical but in the horizontal as well (for $\theta_{inc} \neq 0°$ or 90°), and because of this the local horizontal wavenumber directions associated with Bragg scattering can be considerably deflected azimuthally from that of the main radar beam, Figure 5(b).

From Figure 6(a), it can be seen that, with z upwards, the radar crest that just intersects the sea surface, as shown, does so at

$$z_1 = (x_r - z\tan\theta_{inc})\tan\theta_{inc} \tag{36}$$

If the surface height is specified by a long surface wave with height ς given by

$$\varsigma = \varsigma_o\cos(\vec{k}\cdot\vec{x}+\phi) \tag{37}$$

where ς_o is the wave amplitude, $\vec{k}$ is its wavenumber, $\vec{x}=(x_r,x_a)$ are range and azimuth coordinates, and ϕ is an arbitrary phase angle, then the intersection takes place at

$$z_1 = \varsigma \tag{38}$$

which, by Equations (36) and (37) is at

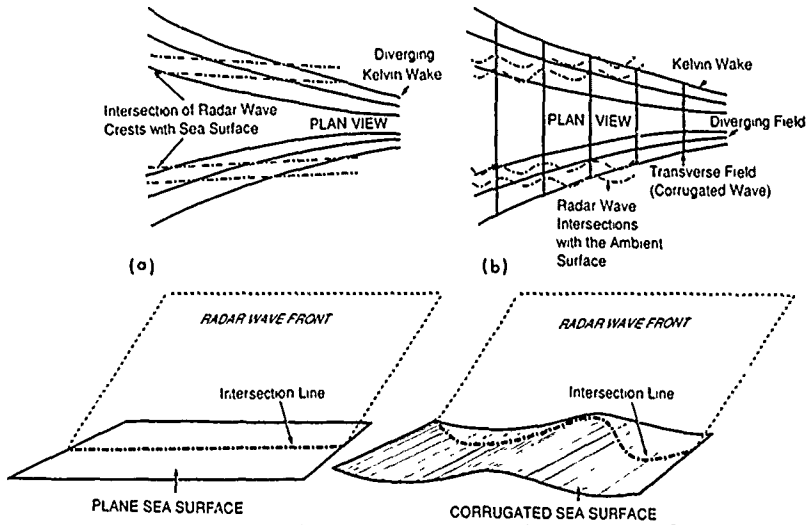


Figure 5. Intersection of the radar beam and the ocean surface (a) non-corrugated, (b) corrugated.

$$x_r - z\tan\theta_{inc} = \frac{\varsigma_o}{\tan\theta_{inc}}\cos(\vec{k}\cdot\vec{x}+\phi) \quad . \tag{39}$$

This last can be solved for x_r in terms of x_a (in principle) giving the horizontal shape of the intersection lines. For a corrugation propagating along the azimuth, $\vec{k}\cdot\vec{x} = k_a x_a$ only, and

$$x_r = \frac{\varsigma_o}{\tan\theta_{inc}}\cos(k_a x_a - \phi) + z\tan\theta_{inc} \quad . \tag{40}$$

Typical plan view intersection lines are shown for this case in Figure 6(b).

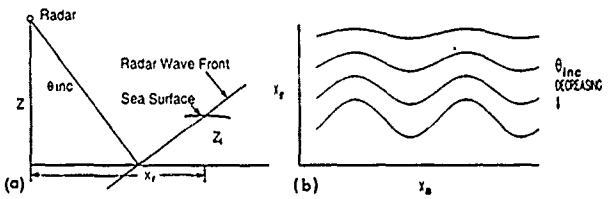


Figure 6. (a) Geometry of intersection (elevation view). (b) Typical horizontal intersection lines for a radar field in the presence of an azimuthally propagating corrugation (plan view).

The slopes of these lines provide estimates of the amounts of azimuthal deflection for the effective radar beam in the Bragg scattering process. From Equation (40) the slopes are

$$\frac{dx_r}{dx_a} = -\frac{2\pi\varsigma_o}{\lambda_s\tan\theta_{inc}}\sin\left(\frac{2\pi x_a}{\lambda_s} + \phi\right) \tag{41}$$

where λ_s is the wavelength. For the values of Equation (32a-m) but with $\varsigma_o = 0.5$ m,

$$\left|\frac{dx_r}{dx_a}\right| = 0.177 = \tan 10^\circ \tag{42}$$

This is in accordance with the results of Section 3 where, from Figure 3, the centre of the SAR processor at $\theta=0^\circ$ is aligned with waves propagating at 9.6° (at the maximum deviation).

5. Conclusion

The main conclusion is that large scale surface curvatures can exert a major influence on the processed SAR imagery of Kelvin wakes. For particular cases, as examined in Section 2, the curvatures associated with the corrugating effect of the wake's transverse wave component can readily rotate the processor Bragg acceptance angle from being boresighted with the radar beam to being ~10° fore or aft of the beam. In modelling SAR imaging of Kelvin wakes, particularly for narrow-V effects, this mechanism should be included.

References

1. Hammond R.R., R.R. Buntzen and E.E. Floren. "Using Ship Wake Patterns to Evaluate SAR Ocean Wave Imaging Mechanisms", Technical Report 978, Naval Ocean Systems Center, San Diego, CA, 92152-5000, 1985.

2. Hughes B.A. and T.W. Dawson. "Joint CAN/US Ocean Wave Investigation Project (JOWIP), the Georgia Strait Experiment: An Overview", J. Geophys. Res., 93, C10, 12219-12234, 1988.

3. Hughes B.A. "Speckle Noise and Pattern Detection in SAR Imagery of Internal Waves", Technical Memorandum 81-12, Def. Res. Est. Pac., FMO Victoria, B.C., Canada, VOS 1BO, 1981.

4. Wright J.W. "A New Model for Sea Clutter", IEEE Trans. Ant. and Prop., AP-16, 217-233, 1968.

5. Hughes B.A. "First-Bragg SAR Imagery of Moving Internal Wave Fields, and Kelvin Wakes", Def. Res. Est. Pac. Technical Memorandum, in preparation.

6. Hughes B.A. and T.W. Dawson. "DREP Measurements of Surface Ship Wakes in Dabob Bay, WN, July 1983", Report 85-1, Def. Res. Est. Pac., FMO Victoria, B.C., Canada, VOS 1BO, 1985.

7. Wyatt D.C. and R.E. Hall. "Analysis of Ship-Generated Surface Waves Using a Method Based Upon the Local Fourier Transform", J. Geophys. Res., 93, C11, 14133-14164, 1988.

THE USE OF THE RADON TRANSFORM FOR WAKE DETECTION IN SEASAT IMAGES

M.T. Roy, J.T. Folinsbee, R.D. Ball, J.K. Tunaley, M.R. Vant

M.T. Roy, Defence Research Establishment Ottawa, J.T. Folinsbee, Imago Manufacturing Limited, R.D. Ball, Imago Manufacturing Limited, J.K. Tunaley, London Research and Development, M.R. Vant, Defence Research Establishment Ottawa

ABSTRACT

A moving ship produces a set of waves in a characteristic linear 'V' pattern. This pattern, or some of its components, can often be detected in ocean imagery produced by satellite-borne Synthetic Aperture Radar (SAR) sensors operating at L-band. As ship wake detection can provide information such as ship direction and speed, the detection of these wakes can play an important role in satellite surveillance of shipping.

This paper describes research done on the use of the Radon Transform to automatically detect ship wakes in SEASAT-A SAR ocean images. In this work, ship wakes and surrounding ocean scenes were extracted from SEASAT-A SAR imagery. These images included other naturally-occurring ocean phenomena. A variety of Automatic Detection Algorithms (ADA) based on the Radon Transform were developed and tested to improve the probability of detection of ship wakes and reduce the false alarm rate. These included line length normalization, automatically analyzing the shape of the detected peaks in the Transform, applying various semblance filters to the image, and high-pass and Weiner filtering of the data in the Transform domain. To date the use of an ADA which incorporates a high-pass filter followed first by a Radon Transform and then by a Wiener filter has been shown to reliably distinguish wake peaks from false alarm peaks. Keywords: Radon, SEASAT, Wake

1.0 INTRODUCTION

Ship wakes can often be detected in ocean imagery produced by satellite-borne Synthetic Aperture Radar (SAR) sensors operating at L-band. SEASAT-A SAR data show these wakes as straight lines on the ocean background, often extending for 5 to 15 kilometers aft of the ship. Because of the motion of the ship relative to the SAR platform, the ship often appears displaced in azimuth from the wake.

The wakes seen in SAR images can be divided into components consisting of the classical Kelvin wake and the turbulent wake. The turbulent wake comprises the turbulent region directly behind the ship.

Ship wake characteristics can provide information such as ship direction and speed. Detection of the wake may also be useful as a means of inferring the presence of a ship when the ship image itself is undetectable.

The evaluation of ship wake detection techniques was based on the probability of detecting a wake (PD), and the false alarm rate (FAR) of the detection algorithm. False alarms can be caused by: internal waves generated by salinity or thermal gradients, features of the underwater topography, and wind waves. However, naturally occurring ocean phenomena are unlikely to remain linear for lengths on the order of the ship wake length.

1.1 THE RADON TRANSFORM

The Radon Transform of a continuous image is defined as:

$$f(\rho,\theta) = \int_D g(x,y)\,\delta(\rho - x\cos\theta - y\sin\theta)\,dx\,dy$$

where D is the entire image plane [1]
$g(x,y)$ is the grey level at position (x,y)
$\delta(\cdot)$ is the Dirac delta function
ρ is the radius coordinate of a straight line and
θ is the angle coordinate of a straight line.

The Radon Transform converts a point in image space to a sinusoid in the Transform space representing all the lines to which the point could belong. The sinusoids of collinear points in the image space intersect at a point in the transform space. This property allows the enhancement of linear features since collinear pixels in the image, which are on average of different intensity than the background, produce a bright (higher than average) or dark (lower than average) peak in the transform space. Therefore a bright or dark spot in the transform domain corresponds to a bright or dark line in the image. The integration process averages out noise in the Transform domain, so the Signal to Noise Ratio is greater in the Transform domain than in the image. Also, the process does not depend on a line being continuous in nature.

1.2 RESEARCH OBJECTIVE

As ship wakes are linear features, it was decided that the Radon Transform would be an appropriate algorithm to apply to their detection. However, wakes must be distinguished from naturally-occurring linear features which will also be detected by the application of the Radon Transform. Therefore the objectives of this work were twofold: first, to determine the utility of the Radon Transform for detection of linear features in SEASAT wake imagery,

and secondly, to reliably distinguish wakes from other linear ocean features.

2.0 SEASAT SAR IMAGERY

The images used in this study were extracted from digitally processed SEASAT ocean imagery, supplied as complex 32 bit image data. These were converted to magnitude data, then linearly compressed to a 256 grey level image for analysis.

The set included both images containing ship wakes and images showing only the ocean background. When acquiring the data, land features were used to determine the coordinates of ship wakes of interest in the original SEASAT data. Hence, the wake data used in this study consist mostly of images from coastal regions, where linear features induced by the underwater topography are abundant.

This study was conducted in two parts. The first part [2], involved the evaluation of the use of the Radon Transform and Automatic Detection techniques. During the evaluation the techniques were applied to 19 wake images. The second part [3], currently underway, involves the evaluation of the use of an Automatic Detection scheme based on high-pass and Weiner filtering. These techniques are being tested on a set of 32 distinct wake images.

Ships appear in the images as small areas of extremely bright returns, usually associated with one or more visible wake components. These wakes are frequently displaced from the ship return. This occurs because SAR uses the Doppler shift of the return target signal to determine the position of the target in azimuth, hence the Doppler shift due to the velocity of the ship may introduce an error into its calculated position. The exact offset depends on the speed and direction of the ship's travel.

The wakes themselves appear as bright or dark lines, often discontinuous, generally in a V-shape pattern. Not all wake components are visible in all images, possibly because of the orientation of the ship wake to the SEASAT sensor, or varying ocean conditions. The most frequently visible wake components are the dark turbulent wake and the bright wake arms.

The evaluation of the wake detection algorithms ideally requires comparison with independent data on the position and characteristics of the ships and wakes in each image. Such data were unavailable for the images, so the comparison was made against the positions of visually detected wake components. The positions of automatically detected wake components were compared with those of the manually detected components. If they corresponded, the Automatic Detection was considered to have detected the component. If they did not correspond, it was considered to be a false alarm. The visual inspection could not differentiate between bright and dark wake components, since some components consisted of closely spaced bright and dark lines. Therefore in the evaluation, a distinction was not made between bright and dark components. The Automatic Detection Algorithm (ADA) would consider a detection as being correct regardless of whether it was truly dark, or truly bright. It could not tell the difference.

3.0 EVALUATION OF THE RADON TRANSFORM

3.1 DETERMINATION OF OPTIMUM IMAGE SIZE

The performance of the Radon Transform in the detection of linear features is optimised when the length of the feature of interest is comparable to the size of the image. It was determined for our data that an image size of approximately 400x400 pixels (6.4 km x 6.4 km) gave optimum results. Based on these results, a standard set of 19 400x400 pixel wake images was chosen for use in the preliminary testing. Another set of 400x400 adjacent ocean images containing no visually obvious wake components were also extracted to determine the performance of the Radon Transform on apparently uniform ocean regions.

3.2 EFFECT OF SHIP RETURNS ON TRANSFORM PERFORMANCE

The ships appear in the imagery as small areas of extremely bright returns. These 'swamp' the transform with very bright sinusoids, making detection of wake peaks difficult. To alleviate this problem a threshold was determined and incorporated into the Radon Transform, so that these very bright pixel values would be ignored in determining the transform.

3.3 RADON TRANSFORM PERFORMANCE

A Radon Transform was performed on the preliminary image set, with image mean subtracted and threshold level set to a predetermined value to eliminate ship returns. An ADA was then applied to the data. The ADA first calculated the mean, m, and standard deviation, σ, of the Radon transform of an image. A threshold was calculated as $m \pm (K*\sigma)$, the plus or minus sign depending on whether bright or dark peaks were to be detected. In the threshold equation K is a user-defined parameter, generally set equal to 4. The transform was then processed with this threshold to produce a bitmap of the bright or dark region of interest. Each connected bright or dark region represented a detected wake component. The position and direction of the component were defined by the position of the maximum or minimum of the transform in the region under the bitmap.

The transforms of the uniform ocean areas did not show many features, although the sinusoidal nature of the transformation of each image point was usually visible in the transform space.

The transforms of the ship wake data showed, as expected, bright or dark sinusoids converging at a point corresponding to each of the ship wake components. The probability of detecting a component as a bright wake was 0.5, and as a dark wake, 0.25. Overall the ADA detected 21 out of 30 wake components (bright and dark wakes combined) for an overall PD of 0.7. The sum of the bright and dark wake PD is greater than the overall PD since some components were detected as both bright and dark wakes. Visual examination of the transforms detected 29 out of 30 wake components from their associated peaks. In general, the ADA did not detect the missed peaks found in the visual inspection because they fell below the selected 4σ threshold. Lowering this threshold improved the ADA performance for these peaks, but also increased FAR substantially. It was found that at least one of the wake components, in any wake, could be detected with a PD = 1 in all the images. The FAR was typically 3.8 false alarms per 400x400 pixel image for bright wakes, and 1.6 for dark wakes. As the use of the Radon Transform for wake detection is

ultimately under consideration for satellite surveillance of shipping applications, where large ocean areas must be rapidly searched, the FAR values were unacceptably high. It was therefore concluded that improvement must be made to the ADA to increase the PD and reduce the FAR.

4.0 IMPROVEMENTS TO THE BASIC ADA

Based on the conclusions drawn above, several techniques were developed and tested as methods of increasing the PD and reducing FAR. These techniques included: normalizing the transform, doing peak shape measurements, using directional semblance, and using high-pass and Weiner filtering.

4.1 LINE LENGTH NORMALIZATION

The lengths of the lines that are integrated in the previously defined Radon transform depend on the position in the image; lines crossing a corner of the image are shorter than lines through the middle. This produces a position bias in the ADA; wakes along shorter lines will be less likely to produce peaks exceeding the threshold. To reduce this problem the transform values were normalized to the lengths of the associated images lines, so that the transform value was the average grey level along each line rather than the sum of the grey levels.

This process produced a slight decrease in FAR with no change in PD. The sample size was not large enough to determine if the changes were significant.

4.2 PEAK SHAPE MEASUREMENTS

In the data analysis thin linear ocean features showed up as a 'bowtie' shape around the central peak in the Radon transform while broader linear features and non-linear features showed up poorly or not at all. Therefore it was believed that algorithms which examine the peak shape in the Radon Transform might assist in distinguishing ship wakes from false alarms.

An algorithm to measure the peak shape of the detected transform peaks was developed and tested on the initial image set. Using this algorithm the strength of the lines through the transform peak was found by averaging the two pixels on either side of the peak in each of the six possible directions. For true wake components the maximum strength directions and minimum strength directions tended to be at right angles, and the differences were quite large. These results were used to determine a simple set of rules for discrimination of peak shape. These rules were incorporated into the ADA. The results of the application of the ADA to the preliminary data set showed a strong decrease in FAR, and hence looked promising as a means of enhancing ADA performance. However, when the technique was examined for the more recently acquired data set, the peaks associated with less obvious wakes or wake components did not exhibit a well-defined bowtie shape, and the technique was therefore abandoned.

4.3 SEMBLANCE TECHNIQUES

Semblance is a second order statistic defined as:

$$S = (\sum_{W} g)^2 \; / \; N \sum_{W} g^2$$

where: g is the pixel grey level value
w is the region or window the sum is taken over, and
N is the number of samples in the window.

Semblance ranges between 0 and 1, and is a measure of the similarity of the data within the sample window.

Previous work indicated that semblance techniques may be useful for enhancing the visibility of wakes in SEASAT images. Hence it was decided to investigate the use of semblance filters on our data.

Various semblance preprocessors were developed and tried on our data. The most promising of these techniques was the Directional Semblance Filter (DSF). This filter is a modification of the semblance algorithm tailored to preferentially detect the long narrow linear wake features. It determines the semblance for a set of long narrow windows centred on each pixel, covering the entire range of directions. It can be combined with a Radon transform to produce an output similar to that of the basic Radon Transform, but containing more directional information. The definition of the DSF is:

$$S(K,1,D) = [\sum_{W(d)} f(i,j)]^2 / (N \sum_{W(d)} [f^2(i,j)])$$

where W(d) is a window in the image centred at (k,1) and lying along a direction defined by d.

Comparison of peak heights for detected wakes indicated that the semblance filtering enhanced dark wakes but degraded very bright wakes. Faint bright wakes were not degraded and in some cases were enhanced by the semblance process. However, application of the semblance techniques also increased FAR, because the dark line sensitivity was drastically increased. This caused the detection algorithm to find many spurious dark lines in the image, which were not associated with the ship wake.

The application to the data of line length normalization and peak shape detection filters offset the increased FAR values associated with the application of directional semblance techniques.

4.4 HIGH-PASS AND WEINER FILTERING

The techniques developed on the preliminary data set were reevaluated on a new data set of 37 distinct SEASAT wakes. The use of peak shape detection algorithms did not produce the results hoped for when applied to the larger data set.

The wake data were first inspected to determine the Signal to Noise (S/N) Ratio of the wake features vs. the ocean background in the Radon domain. Although the wake components are detectable in that they are the highest or lowest points in the transform, the (S/N) Ratio, defined as the ratio of the wake maxima or minima to the overall standard deviation of the transform, was small.

Preliminary testing using a simple running mean filter with varying window sizes on a few wake images showed an increase in the S/N ratio for some window sizes. This indicated that there existed significant broad features in the ocean background. Therefore it was decided to test a high-pass filter on several wake images, which were first compressed by two to increase processing speed. A broad range of window sizes was used for the filtering, ranging from 10x10 pixels to 30x30 pixels. This filtering produced a significant increase in the SNR of the filtered vs. the unfiltered images, and hence a significant increase in the detectability of the wake components. It was found

that detectability was dependent on the size of the window relative to the length of the wake components in the image. As the window size approaches the size of the image, the effect of the filtering disappears. The detectability of dark and bright wakes was enhanced using this technique.

The improvement in detectability of wake components vs. false alarms, obtained using the high pass filter, is very significant because the PFA falls off very rapidly with an increase in the SNR. For example, a 10% increase in the S/N ratio (defined on the previous page) from 6.0 to 6.6 standard deviations of the summed amplitude in the Radon domain, brings an order of magnitude reduction in the Probability of False Alarm.

The above results led to a new Radon Transform system for detecting wakes in an ocean background and distinguishing them from false alarm peaks. The individual steps in this system are discussed below.

If the image is large, an image compression program is run; this replaces the pixels in a square window by a single pixel containing their sum. This dramatically reduces the time and complexity of the subsequent steps without significantly altering the results obtained. Next, a high pass or running mean filter is used to remove broad features in the sea background. Then a normalized Radon transform is applied. Following this, an automatic peak detection algorithm is employed, which consists of a simple threshold combined with a local maximum or minimum detector. The peaks returned by this detection algorithm are Wiener filtered using pixels in a 7x7 square window around the peak. The effect of the filter is to produce a further separation in the S/N ratio between wake and false alarm peaks of between 10 to 20%. To date, of the 37 wake images evaluated with this technique, all wake components are detected with a false alarm rate of zero.

5.0 CONCLUSIONS

The Radon Transform combined with a simple Automatic Peak shape detection algorithm can detect almost all visible wakes in an image. However the False Alarm Rate associated with the use of these techniques is unacceptably high for satellite ship surveillance applications.

The use of Directional Semblance combined with line normalization and peak shape detection was of use in detecting wake components reliably, however, the associated false alarm rates were still unacceptably high.

The use of an ADA based on the application of a high-pass filter, Radon Transform and Weiner filter is the most promising technique to date, giving a PD of 1 and FAR of zero for the 37 wakes to which it was applied. The high-pass filter significantly improved the SNR of the filtered vs. unfiltered wake images by removing many of the broad features in the ocean background. The Weiner filter further enhanced the separation in magnitude between wake and false alarm peaks.

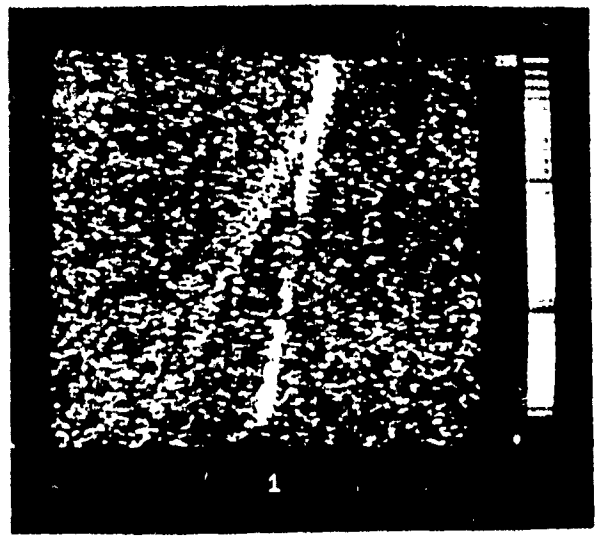


Figure 1. Image of ship wake, showing bright turbulent wake and other wake arms. Note ship is at convergence of wake arms.

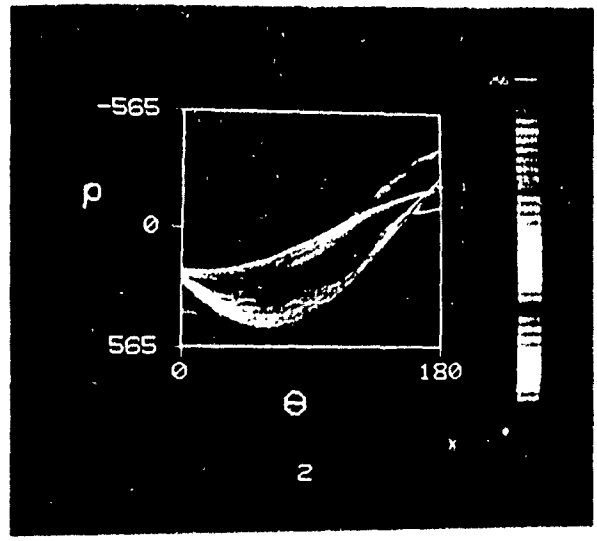


Figure 2. Radon Transform of 1. Bright sinusoids are produced by ship returns.

6.0 REFERENCES

1. Deans, S.R. "Hough Transform from the Radon Transform", IEEE Trans. on Pattern Analysis and Machine Intelligence, PAM-3, 2, 1981.

2. Folinsbee, J.T. "Radon Transform Techniques for Wake Detection in SAR Images", for Defence Research Establishment Ottawa under Contract File No. W7714-6-5226/01-SV March 31, 1988.

3. Tunaley, J.K.E.; Jahans, P.A. and Dixon, J.A. "Evaluation of Radon Transform Techniques in the Detection and Enhancement of Linear Wake Features in SAR Wake Data - Part 2", for Defence Research Establishment Ottawa under Contract File No. W7714-8-5562/01-SV March 31, 1989.

PRINCIPLES AND PERFORMANCE OF AN AUTOMATED SHIP DETECTION SYSTEM FOR SAR IMAGES

K Eldhuset

Norwegian Defence Research Establishment
(NDRE), Kjeller, Norway

ABSTRACT

A demonstration system for detection of ships and ship wakes in SAR images has been developed on an Apollo DN 10000 workstation. The main principles of the system is to distinguish land from sea by using accurate pixel location algorithms and digital terrain models. Then an adaptive point detector is used to extract potential ship targets. The system searches around each potential ship target for the wake. In addition, a homogeneity test and a detailed wake analysis are performed to reduce false alarms if potential ships and wakes actually are oceanic phenomena such as fronts, eddies, or internal waves.

1. INTRODUCTION

A prototype automatic system has been developed at Norwegian Defence Research Establishment for detection and analysis of ships and ship wakes in Seasat SAR images. This is an experimental system and shall be further developed during the ERS-1 mission. An ESA contract study was performed at NDRE in 1986 (Aksnes, 1988) on developing detection methods for ships and ship wakes in Seasat SAR images. Based on some of these techniques and some recently developed methods, a demonstration system for automatic detection has been implemented on an Apollo DN 10000 workstation with only one of the CPU's. All fiords, islands, and rocks in Norwegian waters complicate the automation of a ship detection system. Figure 1 shows a typical coastal region from Western Norway. To overcome these problems a digital terrain model has been included to locate open waters where a ship detector may operate, followed by a wake detector. The ship and wake detectors may work satisfactorily under normal or slowly varying sea conditions, but strong scatterers in the sea such as eddies and fronts may be interpreted as ships and wakes. To avoid false alarms in such regions, a homogeneity test has been developed to investigate the region in front of all potential ships and wakes. Also the wake scan curve (see Chapter 3) generated around the potental ship is analysed and has to be sufficiently smooth. If the region near the ship is inhomogeneous or the scan curve is too steep where a potential wake was detected, the ship and wake are rejected.

Since the sea and land detection problem has been treated previously by the author (Eldhuset, 1987, 1988), we shall concentrate on the ship and wake detection. A ship detector using a Wiener filter and a highpass filter has also been described by Eldhuset (1987). A much faster algorithm has now been developed and is treated in Chapter 2. An important part of this paper will be the homogeneity testing (Chapter 4), which is a necessary tool to reduce false alarms.

2. SHIP DETECTION

In Figure 2 is shown a typical situation for a ship detector. We assume that land and islands have been detected by our system (shaded in Figure 2). The principle of the ship detector is to estimate the mean, μ, in the frame region B (b=20) around the small window denoted A (a=10):

$$\mu = (\Sigma_i \Sigma_j X(i,j))/K \ , \qquad (1)$$

where K is the number of pixels in region B. X(i,j) is the intensity or the amplitude with coordinates (i,j). A new image is generated by the formula:

$$C(i,j)=(X(i,j)+X(i,j+1)+X(i+1,j)+X(i+1,j+1)-4\mu), \qquad (2)$$

for all (i,j) inside window A. Window A and frame B are moved with steps equal to the dimension of window A between each computation of Eq. 1. For an N-look SAR image the standard deviation, σ, is related to the mean, μ, by (Ulaby,1976 and Andås,1988):

$$\sigma=\mu\cdot((4/\pi-1)/N)^{1/2} \quad \text{(N-look amplitude image)} \ , \qquad (3)$$

$$\sigma=\mu/\sqrt{N} \quad \text{(N-look intensity image)} \ . \qquad (4)$$

Eq. 3 is valid if N amplitude looks are added, while Eq. 4 is valid if N intensity looks are added. When Eqs. 3 and 4 are used the effective number of looks, N', has to be estimated. For the NDRE processor N'=3.6 and for the DFVLR processor N'=3.3. If C(i,j) in Eq. 2 is greater than a threshold given by $T=q\cdot\sigma$, the pixel (i,j) is accepted as a ship pixel. The number q is typically 4.5. Then all such pixels in a given image (512x512) are grouped together. If an accepted pixel is too far from another pixel it belongs to another ship. The advantage with this filter is that it is rapid and the frame region (B) may be defined only once. The filter is adaptive, because the threshold, T, is dependent on the mean in frame B. Hence, a ship with a given backscatter is more easily detected in dark sea than in bright sea. In principle, one should move the frame when (i,j) is

changed in Eq. 2. Instead, the frame is moved with steps of size a. This of course affects the operation of the filter, but the computing time is reduced substantially. Experience shows that if the frame region is inhomogeneous, the number of pixels above the threshold, T, may depend on the starting point of the filter. This is typically the case at fronts and eddies, but if a ship is present in the window A, the number of pixels above the threshold is not dependent on the starting point of the filter when frame B is approximately homogeneous.

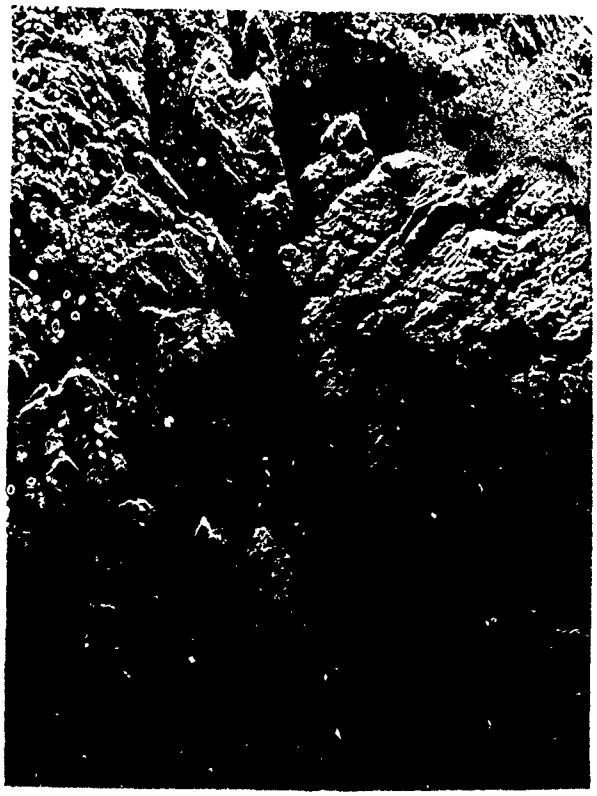


Figure 1. Seasat SAR image from Sognefjorden in Norway processed at NDRE (orbit 1502)

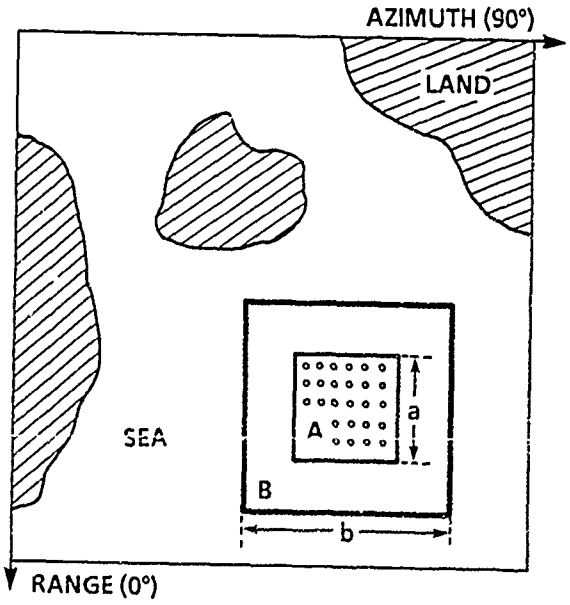


Figure 2. Principle of the ship detector

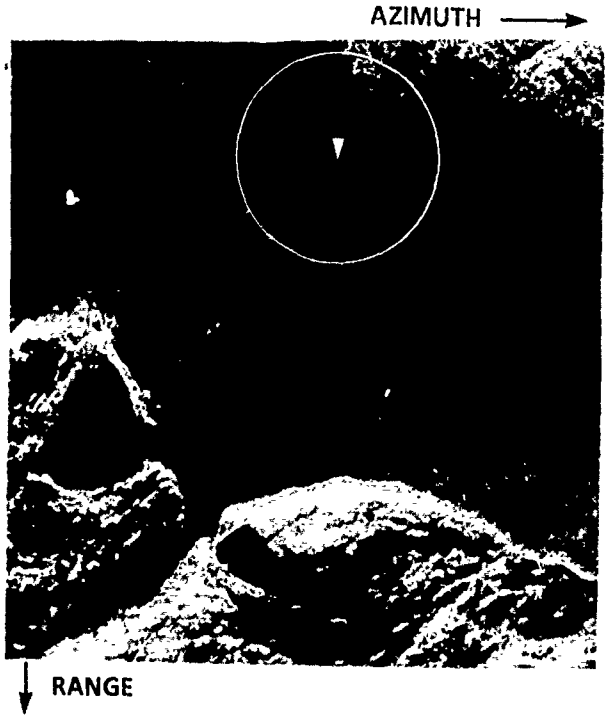


Figure 3. Detection of a ship and wake in coastal regions

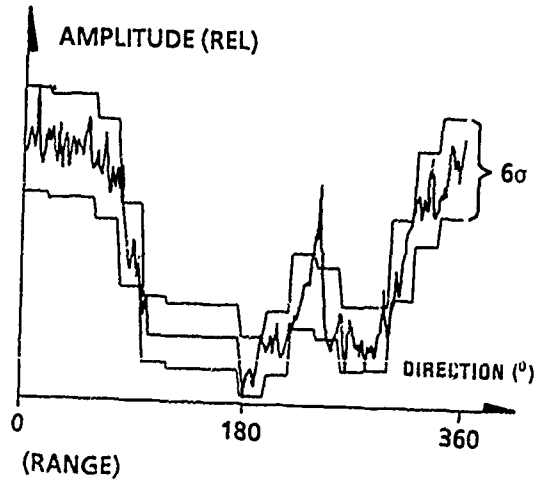


Figure 4. Scan curve showing the wake in Figure 3

3. WAKE DETECTION

In Figure 3 is shown an image from Sognefjorden in Western Norway with a mixture of land and islands which are effectively classified by means of the digital terrain model and the pixel location algorithm. Note the two small islands in the middle of the left part of the image which are effectively classified as land. The overlayed map is not shown here. Before wake detection is performed, all detected ships are replaced by homogeneous sea pixels. In Figure 3 we see a small ship with a bright wake. Figure 4 shows the scan curve which is generated by averaging pixels along lines with a fixed angular distance starting in a given center. This center is moved in azimuth direction on both sides of the ship. The curve with the most significant wake features is chosen, hence the displacement of the ship is also estimated. The real position of the detected ship in this image is indicated by the arrow in Figure 3, and the scan curve is generated inside the circle shown. In Figure 4 we have used step functions to indicate a given threshold (3σ or more) on both sides of the curve. We observe that the bright wake appears as a peak in the curve with very high significance at 240 degrees measured from range direction. The method of least squares combined with Chebyshev polynomials has turned out to be more flexible in many cases to estimate the threshold above and below the scan curve. Eq. 3 or 4 is used to estimate the threshold with N replaced by $N' \cdot L$, where L is the number of independent pixels averaged along each line inside the circle in Figure 3. The straight line in the scan curve from about 90 to 180 degrees in Figure 4 indicates the direction towards land, where we should not accept wakes because land pixels are inside the search radius. The wake indicates the direction of the moving ship, but the system also computes the direction of the ship itself by minimizing the moment of inertia of the ship. Hence, the two extracted directions can be compared.

4. HOMOGENEITY TESTING

The ship and wake detector may be confused in regions with fronts, eddies, or internal waves. Figure 5 shows a front with a very abrupt change in the backscatter. Figure 6 shows a very strong eddy in a quite dark sea. In such cases false ships and wakes are detected. The scan curve around the ship may have a strong drop in the transition between bright and dark sea. If this drop is too large the ship and wake are rejected. If the drop is not sufficiently significant a homogeneity test developed at NDRE often reveals that fronts or eddies are present. Experiments show that a combination of the scan curve test and the homogeneity test reveals false ships and wakes very effectively. In the next section we shall sketch the principle of our homogeneity test.

4.1 Moment estimation for homogeneity testing

The moment estimator for the n'th central moment is:

$$\underline{\sigma}^n = \sum_{j=1}^{K} (X_j - \underline{\mu})^n / K \quad , \tag{5}$$

$$\underline{\mu} = \left(\sum_{j=1}^{K} X_j \right) / K \quad \text{and} \quad X_j \text{ is the pixel value of}$$

pixel number j in a defined region. The expectation and variance of the estimator in Eq. 5 is very complicated if the mean, $\underline{\mu}$, is considered as an estimator. We have therefore assumed that the mean is a given constant which becomes a good approximation when the number of pixels K increases. The expectation and variance of the estimator in Eq. 5 may be expressed (Andås,1988):

$$E(\underline{\sigma}^n) = \sum_{j=0}^{n} \binom{n}{j} (-\mu)^j E(X^{n-j}) \quad , \tag{6}$$

$$\mathrm{var}(\underline{\sigma}^n) = \{E((X-\mu)^{2n}) - (E((X-\mu)^n))^2\}/K \quad , \tag{7}$$

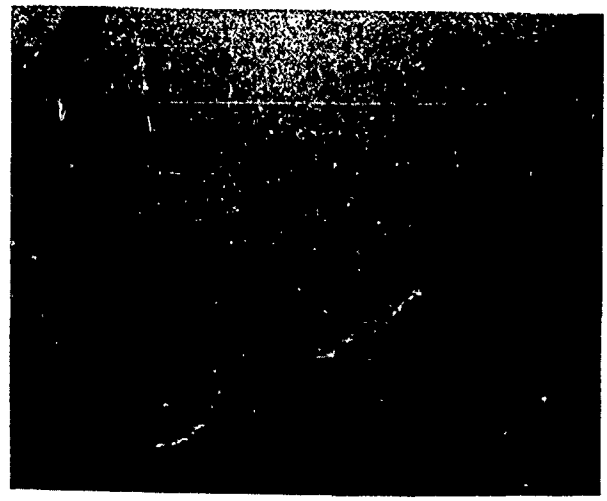


Figure 5. Ocean front and a ship from Corsica, processed at DFVLR, orbit 762



Figure 6. Eddies and a ship in Oslofjorden, Norway, processed at NDRE, orbit 1473

$$E((X-\mu)^L) = \sum_{j=0}^{L} \left[\begin{matrix} L \\ j \end{matrix} \right] (-\mu)^j \, E(X^{L-j}) \quad , \; L=2n \text{ or } L=n. \qquad (8)$$

The square bracket is the Binomial coefficient. The formulas in Eqs. 6 and 7 are used to define the confidence interval for the estimator in Eq. 5. If the estimated central moment for a given region lies in this interval the region is said to be homogeneous. From Eq. 8 we see that we need an expression for the L'th moment of a distribution. It can be shown (Andås, 1988) that that the L'th moment for a gamma distribution corresponding to an N-look intensity image (see Eq. 4) is given by:

$$E_N(X^L) = \beta^L \prod_{j=1}^{L} (N+j-1) \quad , \; \beta=\mu/N \; . \qquad (9)$$

The L'th moments for the distribution in amplitude images have also been estimated which is rather cumbersome (Andås,1988). Only moments for L=1,2,3,4, 5,6 have been deduced. These expressions are finite although the corresponding distribution has to be expressed by infinite sums. Estimation of the second and third central moment (n=2,3) is usually sufficient for homogeneity testing in SAR images.

In Figure 5 some false ships and wakes are detected on the front, but application of the tests above rejects all of them effectively. The ship with the bright wake is detected with very high significance. In Figure 6 about 20 false ships and wakes are detected in the eddy. The tests reject at least 90 % of them. The processing time is of course increased if the phenomena mentioned above arise. The small ship between land and the eddy in Figure 6 is well detected. Strong internal waves are often detected by the homogeneity test alone. Figure 7 shows some strong internal waves. The adaptive property of the ship detector results in only two potential ships being detected in this image. The homogeneity test tells us that the regions in the vicinity of the ships are not sufficiently homogeneous, hence the two ships are rejected.



Figure 7. Internal waves in the English Channel, processed at NDRE, orbit 762

5. PERFORMANCE OF THE SHIP DETECTION SYSTEM

The philosophy behind the system is that the computing time is minimized by searching for point targets first, because point detection may be performed rapidly and effectively. The filter constructed in our system is in principle very simple, but is based on statistical theory, and its performance is comparable to more complicated and time consuming filters. Point target detection is completed in about 6 seconds on a 512x512 image with one CPU on DN 10000 (18 MIPS). Under normal sea conditions as in Figure 3, usually no false ship candidates are detected. Hence, wake detection is not superfluous. The wake detector needs between 1 and 2 seconds to locate one wake around a point target. The time is of course dependent on search radius, angular distance between the scan lines etc. The homogeneity test is very rapid because of the small regions which are tested. If there is land in the SAR image the land and ship detection is faster than ship detection over open sea, because the land detection is very fast, and the ship detector need not work on land.

6. CONCLUSIONS AND FUTURE PLANS

Our experience with the automated ship detection system is very encouraging, both with regard to time consumption and false alarm rate. The current system detects most ships and wakes, and the additional computations performed by the system appear to reveal the greater part of false ships and wakes which actually are oceanic phenomena. During the ERS-1 mission we intend to investigate to what extent an automated system may work without the influence of an operator. The ship detection process is well suited for parallell computing, and the DN 10000 may work with 4 CPU's. A digital terrain model may not reveal rocks near the sea level, but we plan to include also digital models of the sea depth to complement the information which digital terrain models provide. In shallow water there are of course no ships and internal waves may arise, so sea depth models may be a valuable information to the system. We have not yet completed the performance analysis concerning false alarms with the current system, but we intend to work out a survey of true and false alarms in as many Seasat scenes as possible.

7. REFERENCES

1. Aksnes et al, "SAR Detection of Ships and Ship Wakes", ESA Contract Study Report, Vol 2: Main Volume, NDRE 1988.

2. Andås H. and Eldhuset K., "Estimation of central moments in SAR images for homogeneity testing" (in Norwegian), FFI/NOTAT-88/9003 , NDRE, Kjeller, Norway, 1988.

3. Eldhuset K.,"Digital Image Processing for Maritime Surveillance with Spaceborne Synthetic Aperture Radar", FFI/RAPPORT-87/9001. Norwegian Defence Research Establishment, Kjeller, Norway, 1987.

4. Eldhuset K., "Automatic Ship and Ship Wake Detection in Spaceborne SAR Images from Coastal Regions", Proceedings of IGARSS'88, (ESA SP-284).

5. Ulaby et al, "Textural Information in SAR Images", IEEE Tr on Geoscience and Remote Sensing, vol GE-24, No 2, March 1976.

THE EFFECT OF SHIP'S SCREWS ON THE SHIP WAKE AND ITS IMPLICATIONS FOR THE RADAR IMAGE OF THE WAKE

Eric H. Buller and James K.E. Tunaley

London Research and Development
755 Queens Ave., London, Ontario, Canada

ABSTRACT

The radar image astern of a ship is dependent on the wake produced by the ship's linear motion. Contributions to the turbulent wake by the ship propeller have been studied so that the effects on radar backscattering can be estimated. Both linear momentum and angular momentum wake components may be produced and the diffusion of each will affect the width and intensity of the radar image. Linear momentum tends to diffuse so that the wake radius varies as the cube root of distance behind the ship while it is shown that angular momentum diffuses as the fourth root of distance behind the ship. It turns out that for surface ships the radar image of a wake far astern will be dominated by the linear momentum.

KEY WORDS: RADAR, WAKE, PROPELLER

I. INTRODUCTION

Radar images of ship wakes have been described by a number of authors (Hammond et al, 1985, Lyden et al, 1985, Case et al, 1985 and Lyden et al 1985). In an effort to obtain a more thorough understanding of the radar image astern of a ship, the wake produced by the propeller has been studied. The propeller will introduce various flows astern of the ship which will include both a steady component of fluid flow behind the vessel and a rotational movement. The shear layer created at the boundary of the wake may be expected to cause a turbulent diffusion which broadens the wake. The wake flows are capable of modifying the local radar cross-section of the ocean surface as described for the linear component by Tunaley et al (1987), in which the interaction of surface waves with the wake is simulated. The magnitudes of the flow velocities are dependent on the efficiency of the screws.

Propeller design has been widely considered (Clancy, 1975, Comstock, 1967 and O'Brien, 1962). In a simple theory, the propeller is treated as an actuator disk and the energy and linear and angular momenta produced in its neighbourhood are considered. However even this simplification is not sufficient to understand the operation completely. Aerodynamic theory must be used and the propeller blades viewed as aerofoils. For example, thrust is provided by the lift force acting on the blades. In this study, the principles involved in screw propellers have been used to arrive at a set of non-linear equations that may be solved numerically to estimate the magnitude of the wake momenta.

Prandtl's mixing theory may be used to study the diffusion of the momenta (both linear and angular) produced by the propeller. It will be shown that angular momentum tends to diffuse much more slowly than linear momentum but the associated velocity field decreases more rapidly: as a result it is not expected to be significant except in cases where little linear momentum is produced.

II. ACTUATOR DISK THEORY

To treat the linear motion, we consider the propeller as a disk of area A_D and radius r_D, immersed in a fluid of density, ρ as shown in fig. 1. From the point of view of an observer moving with a coordinate frame attached to the ship, if the velocity of the fluid well in front of the propeller is V_A, then, in order for some thrust to be obtained, the fluid must acquire some linear momentum in the form of an increase in its velocity. The velocity well behind the propeller will rise to $V_A(1+b)$ and at the disk to $V_A(1+a)$, say. Well behind the propeller but before turbulence sets in, the affected fluid has a smaller radius, r_w. The mass of fluid that flows past the propeller in a time t experiences a change in linear momentum. During the same time, the mass flowing past a disk in the wake must equal the amount of mass flowing past the propeller. From the conservation of mass and the fact that the thrust is equal to the change in momentum per unit time, the thrust can be determined. This can be compared to the result obtained by applying Bernoulli's equation. In order for the results to agree, the inflow factor, a must be equal to half of the outflow factor, b.

In order to see the effects of angular momentum, we assume that the fluid has no initial rotational velocity. The propeller is rotating with angular velocity, ω, the angular velocity of the fluid at the propeller is a' and far behind the propeller it is b'. In a time t, the mass of fluid flowing past a disk in the wake has a moment of inertia and a related increase in angular momentum. The torque is simply the rate of change of angular momentum and this can be combined with the conservation of mass condition. In the same time, the propeller delivers an amount of energy, some of which the fluid obtains as kinetic energy of rotation and some as translational kinetic energy. The useful energy obtained corresponds to the thrust. Assuming that the production of heat is small, the conservation of energy combined with the previous results gives a relation that is usually combined with Kelvin's

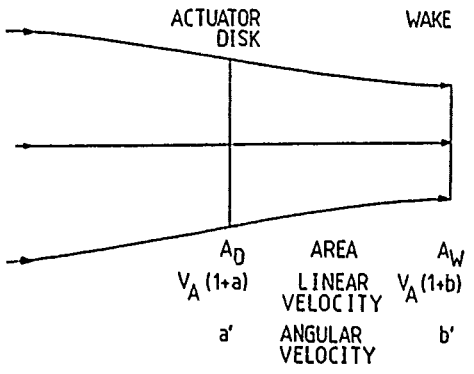


Figure 1. The screw propeller as an actuator disk.

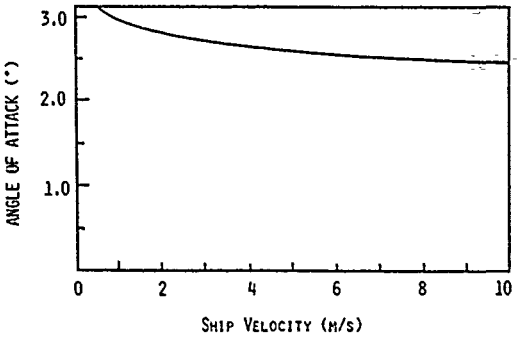


Figure 2. Inclination angle as a function of the ship velocity.

circulation theorem to show that the rotational inflow velocity, b' is twice the rotational inflow velocity, a'.

For an aerofoil of area, A immersed in fluid of density, ρ and flowing past at a velocity, V there will be a lift force, L perpendicular to the fluid flow direction and a drag force, D parallel to the fluid flow. The lift and drag are directly proportional to the lift and drag coefficients, C_L and C_D, respectively.

In most cases, C_D is nearly constant when the direction of fluid flow is only slightly inclined to the aerofoil. The angle between the fluid flow direction and the major axis of the aerofoil is called the inclination angle and is usually denoted as α. For an aerofoil of elliptical cross section, the lift coefficient is proportional to $\sin\alpha$ which is approximately proportional to α for small angles (Comstock, 1967).

The angle of inclination, α is equal to the difference between the pitch angle, ϕ, and the hydrodynamic angle, β. The pitch angle is the angle the propeller blade cross section makes with the direction of advance through the fluid. The hydrodynamic angle is the angle between the lift and the thrust. The thrust is obtained by summing the components of the lift and drag in the direction of advance of the propeller through the fluid. The torque is obtained in a similar way using the "effective radius" of the propeller (about $0.7r_p$).

The area of contact between one blade and the fluid is multiplied by the number of blades to obtain the total contact area. The relative velocity of fluid flowing past the blades is used to calculate the lift and drag. These relations can be used in the thrust and torque relations along with the momentum relations to solve for the linear inflow factor and the rotational velocity of the fluid at the propeller.

The final information necessary to form a complete set of equations is the value of C_D for small inclination angles. A value of 0.0085 has been taken from Comstock (1967) on page 425.

III. NUMERICAL SOLUTIONS

The secant method of finding roots (Press et al, 1986) was used to solve the equations of motion described in the previous section. The inner loop of the procedure found the required propeller rotation to deliver the necessary thrust to propel a ship at a constant velocity.

The outer loop involved a similar secant method. The thrust required to propel a ship of a certain beam and length at constant velocity is equal to the drag on the hull. A simple estimation of the hull drag for a given ship dimensions and velocity has been described (Tunaley et al, 1987). For a certain velocity, two initial estimates of the propeller rotational velocity were made. Each estimate was used to determine the hydrodynamic angle and thrust. The thrust so obtained was compared to the required thrust and a new propeller rotational velocity interpolated. A solution was accepted when the thrust obtained for a certain propeller rotation velocity was equal to the hull drag to within 0.1%.

An Example

In this example, a single propeller is driving a ship of length 150 m and beam 20 m. The propeller radius is one metre and the expanded area ratio, chosen to represent a four blade propeller, has a value of one. The pitch is 45 degrees. From these ship dimensions, the thrust required to maintain a constant velocity was calculated using the method of Tunaley et al (1987).

The rotational frequency of the propeller increases nearly linearly with ship velocity. An interesting feature of the angle of attack curve (figure 2) is that it actually decreases as the velocity increases. This means that the propeller loading also decreases with increasing velocity and the efficiency of the propeller increases slightly (figure 3). The efficiency is defined as the ratio of useful work to work delivered by the engine.

The linear and rotational outflow factors, as shown in figures 4 and 5 respectively, decrease with increasing velocity in a similar fashion to the angle of attack. Both outflow factors show a significant production of fluid motion in comparison with the propeller motion. The linear velocity in the wake is nearly 60% greater than the propeller's linear velocity of advance. In other words, the linear velocity of the wake with respect to the ocean is 60% of the velocity of the ship but in the reverse direction. The angular velocity imparted to the water is approximately 40% of the propeller's angular velocity.

IV. DIFFUSION OF THE WAKE

Though the turbulent wake from a ship moving at constant speed in a straight line maintains its shape in a frame of reference attached to the ship, in a

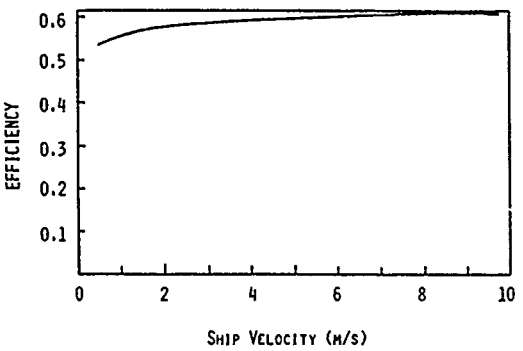


Figure 3. Efficiency of the propeller as a function of the ship velocity.

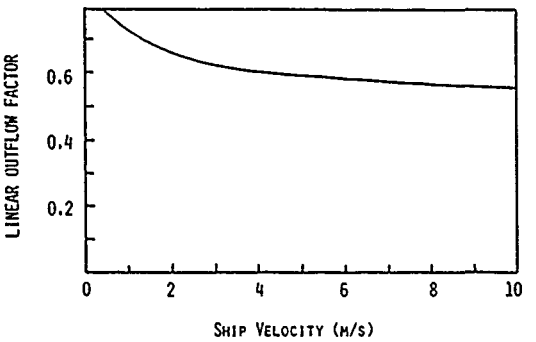


Figure 4. Linear outflow factor as a function of the ship velocity.

frame of reference stationary with respect to the ocean, it broadens with time. This spreading occurs because of turbulent diffusion which arises from shear flows within the wake. A shear flow may result from a mean flow parallel to the wake which itself is a result of a net production of linear momentum by the ship. Shear flow can also occur if the screws produce a net angular momentum. With a single screw, there will be a persistent rotational motion imparted to the water. Thus a gradient in the velocity field must occur where the wake meets the undisturbed ocean. The theory of wakes has been discussed by Schlichting (1979) and for simplicity it is assumed that the wake is circular in cross-section. However the effect of the presence of the free surface could be modelled by introducing an image which will affect the shape of the wake.

This study is limited to the "self similar" region well behind the ship where the dynamics are constant except for scale factors.

The diffusion in the transverse direction is related to the mean velocity in the transverse direction, v'. If the radius of the wake is R, then, according to Schlichting, with Prandtl's mixing length hypothesis, the velocity in the wake decreases as R^{-2} and the width increases as the cube root of the distance behind the ship. For a wake with no linear momentum, $_{m}$ must be related to the value in the propeller race. The principle of the conservation of angular momentum can be applied to a disk of fluid at a fixed position in the wake and rotating at a constant angular velocity, $_{m}$; the angular velocity is inversely proportional to the fourth power of the width of the wake. In the rotational form of Prandtl's mixing length theory, the width of the wake increases as the fourth root of the distance behind the ship. These results resemble those quoted by Case et al (1985), except that for a zero momentum wake we do not find a velocity defect along the wake axis. Also the present results are expressed in terms which reflect the propeller efficiency whilst those by Case et al (1985) do not.

In general a surface ship will produce linear momentum in the water. It will also produce angular momentum if it does not have contra-rotating screws. It is easy to show from the data in the previous section that the diffusion is dominated typically by the linear momentum of the wake. From the example discussed earlier, at a ship velocity of 5 m/s the propeller rotates at about 2 rev/s and the angular outflow factor is 0.4; this gives an angular velocity downstream of the propeller of about 2.5 rad/s. The linear outflow factor is about 0.6 and it can be deduced that the radius of the propeller wake before the onset of turbulence is about 0.85 m. Assuming a drag coefficient of 0.1, the distance from the ship at which the linear momentum wake diffusion would start to dominate that of angular momentum may be found. It turns out that this is much less than 1 m.

V. THE RADAR IMAGE

A radar image of the ship wake may be produced because the Bragg wavelets, responsible for the radar scattering, are modified by surface flows associated with the passage of the ship. The flows perturb the propagation of wind generated wavelets so that their directions are altered: in extreme cases, refraction can be so severe as to exclude Bragg waves from the wake region and this causes the wake to appear as a dark streak on a radar image. As well as being refracted, the amplitudes of the wavelets will change and this may be handled using the principle of the conservation of wave action as described for example by Tunaley and Mitchell (1987). For radar frequencies at L or X bands and under favourable sea conditions and wake and radar geometries, images may be produced with mean flows of a few cm/s.

In the case of a surface ship with a wake having both angular and linear momenta, the mean flow at the surface will have a component of velocity usually in the opposite direction to the ship velocity. It will also have a transverse component because of the rotation. However, because the transverse component falls off as R^{-4}, whilst the linear velocity only falls off as R^{-2}, the transverse component is quite negligible just a few metres behind the ship (see the example in section IV). Therefore, unless the wind drag and Kelvin wake drag fortuitously cancel, which is somewhat unlikely, the propeller wake is probably small.

It is worth noting that the circulation (defined as the line integral of the velocity around a closed path) around a circle concentric with the propeller wake rapidly falls to small values for radii greater than that of the wake. Therefore an image of the propeller wake placed above the water surface will not produce any significant self induced motion until the wake intersects the surface. When this occurs, there may be a general broadening. The effect is worth further study.

VI. CONCLUSIONS

The equations of motion of a propeller have been used to examine the wake produced by the propeller. Both linear momentum and angular momentum are produced and the spreading of each, on its own, is quite different. Prandtl's mixing length theory suggests that the width of an angular momentum wake

increases as the fourth root of the distance behind the propeller while the width of a linear momentum wake increases as the cube root of the distance behind the propeller. As a result of the dependence of the velocity fields on the distance behind the ship, angular momentum wakes could be most important when the net linear momentum produced by the propeller is offset by the linear momentum produced by the ship hull. In this case a radar image could not be produced directly by the turbulent wake.

VII. ACKNOWLEDGEMENT

This work has been conducted for the Canadian Department of National Defence in connection with a contract between London Research and Development and Supply and Services Canada. The authors wish to thank K.H. Wu of the Communications Research Centre, Shirley Bay, Ottawa, for acting as the scientific authority for this project and for proposing, facilitating and encouraging the work.

REFERENCES

Case, K., R. Dashen, W. Munk, J. Vesecky, K. Watson and F. Zachariasen. "SEASAT Report", Mitre Corporation Report JSR-83-203, January, 1985.

Clancy, L.J. Aerodynamics. Pitman Publishing Limited, 1975

Comstock, J.P. (editor). "Principles of Naval Architecture". The Society of Naval Architects and Marine Engineers, 1967.

Hammond, R.R., R.R. Buntzen and E.E. Floren. "Using Ship Wake Patterns to Evaluate SAR Ocean Wave Imaging Mechanisms", Joint U.S. - Canadian Ocean Wave Investigation Project, Naval Ocean Systems Center (San Diego, CA) Technical Report 978, February, 1985.

Lyden, J.D., D.R. Lyzenga, R.A. Shuchman and E.S. Kasischke. "Analysis of Narrow Ship Wakes in Georgia Strait SAR Data", Environmental Research Institute of Michigan Report, January, 1985.

Lyden, J.D., D.R. Lyzenga and R.A. Shuchman. "SAR Detection of Ship Generated Turbulent and Vortex Wakes", Environmental Research Institute of Michigan Report, Report RR-86-112, 1985.

O'Brien, T.P. The Design of Marine Screw Propellers. Hutchinson & Company Limited, 1962.

Paterson, A.R. A First Course in Fluid Dynamics. Cambridge University Press, 1983.

Press, W.H., B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling. Numerical Recipes: The Art of Scientific Computing. Cambridge University Press, 1986.

Schlichting, H. Boundary-Layer Theory. McGraw-Hill Book Company, 1979.

Tunaley, J.K.E., P.A. Rochefort and J.B.A. Mitchell. "The Modelling and Classification of SAR and ISAR Ship Images Including Their Wakes: Part 2." London Research and Development Report to Defence Research Establishment Ottawa, March, 1986.

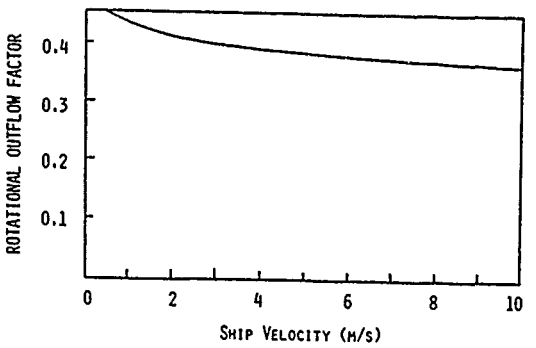


Figure 5. Rotational outflow factor as a function of the ship velocity.

SAR IMAGING OF SHIPS AND SHIP WAKES DURING NORCSEX'88

Terje Wahl and Åge Skøelv
Norwegian Defence Research Establishment
Box 25, N-2007 Kjeller, Norway
Phone: 47 6 80 70 00
Fax: 47 6 80 72 12
Telex: 21682 elect n

ABSTRACT:

The NORCSEX'88 experiment was carried out at Haltenbanken (64°N, 9°E) outside the coast of Norway in March 1988. The Canadian CV580 provided C-band SAR data, and SAR imaging of ships and ship wakes was one of the study objectives. The SAR data show much greater variations in radar backscatter from the ocean surface than observed in Seasat images from the same areas. Some interesting ship wakes have been analysed. Despite the high resolution of the SAR, it turned out to be difficult to see wakes behind slow-moving ships. The results support the view that prime search target in automatic analysis of ERS-1 SAR images must be the ship itself, not the wake.

Key Words: SAR, Ship Wakes, ERS-1, Backscatter

1. INTRODUCTION

For many years, there has been a strong Norwegian interest in the possibility of using satellite-based SAR for monitoring human activities in the large Norwegian economic ocean zones. In 1986, a Norwegian proposal termed "Ship Traffic Monitoring Using the ERS-1 SAR" was submitted to ESA in response to the ERS-1 Announcement of Opportunity. The proposal has been officially accepted by ESA, and quite some activity is currently going on in Norway to meet this challenge. In order to gain further experience before the launch of ERS-1, ship detection was included as one of the topics in the 1988 NORCSEX campaign at Haltenbanken. The Canadian CV580 aircraft here offered real time C-band SAR capability in an area on the continental shelf that may be typical for what ERS-1 will find along the Norwegian coast during winter and spring time.

The main goals of the ship and wake experiment were:

- To study SAR imaging of "typical" fishing vessels and their wakes.
- To study background backscatter variations in C-band and their effects on the detectability of ships and wakes.
- To obtain some well documented wake scenes for later comparison with model simulations.

Several ships were imaged by the CV580 SAR during the NORCSEX'88 campaign. We have focussed mainly on "R/V Håkon Mosby", a research vessel which was covered by the SAR more than 10 times. Being a former fishing vessel, this is a very relevant target. It also provides excellent weather, sea state, and ship motion data.

2. RADAR BACKSCATTER DURING NORCSEX'88

Seasat SAR images from Norwegian coastal waters generally show very homogeneous backscatter from the ocean surface. Apart from in the fiords, ship wakes, internal waves and current shears appear against a very homogeneous background, making detection easy and automatic analysis feasible. However, Seasat scenes are only available from the period August-October 1978. This is a part of the year with small temperature differences between the Atlantic and the Coastal water masses, stable atmospheric conditions, and generally low wave heights. Thus it is dangerous to base ship detection and wake pattern recognition algorithms on Seasat images only. As could be expected, the March 1988 CV580 SAR data from the NORCSEX campaign at Haltenbanken showed a much larger dynamic range within each SAR image (Johannessen & Pettersson, 1988). Concerning ship detection, this is a clear indication that we must expect much more complicated ERS-1 SAR scenes than is offered by the present Seasat training set.

One of the most interesting backscatter situations occurred on March 14th, at rather low wind speed. On that day large sea areas showed a granular SAR texture, with "wind patches" sized 1-2 km^2 spread around everywhere, making 50-70 % of the sea bright and the rest dark. Swell could be seen inside the bright patches only.

3. SOME WAKE EXAMPLES

On March 21st 1988, the CV580 SAR crossed R/V Håkon Mosby (now going at 6 m/s) five times with about 25 minutes intervals, the ship making a 90^0 turn after the third passage. The wind was low and decaying, coming down below 3 m/s. Significant waveheight measured by a nearby buoy (OCEANOR, 1988) was just below 3 meters, with wave period at spectral peak about 9 seconds. The total spectrum was, however, rather complicated, with both south westerly and northerly swells in addition to a low south easterly wind sea. For the two wake scenes to be discussed here (#1 and #5), the aircraft was flying on a heading of 305^0. Thus the SAR mainly imaged the south westerly swell, with almost azimuthally aligned wave crests.

Scene #1:

The ship track was perpedicular to the flight direction, with the ship located almost at nadir. A dark turbulent wake is visible on the SAR image, extending for about 4 km in ground range. Backscatter values inside the wake typically are 1.0 - 1.5 dB below the level of the surrounding sea. On the upwind side, the dark turbulent wake is limited by a narrow, bright edge, resulting from steepening or breaking of the wind sea. Backscatter values on the edge typically are 0.5 dB above the level of the surrounding sea. No Kelvin-like wake features can be seen. It is remarkable, however, that the dark turbulent wake widens out significantly on the downwind side, from 70 m near nadir to 200 m before the wake becomes invisible. This could, of course, be due to a widening of the surface current pattern behind the ship. However, we rather tend to believe that this phenomenon has to do with the variation in incidence angle along the wake, which gives Bragg resonant wavelengths ranging from 20 cm (near nadir) to 5 cm.

Scene #5:

In this case the heading of the ship was parallell to that of the aircraft, and the wind was very weak. No turbulent wake is visible, and the only significant wake feature present is the one Kelvin arm facing the aircraft. The length is about 2 km, with backscatter values 1.5 dB (near the ship) and 1.1 dB (overall value) above the background level of the sea. The absence of a dark turbulent wake in this case would usually be explained by too low wind speed. However, there clearly is enough surface roughness present in the ambient sea for imaging of swell.

Both wakes, however, are rather faint compared to a giant soliton-like feature present in scene #1. This feature is ca 12 km long, it has a radius of curvature of ca 13 km, and a "wavelength" of 400 m. The crest-to-trough backscatter variation is 2.8 dB.

4. DISCUSSION

Most wake observations in SAR images around the world are classified as dark turbulent wakes (Aksnes et al, 1988). The imaging of such wakes is a very complex process, involving ship generated currents, turbulence, foam, upwelling of cold water, monolayers, and wind and wave interactions (see Peltzer et al, 1987, and references therein). Our main intention with the wake part of NORCSEX'88 was not to produce high-quality wake measurements for refinement of state-of-the-art models, but rather to get a broader view of the challenges that ERS-1 will face in 1991. One main result so far (which confirmed model predictions) is the problem of imaging wakes behind vessels going at low speed (which typically is the case when fishing). Thus the prime target when searching for fishing vessels in ERS-1 data should be the ship itself, not its wake. Of course, the wake may be an excellent source of information when monitoring merchant ship traffic.

A major difference between aircraft and satellite SAR is the wide range of incidence angles present in aircraft SAR images. Of course, ships and oil rigs stand out more clearly at larger incidence angles in the NORCSEX'88 SAR data. The incidence angle chosen for ERS-1 is far from ideal for ship detection purposes.

Another very important difference between aircraft and satellite SAR is the effect of velocity bunching. The larger R/V ratio of satellites makes them much more sensitive to wave motion. During NORCSEX'88 the wind speed often dropped to a favourable level for wake imaging, but the significant waveheight seldom was below 1.5-2 m due to swell. For ships in azimuthal motion, velocity bunching will not matter much, as an azimuthally aligned turbulent wake is invariant under this mechanism. But turbulent wakes behind range-going ships may become severely degraded in ERS-1 SAR images because of scatterers "thrown" into the wake from the ambient sea through the velocity bunching mechanism.

REFERENCES
---

1. J A Johannessen, L H Pettersson: NORCSEX'88 Experiment Report. Technical Report no 15, Nansen Remote Sensing Center, 1988.

2. OCEANOR: Wave Measurements from Moored Directional Wave Buoys During NORCSEX'88. ODAP Report no 108, 1988.

3. K Aksnes, K Eldhuset, T Wahl, et al: SAR Detection of Ships and Ship Wakes. ESA Report, Vol 2. NDRE 1988.

4. R D Peltzer, W D Garrett, P M Smith: A remote sensing study of a surface ship wake. Int J Remote Sensing, 1987, Vol 8, No 5.

Acknowledgements

Thanks to the other NORCSEX'88 participants for valuable assistance.

This work was partly funded by the Norwegian Space Centre.

ACTIVE MICROWAVE CLASSIFICATION OF SEA ICE

Robert G. Onstott
Environmental Research Institute of Michigan
Advanced Concepts Division
Radar Science Laboratory
P.O. Box 8618
Ann Arbor, MI 48107

ABSTRACT

Radar backscatter studies of Arctic sea ice have been carried out over a number of years now with the intent to acquire physical property information through by the examination of microwave signatures. The breadth of these studies continues to expand. As an example, measurements are now conducted at frequencies from 500 MHz to about 100 GHz. One of the scientific goals of this work has been to develop an improved understanding of the scattering processes at play. A second, equally important, goal has been to apply the knowledge gained in examining the backscatter response of ice and snow made in conjunction with detailed scene characterizations, the insight gained through theoretical modeling and parametric study, and the data entered into the radar signature library to develop procedures to convert microwave signal information, available in the very near future, into valuable data products, ultimately providing the opportunity for us to better understand our environment. The purpose of this paper is to provide a discussion of what has been learned through the many efforts associated with the near-surface scatterometer measurement programs and how the knowledge gained is assisting in the development of future sea ice type satellite algorithms.

INTRODUCTION

One of the most important pieces of geophysical information that must be obtained from satellite observation of the frozen ocean is ice type. A wide array of categories have been defined by the World Meteorological Organization and are related to thickness, age or formation characteristic. Basic categories include new (0 to 10 cm thick), young (10 to 30 cm thick), first year (> 30 cm thick), second year (survived one melt season) and multiyear (survived multiple melt seasons). The ability of determining ice type using space-based sensors is dependent on spatial resolution and the region under observation. In the marginal ice zone (MIZ) the mixture of floes of various ice types have diameters which are small, ranging from 10 to 200 m, and at some point have experienced moderate deformation. Ice tends to be thick either because ice must be thick to survive in this dynamic region, because thin ice is rapidly crushed during the constant collisions with other floes, or because of the rafting of thin ice floes. Small ice pieces between floes, and the ridges or blocks strewn around floe rims, and the moderately deformed floe interiors produce significant backscatter. In regions of constant shear, such as between shorefast and pack ice, processes are very dynamic and result in topographically rough, spatially variable, and very thick formations. Backscatter is strong and variable regardless of original ice type. The region which is spatially the most significant is the pack ice portion of the Arctic Ocean. Here floes are frozen into fields which float as sheets with sizes often many ten's of kilometers. Ridge building due to pressure is the major deformation event.

IN-SITU SCATTEROMETER PROGRAMS

By 1976 the need to coordinate detailed microwave and surface measurements was well recognized. In addition, because existing sensor parameters varied and future sensors were still undefined it was important to acquire data over a wide range of frequencies, polarizations, and incidence angles. It is interesting to note that the first measurements were made at 1.5 GHz and multiple frequencies between 8 and 18 GHz [1]. Today they are made from 0.5 to 100 GHz (i.e. CEAREX). The sea ice measurement program began at the University of Kansas by Richard K. Moore and the author with a tripod-type structure (1976) whose characteristic was to provide a constant range. It has been extended to operation from helicopter, surface ship, sled, and gantry (laboratory facility). Microwave scene characterization experiments began in 1977 at Point Barrow in conjunction with W.F. Weeks (CRREL and Univ. of Alaska) who served as mentor for sea ice property measurements. Numerous measurement programs were then conducted with R. Ramseier (AES Canada) with whom measurement of physical-chemical properties would be finely honed with the goal of relating specific microwave responses and processes to ice properties. Measurements were made of the surface for the calculation of roughness statistics (rms height and correlation length), of the salinity and density profiles with fine depth resolution (1 cm spacings in the top 20 cm of the ice sheet) to describe a rapidly changing permittivity, and of the brine or brine-slush layer which resides on all new and first-year ice forms. Coordinated active (near-surface scatterometer and SLAR) and passive microwave (aircraft imager and profilers) measurements began with those made at Mould Bay in 1981 as did the use of radar (SLAR) imagery in-situ experiment planning. Coordination of surface active and passive microwave

measurements which began at Mould Bay in 1982 have continued with CEAREX and CRRELEX being examples of the most recent investigations. Basic information about these measurements programs is provided in Table 1. It is important to note that as of this year significant studies have been conducted during all seasons, except perhaps the very end of summer when the ice drains, remains drained, and then begins to freeze. All major ice types have now been observed. Possibly the greatest weakness in the existing observations is that of region. Studies have been made in the Chuckchi, Beaufort, Greenland, and Barents Seas, with almost no study of the Central Arctic.

OBSERVATIONS

In examining under what conditions ice types may be discriminated using radar it was determined that it was important to (a) have data acquired over a wide range of conditions where both the microwave and physical-electrical properties are well-characterized, (b) be able to interpret the empirical observations with supporting electromagnetic scattering theory and model predictions, and (c) to study the variation in sea ice physical-electrical properties. Early on it was hypothesized that first year ice backscatter is dominated by surface scattering and multiyear ice by volume scatter. The critical issues have been in documenting under what conditions this is true, at what frequencies and polarizations, can this difference be exploited, what are the optimum radar parameters for ice type discrimination, and given a radar parameter set what is the anticipated performance.

The parameters and conditions which have a critical influence on backscatter intensity and are typical are listed in Table 2 as a function of major ice type. The attempt here is to list these parameters in their relative order of importance. When optimizing for ice type discrimination, the keys to information exploitation are (a) multiyear ice produces a strong volume scatter due to the presence of a low density ice layer in the upper portion of the ice sheet (a minimum thickness of about 2 cm is required), (b) first year ice returns are dominated by surface scatter and variations in surface roughness scales produce large changes in backscatter levels, and (c) new ice returns are weak (off-vertical) because surfaces are very smooth.

For multiyear ice during winter, volume scatter dominates at frequencies beyond 10 GHz where as at some frequency below about 5 GHz surface scattering begins to dominate. For first year ice surface scattering dominates below 15 GHz. Snow has little impact on multiyear ice backscatter but enhances first year and new ice backscatter [2].

The evolution of the microwave signatures during summer is quite complex. The reader is referred to the discussion provided in the paper by Onstott et al [3]. During the first part of summer the wet snowpack and warm ice sheet produces no significant volume scatter. First-year and multiyear ice backscatter becomes difficult to distinguish. By midsummer, thin first year ice backscatter is enhanced due to an increased small scale surface roughness from a superimposed ice layer which forms at the snow-ice interface and a snow thickness reduced by melt. After midsummer, the backscatter contrast between first year and multiyear is improved because first year ice roughness elements are smoothed by melt and multiyear ice continues to have a complex topography.

Data have been acquired to verify empirically the importance of volume scattering and the low density ice layer to multiyear ice backscattering and of surface scattering as the dominant process for first year ice. Spatial scanning a surface-based scatterometer operating at 5 GHz produced the radar maps provided in Figure 1 and 2. Range is inversely proportional to the (intermediate) frequency shown. The key features are that (a) the returns for first year ice backscatter are confined to its surface and few emanate from within the ice sheet, (b) returns are strong in the region where the low density ice layer is present and are produced below the snow-ice interface, and (c) the returns in the meltpool region of the multiyear ice sheet where the 3-m thick ice sheet is topped by 40 cm of fresh ice the returns are weak even though the ice contained larger gas bubbles (3 to 4 mm diameters) but in numbers not large enough to create a significant backscatter. In Figure 2 is a 24m x 20 m ground map of backscatter intensities acquired for these same multiyear ice features. These data show that the returns are 20 dB lower in the region of the meltpool. What is demonstrated here is the important of the low density ice layer in producing the enhanced return observed from multiyear ice. In addition, it not enough that discontinuities are present, they have to be present in sufficient numbers, hence, upper ice sheet density is a critical factor.

Based upon measured data, physical properties, and a theoretical foundation, predictions based on electromagnetic modeling were made. The role of volume and surface scattering were examined in the context of multiyear and first year ice. Characterizations were made in terms of the low density ice layer (i.e. thickness, air bubble size, density, and complex dielectric constant) of the multiyear ice and the surface of the first year ice (rms height, correlation length, and dielectric constant). In addition, the importance of the surface roughness of multiyear ice was examined as illustrated in Figure 3. Multiyear ice with a rough surface produces an enhanced backscatter at the low frequencies, but a slightly reduced volume scatter at high frequencies. This figure shows the range of surface scatter which may be occur for reasonable surface roughness values.

One of the key studies made during CEAREX by Onstott was of the variation in the thickness, density and air bubble composition of the multiyear low density ice layer. Based upon these in-situ observations predictions of backscatter levels as a function of density and layer thickness have been made for intercomparison with measured data. As is supported by the model predictions shown in Figure 4, backscatter intensity should be very sensitive to layer thickness for surfaces that are not rough. This response is observed in the CEAREX data. Layer thicknesses range from 0 to 15 cm and air bubble sizes of about 1.5 mm is typical.

Significant variations in the backscatter of both multiyear and first year ice can occur due to variations in surface roughness and properties of the low density ice layer of multiyear. The range of variations illustrated in Figure 5 are predicted values based on measured physical properties values. Observations shown in Figure 6 were made during CRRELEX-89 [4] and serve to illustrate the extreme range of possible signatures, in this case, those of thin first-year ice. Data have been acquired to address the important issue as to what is typical and what is the variation in the physical properties?

For this discussion it is assumed that the number of spatial samples included in an observation is large enough that sampling is not the dominant issue and we can examine floe-to-floe variability and differences in mean cross-sections. In discriminating multiyear ice from thinner ice the enhancement gained through volume scattering is exploited. Air bubbles are

typically about 1mm to 2mm in diameter. Operation at frequencies greater than about 4 GHz is necessary and contrast is known to improve with frequency up to at least 18 GHz, with optimum frequencies probably from 8 to 18 GHz. Operation at L-band (i.e. 1.25 GHz) has been demonstrated to be too low and at Ka-band (i.e. 35 GHz) all surfaces may appear rough. The enhancement in volume scatter by choosing a smaller wavelength may be offset by an enhanced surface scatter. Operation at C-band is a reasonable compromise because during the period after midsummer it is important to penetrate the thin snow pack and produce scatter from the rough snow-ice interface which allows the discrimination of thin and thick ice.

Discrimination of second year from multiyear ice depends on differences in backscatter intensity which arises due to differences in the thickness of the low density ice layer. Observations at Mould Bay in 1983 and CEAREX-88 suggest that the second-year low-density ice layer is less well developed in both thickness and size on air bubbles. Returns fall between multiyear ice and moderately-rough first-year.

To discriminate between first year and thin ice types it is necessary to exploit the differences between surface roughness and dielectric constant. There are two major dielectric constant regimes; one when the ice is very new and the other contains all remaining cases. The case in which a slush layer is present on thin-to-thick ice probably falls into the second category due to the presence of snow and its expected effect. Observations during CEAREX suggest that a slush and snow layer composite tends to reduce backscatter intensities. New ice can be discriminated older first year ice because of its very smooth surface. As first year ice ages the key consideration is a canopy composed potentially of frost flower remnants, snow, and metamorphosed snow. As the canopy becomes thicker, ice crystals enlarge, and the ice surface roughens backscatter will be enhanced. Analysis of CEAREX data will provide the additional information needed to determine if medium and thick first-year ice are discriminable.

SUMMARY

This paper presents the logic and mechanisms used in discriminating sea ice types. The quantitative analysis was reserved for the presentation at the symposium. The ability to discriminate ice types have been summarized in Figures 7 for winter conditions and Figure 8 for summer conditions. These curves are based on measured data and model predictions.

ACKNOWLEDGEMENTS

This work was supported under the Office of Naval Research (ONR) and the National Aeronautics and Space Administration (NASA) Contract N00014-86-C-0469. The technical monitors for this work were Dr. R.H. Thomas (NASA) and Mr. C.A. Luther (ONR).

REFERENCES

1) Onstott, R. G., R. K. Moore, and W. F. Weeks, "Surface-Band Scatterometer Results of Arctic Sea Ice," IEEE Trans. Geosci. Electron., GE-17 (3), 78-85, 1979.

2) Kim, Y. S., R. G. Onstott, and R. K. Moore, "The Effect of a Snow Cover on Microwave Backscatter from Sea Ice," IEEE Journal of Oceanic Eng., Vol. OE-9, No. 5, December, 1984.

3) Onstott, R. G., T. C. Grenfell, C. Matzler, C. A. Luther, and E. A. Svendsen, "Evolution of Microwave Sea Ice Signatures During Early Summer and Midsummer in the Marginal Ice Zone," Journal of Geophysical, Vol. 92, No. C7, pp. 6825-6835, June, 1987

4) Onstott, R. G. and S. H. Gaboury, "Polarimetric Radar Measurements of Artificial Sea Ice," IGARSS '89 Proceedings, July 1989.

5) Kim, Y. S., R. K. Moore, R. G. Onstott, and S. Gogineni, "Towards Identification of Optimum Radar Parameters for Sea Ice Monitoring," Journal of Glaciology, Vol. 31, No. 109, 1985.

TABLE 1. NEAR-SURFACE SCATTEROMETER INVESTIGATIONS

Year	Investigation	Location	Months	Ice Types	Frequencies	Comments
1989	CEAREX	Fram Str	Mar	OW, N PC, FY MY	0.1 to 100 18 to 35 (FP)	Evol. of OW to New Pancake & Ocean Study
1989	CEAREX	Barents	Feb	OW N PC FY MY	0.5 to 100	Various Scales of Deformation
1988	CRRELEX	Ice Tank	Jan Feb	OW-ThFY	1.8 to 35 (FP)	Evol. of OW to ThFY Multiyear Simulation 1st Pol. Measurements
1988	CEAREX	Ice Tank	Nov Dec	FY, MY	1.5 to 100	Winter Conditions
1988	CEAREX	Ice Tank	Sept Oct	Ow FY, N, SY MY	0.5 to 100	Fall-Freeze Up
1988	CRRELEX	Ice Tank	Jan-Feb	OW-ThFY	1.8 to 10 (FP) 10 to 100	Evol. of Ow to ThFY Surface Roughness Study
1988	BEEPERS	Bornea	Mar	BRACKISH	10	Brackish Ice
1987	MIZEX	Fram Str	Mar	OW FY	0.5 to 35	Winter
1985	CRRELEX	Ice Tank	Jan-Mar	OW-ThFY	5 to 18	Evolution Study
1984	MIZEX	Fram Str	Jun-Jul	OW-ThFY MY	1.5 to 18	Summer
1983	MIZEX	Fram Str	Jun-Jul	OW-ThFY-MY	1.5 to 18	Summer
1983	FIREX	Mould Bay	Apr	FY SY	5 to 18	SY Ice Study
1982	FIREX	Mould Bay	Jun Jul	OW TFY MY	1.5 to 18	Summer
1982	LAB SEA	Labrador	Feb	OW NW ThFY	1.5 to 18	FY Ice Forms
1981	SURSAT	Mould Bay	Sept Oct	OY MY	1.5 to 18	Fall Freeze Up
1980	YMER	E G S	Sept	OW FY MY	1.5 to 18	Fall Freeze Up
1979	SURSAT	TUK	Mar	Brack FY	1.5 to 18	Winter
1978	UKANSAS	Pt Barrow	Apr	FY Lake	1.5 to 18	Winter
1977	UKANSAS	Pt Barrow	May	TFY MY Lake	1.5, 8 to 18	Spring

Legend

Locations are Fram Str (Fram Strait) E G S is (East Greenland Sea) and TUK is (Tuktoyaktuk, N.W.T.)

Ice Types are OW (Open Water), N (New), PC (Pancake) ThFY (Thin First Year), FY (First Year) SY (Second Year) and MY (Multiyear)

TABLE 2. SCENE PARAMETER WHICH INFLUENCE BACKSCATTER INTENSITY

MULTIYEAR ICE

- Presence of LDI Layer in Upper Portion of Ice Sheet
  - LDI Thickness
  - LDI Density
  - LDI Air Bubble Distribution
- Surface Roughness
- Topography
  - Fast Ice
  - Pressure Ridges
  - Meltpools
- Refreezing and Flooding of Ice Sheet

FIRST YEAR ICE

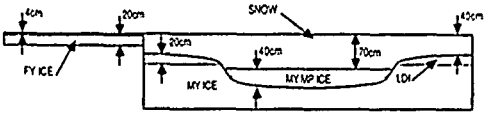
- Surface Roughness
- Brine Slush Layer (Significant Influences on Magnitude of ε_r')
- Snow Cover
- Topography
  - Undisturbed Ice
  - Pressure Ridges and Rubble

NEW ICE

- Brine Layer (significant influence on Magnitude of ε_r')
- Surface Roughness and Presence of Frost Flowers
- Topography
  - Homogeneous Ice
  - Rafting
  - Pancakes

Legend
- LDI - Low Density Ice Layer
- ε_r' - Real Part of Complex Dielectric Constant
- ε_r'' - Imaginary Part of Complex Dielectric Constant

DIMENSIONS OF ICE REGIONS



FREQUENCIES WHERE INTERFACES OCCUR

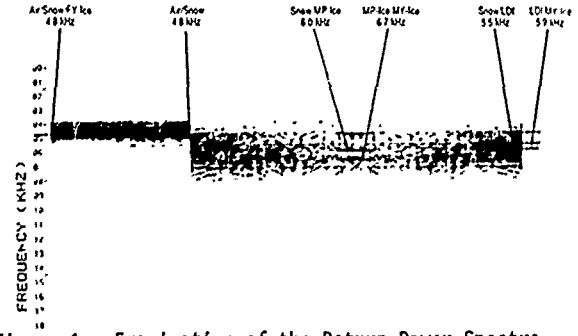


Figure 1. Examination of the Return Power Spectra Have Provided Empirical Validation of the Important Scattering Mechanisms of First-Year (i.e., Surface Scattering) and Multiyear (i.e., Volume Scattering) Ice. These Data were Collected at 5.0 GHz, VV-Polarization.

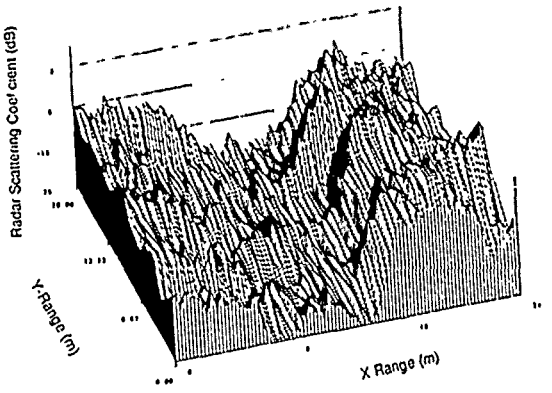


Figure 2. Data Acquired Using the Sled-Based Scatterometer to Create a Radar Backscatter Map of a 24 m x 20 m area. These Data Were Acquired at 5.0 GHz, VH-Polarization, and an Incidence Angle of 40 Degrees. The Scene is Composed of a Frozen Fresh Water Meltpool in a Multiyear Sea Ice Background.

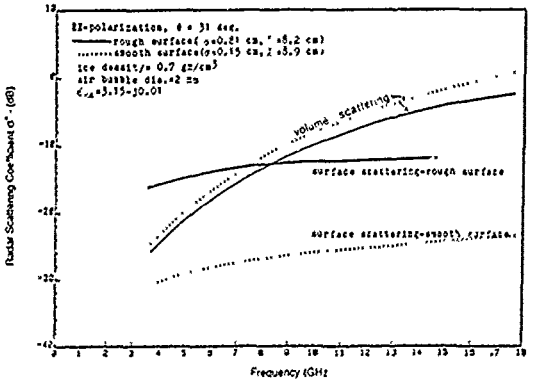


Figure 3. Theoretical Frequency Behavior of Radar Scattering Coefficient for Smooth and Rough Surfaces.

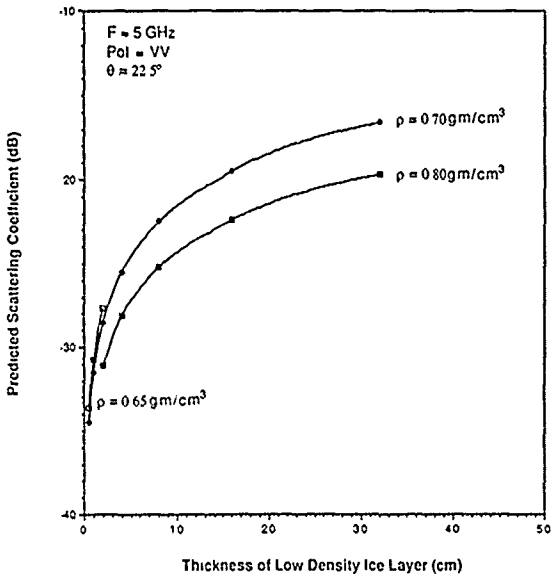


Figure 4. Predicted Scattering Coefficients at 5 GHz for Various Low Density Ice Layer Conditions.

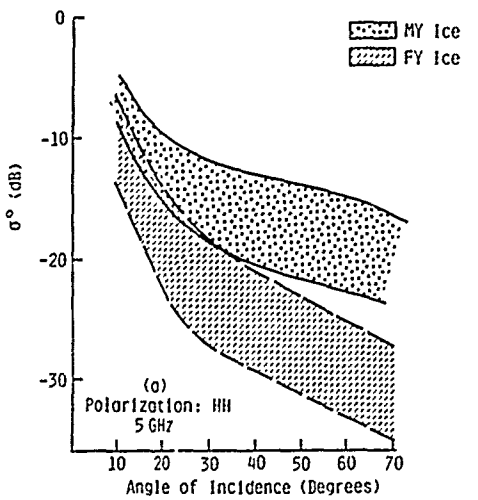


Figure 5. Theoretical Angular Variations of σ° of Multiyear Ice and First-Year Ice Using Measured Physical Property Parameters. [Kim et al, 1985].

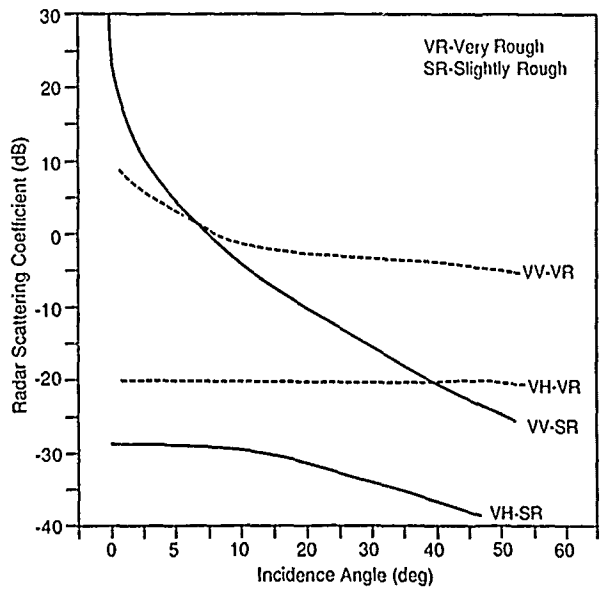


Figure 6. Radar Scattering Coefficients at 5.25 GHz for Very-Rough ($\sigma \approx 0.5$ cm and $\ell \approx 1.5$ cm) and Slightly-Rough ($\sigma \approx 0.05$ cm and $\ell \approx 1.6$ cm) Observed During CRRELEX '89 [Ref. 4].

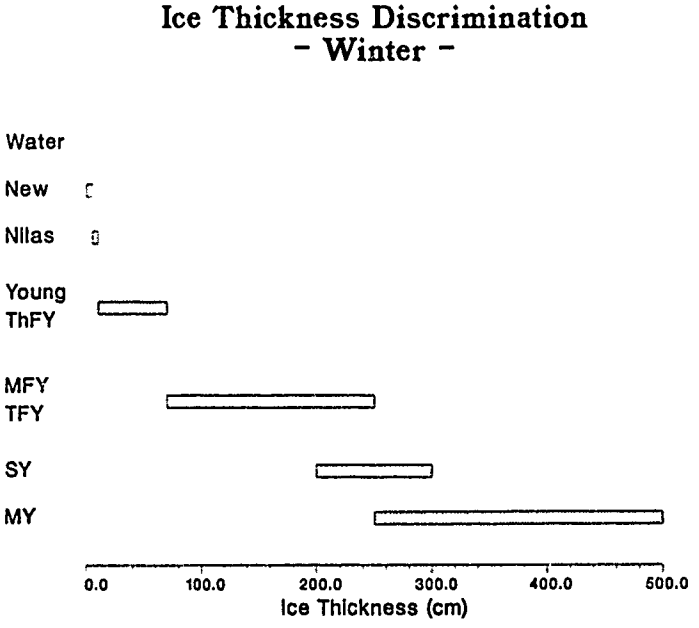


Figure 7. Major Ice Types Which May be Discriminated During Winter.

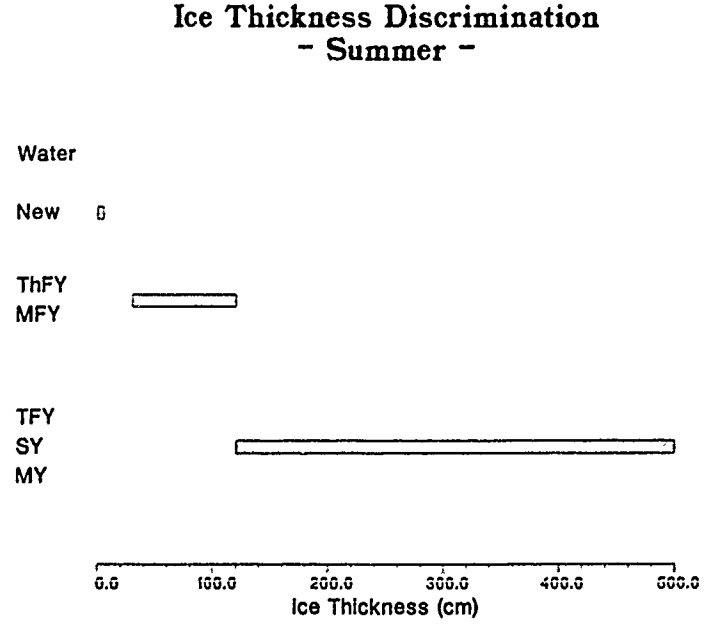


Figure 8. Major Ice Types Which May be Discriminated During Summer.

DISCRIMINATION BETWEEN NEW SEA ICE AND OPEN WATER -- COMPARISON OF C-BAND THEORY WITH MEASUREMENTS

Dale P. Winebrenner
Polar Science Center
Applied Physics Laboratory
University of Washington
Seattle, WA 98195
(206)543-6613

Prasad Gogineni and Jonathan Bredow
Remote Sensing Laboratory
University of Kansas
Lawrence, KS 66045
(913)864-7744

Leung Tsang
Department of Electrical Engineering
University of Washington
Seattle, WA 98195
(206)545-7537

Areal abundances and distributions of thin first year sea ice and open water are of particular importance in modulating heat and salt fluxes in ice-covered oceans. Thus it is of particular interest to discriminate open water, thin new ice (of thickness less than a few tens of centimeters), and thicker (first- and multi-year) ice in remote sensing data. We present a comparison of signature modeling and experimental results for multi-polarization active microwave measurements at C-band. Theory suggests a means of discriminating thin ice from open water using a multipolarization C-band synthetic aperture adar (SAR).

Theoretically, we employ the small perturbation method for rough surface scattering from new sea ice at C-band, taking the model beyond first order so as to account for depolarization. We combine this with a simple model for the dielectric constant of new ice as a function of thickness. We compare predictions from this combination of models with near-backscattering data from artificial new ice gathered as part of the CRRELEX experiment at the Cold Regions Research and Engineering Laboratory during 1988 and 1989. Model predictions for the ratio of like-polarized backscattering cross sections suggest that this signature may be useful for discrimination of open water from ice thicker than approximately 10 cm.

Key Words: new ice/open water discrimination, polarimetry

COMBINED ACTIVE/PASSIVE MICROWAVE CLASSIFICATION OF SEA ICE

C.E. Livingstone
Canada Centre for Remote Sensing
2464 Sheffield Road, Ottawa, Canada K1A 0Y7
Telephone: (613) 998-9060 Fax: (613) 993-5022 Telex: 053 3589

ABSTRACT

As sea ice grows and ages the structure and composition of the upper layers of the ice change in a manner that produces distinctive microwave signatures for a wide range of ice "types" under cold conditions. As temperatures rise above ~-5°C, these signatures undergo transitions due to changes in liquid water content and snow cover recrystallization until they are ultimately dominated by surface roughness and free water in the middle of the melt season.

Ice classification can thus be based on multi-polarization, multi-frequency passive microwave measurements of emissivity (or brightness temperature) or on multi-frequency, multi-polarization active microwave measurements of scattering cross-sections.

The complementary nature of scattering and emissivity make combined active/passive signature measurements a particularly powerful ice classification tool.

This paper reviews previously reported active/passive sea ice signature studies with emphasis on ice classification problems.

1. INTRODUCTION

Two ongoing areas of concern for human activities in sea-ice infested waters are the ice concentration and the spatial distribution of the sea ice types present. Sea ice types are defined by the stage of growth, the mechanical history and thermal histories of individual floes and consolidated areas within the ice cover. When regional information is included, the range of mechanical properties of the elements of the ice pack can be inferred. The ice concentration, combined with the ice type distribution indicates existing and potential navigation hazards in the ice pack for vessels of a given class and also defines useable routes. From a scientific viewpoint the ice type/concentration information has much broader implications (climatology, transport mechanisms, etc.).

The purpose of this paper is to outline some of the work that has been done and some of the problems that have been encountered in sea-ice classification using microwave sensors.

This paper is not a comprehensive review of the art nor are the references used of any special significance, other than providing snapshot views of a far ranging and vigorous field of research.

2. SEA ICE AND ITS SIGNATURES

As sea ice grows from a loose suspension of ice crystals in the sea surface to a massive solid, the surface layers undergo a series of transformations in crystal structure, brine/air inclusion density, shape and size as well as surface granularity and larger scale surface roughness. In mechanically disturbed areas such as marginal ice zones the normal "quiet" evolution of the ice surface is augmented by brash formation prior to consolidation, ice levee build up at the edges of consolidated ice cakes and rafting. Mechanical effects in thicker, "brittle" pack ice include rafting (in thin young ice) and ridging (in thicker ice). As surface temperatures decrease towards -40° the brine volume in the surface layers decreases. Periodic warming to near the freezing point results in recrystallization, brine migration and the growth of air filled voids (bubbles). Accumulating snow cover at various phases of ice growth alters the small scale surface structure in ways that are dependent on the ice growth stage and local temperature. These changes are preserved as the ice grows to its final thick first year mature stage. Scale sizes of surface elements vary from less than 1 mm (air bubbles, brine pockets and crystals) to more than 1 m (ridge rubble).

The ice decay process is characterized by an increase in moisture within the overlying snow pack and recrystallization at the snow-ice interface; an accumulation of water at the snow-ice interface and the growth of superimposed rough ice; melt pond formation and, (in close pack) eventual surface flooding. Thin ice disintegrates, thicker ice loses much of its surface layer salinity through brine drainage and brine inclusions are replaced by air pockets. If the first year ice survives until the melting process is halted by a return to lower temperatures it is transformed to second year ice - which, by virtue of its decreased salinity is much stronger than the mature first year ice. It also has very different surface layer characteristics.

Ice that survives more than one summer's melt becomes multi-year ice and has surface layer characteristics that stabilize (in a statistical sense) with increasing age. Within this simple minded outline of the history of sea ice are many developmental stages with recognizable characteristics - "ice types" - if you like. The youngest consolidated ice, Nilas, is very

smooth and very fragile and has a film of high salinity brine. Left undisturbed Nilas grows to "grey" ice which in turn thickens to "grey white ice". With time and low temperatures, "grey-white ice" becomes first year ice.

A more rigorous discussion of ice types and characteristics can be found in [1]. The foregoing, however, is sufficient to sketch the scope of the classification problem.

In principle the problem of classifying sea ice in a region reduces to the problem of observing the detailed evolution of all ice formed in and transported through the region of interest. Since time sequence observation with high temporal and spatial resolution is at present impractical, signatures that will summarize the salient elements in a snapshot view are required.

The sea ice and its snow cover is in general a three phase medium; ice crystals, air pockets and brine. Each constituent has a very different complex dielectric constant at microwave frequencies and exists in characteristic sizes, shapes, volume distributions and surface distributions at each stage of ice development and thermal history. The surface and volume distributions of the phases contribute to the scattering and extinction of incident electromagnetic waves and to the scattering and source energy of thermal emissions. Since the scale sizes of the constituents (~mm to ~ cm within the ice/snow volume and ~mm to ~m on the ice surface) correspond to the range of wave lengths found in the microwave portion of the spectrum, characteristic microwave signatures are expected.

Because of its large, complex dielectric constant at microwave frequencies, the liquid phase plays a dominant role under warm conditions and is important in establishing radiation penetration depths in the ice at all other times.

In spite of the sometimes varied histories of ice samples comprising an ice type, early airborne experiments with active [2] and passive [3] instruments showed that ice type dependent signatures do exist. As a result, the relationships between the structures of sea ice and its microwave characteristics have been a subject of ongoing research a) in the field [4], [5], [6], [7], [8], [9], [10], [11], [12] and b) in the laboratory [13], [14], [15].

From airborne and surface experiment results, ice type signatures based on single frequency, single polarization active or passive point measurements contain fundamental ambiguities from the ice classification viewpoint [16].

Early multi-sensor, multi-frequency, polarization diversified experimental data sets [17], [18], [19], [20] indicated three possible approaches to resolve the ice classification ambiguity problem in a point by point measurement basis:

(a) spatially and temporally coincident multi-frequency multi-polarization measurement of microwave brightness temperatures;

(b) measurement of spatially and temporally coincident scattering cross sections using multiple frequencies and polarizations;

(c) measurement of spatially and temporally coincident multiple polarization scattering cross sections and microwave brightness temperatures.

All three approaches have been pursued over the past decade with reasonable success.

3. COMBINED ACTIVE-PASSIVE SIGNATURES OF SEA ICE

The active-passive investigations followed from an observation based on figure 4 in [20] that simultaneous scattering cross section and emissivity measurements could indeed be complementary elements of a sea ice signature set. Results of a winter Beaufort Sea experiment in 1979 were particularly exciting when expressed in a scattering cross section, depolarization ratio, emissivity feature space[1] [21], [22]. This became the first of a series of airborne experiments conducted from 1979 to 1982 to investigate the seasonal and regional dependencies of multi-parameter sea ice signatures [23], [24], [25]. The work as summarized in [26] noted:

a) During the fall, winter and early spring major ice types are separated by more than one standard deviation in a feature space whose axes are the HH and HV polarized 13.3 GHz normalized scattering cross sections at 45° incidence angle and the H polarized 19.4 GHz emissivity at 45° incidence angle. Ice type classification accuracies in this space, for this period exceed 95%.

b) In the early spring, transformations within the snow pack produce multiple distinct signatures for the same ice type and regional effects are probably very significant.

c) During the late spring the snow-ice interface is saturated with water and the wet snow cover precludes the measurement of ice type signatures. In mid-summer, the water saturated ice surface and the melt pond distribution on the ice severely limit the classification potential of this technique. The ice is still, however, distinguishable from open water but it must be noted that a melt pond constitutes open water.

The technique explored in this research program used data from co-registered profiling radiometer and scatterometer sensors with similar resolution cell sizes. Its generalization to imaging sensors (either airborne or spaceborne) requires registration of near simultaneous data from spatially offset vehicle tracks for sensors which are either located on a single platform or on separate platforms.

A suggestion by Cavalieri et al [27] would combine spaceborne SAR and radiometer data for ice concentration (and possibly ice type) measurements. Because of the very large mismatch in resolution cell size (12 to 30 km for the radiometers and 15 to 30 m for the SAR) some difficulties are anticipated in generating a simple data combination algorithm.

A recent experiment [28] completed in the Newfoundland ice pack in March 1989 combines spatially overlapped, two frequency airborne SAR data and scatterometer data from one aircraft with two frequency scanning radiometer data from another aircraft over a surface measurement site. The proposed analysis will examine combined active/passive signatures for a range of possible feature spaces.

---

[1] *It is suggested in [16] that the feature space used is in no sense unique for ice signature presentation and that multi-frequency scattering data or multi-frequency radiometer data could be usefully displayed in similar form.*

4. MULTI-PARAMETER PASSIVE SIGNATURE OF SEA ICE

The multi-frequency passive microwave studies of Wilheit et al [3] represent the starting point for a long sequence of airborne experiments aimed at understanding and quantifying the seasonal and regional behaviour of sea ice brightness temperature signatures [17], [19], [29], [30], [31]. In summary, this research has shown that well defined passive microwave signatures exist for a variety of sea ice types under cold conditions and that brightness temperature/ice type ambiguities seen in single frequency data can be resolved by combining measurements at two or more frequencies and both linear polarizations.

The launch of the Electrically Scanning Microwave Radiometer (ESMR-5) aboard the Nimbus 5 satellite in 1972 provided a first global look at the earth's ice cover [32]. However, the single channel, single frequency sensor could not be reliably used for accurate total ice concentration or for multi-year ice fraction measurements from a single pass without auxiliary data [33], [19]. The launch of the Scanning Multi-channel Microwave Radiometer (SMMR) on Nimbus 7 in 1978 overcame the ESMR single channel limitation by providing ten radiometer channels (Horizontal and Vertical polarizations at five frequencies between 6.6 GHz and 37 GHz). Algorithms developed for analyzing SMMR data make use of the multi-channel aspect of the data set to solve for a range of variables [34].

An assessment of total ice concentration and multi-year ice fraction measurement accuracy for freeze up conditions [35] has reported SMMR ice concentration errors ~ ± 3% and multi-year ice concentration errors ~ ± 10% when compared to aircraft measurements in a test area between Greenland and Svalbard in 1979.

To investigate the winter (relatively stable) emissivity signatures measured by SMMR for the arctic, Comiso [36] employed a series of feature spaces where axes were emissivities at specific frequencies and polarizations. Cluster analysis performed in these spaces reveal the existence of 5 different signatures for consolidated ice in the arctic. Some of these have been associated with ice type (first year, multi-year), others are considered to represent variations in surface conditions. Co-planarity tests in the feature spaces failed to reveal unambiguous mixing ratios in the 10.7, 18 and 37 GHZ data that would serve as a basis for reliable extraction of specific cluster signatures within the "multi-year ice" set.

One of the limitations of spaceborne radiometers for sea ice type classification is the large resolution cell size produced by the imaging geometry (30 km x 30 km for the 37 GHz channels of SMMR). As a result the contents of a given cell must be inferred from mixing formulae using available coarse resolution data and a detailed knowledge of sea ice emissivities derived from high resolution surface and airborne measurements. During the late spring and summer, snow moisture and surface water degrade the available major ice type signatures [36], and surface free water degrades the ice concentration measurement accuracy.

The recent launch of Special Sensor Microwave/Imager (SSM/I) package with its improved spatial resolution (12 km) at the 85.5 GHz upper frequency may provide better ice type discrimination from spaceborne sensors.

A digitally recorded two frequency airborne scanning radiometer, AIMR, has recently been commissioned by the Canadian Atmospheric Environment Service for use in SSM/I validation. Early results suggest that this instrument will be a powerful tool for sea ice signature research [37].

5. MULTI-PARAMETER ACTIVE SIGNATURES OF SEA ICE

Since the late 1970's interest in monitoring arctic sea ice by radar has driven a long series of investigations into the spectral properties of the microwave scattering cross section of sea ice. The dominant group in this area was based at the University of Kansas and concentrated on first surface and later near surface (helicopter borne) scatterometer measurements over the 1.0 GHZ to 35 GHZ frequency range. Initial studies concentrated on arctic winter ice in the Beaufort Sea [6]. Subsequent work followed the seasonal cycle and investigated arctic fall and summer conditions [38], [11]. Since this time much of the work done by Onstott has been concentrated in the Norcsex marginal ice zone research area between Greenland and Svalbard over the seasonal cycle [39], [40].

Since the emphasis of this work has been on surface and near surface local measurements, a large body of auxiliary ice surface and snow-ice interface data has been accumulated. The combined radar cross section measurement and surface data sets have been used in the development of a quasi-empirical ice scattering signature model [41] which has proven to be an extremely powerful tool for the interpretation and generalization of the multi-spectral scattering measurements [42]. Results of the accumulated multi-spectral studies have shown that sea ice can be reliably classified by ice type from multi-dimensional scattering cross section measurements under cold conditions (temperature < -5°C). As the high point of the daily temperature cycle approaches and exceeds the freezing point, snow pack transformations (especially at the snow-ice interface) which alter the scattering signatures are observed. It is expected that some of these transformations are responsible for the multiplicity of first year ice signatures reported in the active-passive studies in the early melt period. As snow melt progresses, the wet snow pack and accumulated free water at the snow ice interface dominate the microwave scattering signatures and ice type discrimination by scattering becomes much more difficult. In the marginal ice zone in the Greenland Sea, large areas of melt pond coverage were not observed within the ice pack and ice-water discrimination remained excellent for HH polarized radar throughout the measurement period.

Recent developments at JPL have seen the P, L, C band SAR system applied to the ice classification problem. Preliminary three band composite images [44] are promising for use as a sea ice classification tool.

Onstott et al [43] provide a detailed description of the structural changes that occur in the ice and its snow cover from winter to mid-summer and show how these changes influence active and passive microwave measurements of sea ice properties. The variation of microwave ice type signatures over the summer period for first year and multi-year ice are explained and the significance of the remnant snow pack during this season is emphasized.

6. SUMMARY

All three classes of multi-dimensional microwave signature spaces discussed are fully capable of providing accurate ice type identifications during the portion of the year when the sea ice and its snow cover are not saturated with water. During the summer melt

period, all electromagnetic measurement approaches to sea ice classification degrade or fail.

Carsey [45] analyzed summer measurements from ESMR and SMMR and found that by constructing averages of microwave brightness temperature with degraded resolution and combining this data with synthetic aperture radar imagery (SEASAT in this case) and with spaceborne scatterometer data, the gross behaviour of the polar ice pack was interpretable over the summer period. Sea ice concentrations could be inferred by observing variations of the signatures with time.

This paper has concentrated on the development of "point by point" sea ice classification using quantitative measurements. The accurate interpretation of high resolution radar and radiometer images by skilled professionals is well known. Less well known are the methods required to automate this process to provide accurate, machine interpretation based on spatial structure and texture as well as intensity distribution in an image. This classification area is accumulating a body of literature and holds its own fascination but unfortunately lies beyond the scope of this article.

REFERENCES

1. Weeks, W.F. and S.F. Ackley, "The growth, structure and properties of sea ice", CRREL Monogr., 82-1, US Cold Reg. Res. Eng. Lab., Hanover NH, 136 pp, 1982.

2. Rouse, J.W., "Arctic ice type identification by radar", Proc. IEEE, 57, #4, pp 605-611, 1969.

3. Wilheit, T., Nordberg, W., Blenn, J., Campbell, W. and A. Edgerton, "Aircraft measurement of microwave emissions from Arctic sea ice", Remote Sensing of Environment, 2, pp 129-139, 1972.

4. Vant, M.R., Gray, R.B., Ramseier, R.O. and V. Mahias, "Dielectric properties of fresh and sea ice at 10 and 35 GHZ", J. Appl. Phys., 45 #11, pp 4712-4717, 1974.

5. Vant, M.R., "A combined empirical and theoretical study of the dielectric properties of sea ice over the frequency range 100 MHZ to 40 GHZ", PhD Thesis, Carleton University, Ottawa, 1976.

6. Onstott, R.G., Moore, R.K. and W.F. Weeks, "Surface-based scatterometer results of Arctic sea ice", IEEE Trans., Geosci. Electron., GE-17 #3, pp 78-85, 1979.

7. Matzler, C., "Microwave signatures of young sea ice and its influence on ice concentration algorithms", Proc. ISPRS Commission VII, Toulouse, pp 757-765, Sept. 1982.

8. Lohanick, A.W., "Snow thickness and vertically polarized brightness temperature on multi-year ice", NORDA Technical Note 171, 11 pp, 1982.

9. Grenfell, T. and A.W. Lohanick, "Temporal variations of the microwave signatures of sea ice during the late spring and early summer near Mould Bay, NWT", J. Geophys Res., 90, pp 5063-5074, 1985.

10. Lohanick, A.W. and T.C. Grenfell, "Variations in brightness temperature over cold first year sea ice near Tuktoyaktuk", J. Geophys. Res., 91, C4, pp 5133-5144, 1986.

11. Onstott, R.G. and S.P. Gogineni, "Active microwave measurements of Arctic sea ice under summer conditions", J. Geophys. Res., 90, AC3, pp 5035-5044, 1985.

12. Hallikainen, M.T., Hyyppä, J.M., Taikka, M.V.O., Haapanen, J.A.E., Tares, T.I. and P.J. Ahola, "Backscatter behaviour of low salinity sea ice at C- and X-band", Proc. IGARSS '88, Edinburgh, pp 791-792, Sept. 1988.

13. Stogryn, A. and G. Desargent, "The dielectric properties of brine in sea ice at microwave frequencies", IEEE Trans. Ant. and Prop., AP-33(5), pp 40-48, 1985.

14. Arcone, S.A., Gow, A.J. and S. Grew, "Structure and dielectric properties at 4.8 and 9.5 GHZ of saline ice, J. Geophys. Res., 91, #C12, pp 14281-14303, 1986.

15. Swift, C.T., De Harity, D.C., Tanner, A. and R.E. McIntosh, "Passive microwave emission from saline ice at C-band during the growth phase", IEEE Trans. Geosci. Remote Sens., GE 24, #6, pp 840-848, 1986.

16. Livingstone, C.E., Hawkins, R.K., Gray, A.L., Drapier-Arsenault, L., Okomoto, K., Wilkinson, T.L. and D. Pearson, "The CCRS/SURSAT active-passive experiment 1978-1980: the microwave signatures of sea ice", Canada Centre for Remote Sensing Report, 211 pp, June 1983.

17. Tooma, S.G., Mannella, R.A., Hollinger, J.P. and R.D. Ketchum, "Comparison of sea ice type identification between airborne dual frequency passive microwave radiometry and standard laser/infra red techniques", J. Glaciology, 15, pp 225-239, 1975.

18. Ketchum, R.D. and S.G. Tooma, "Analysis and interpretation of airborne multi-frequency side looking radar sea ice imagery", J. Geophys. Res., 78, #3, pp 520-538, 1973.

19. Gloersen, P.G., Zwally, H.J., Chang, A.T.C., Hall, D.K., Campbell, W.J. and R.O. Ramseier, "Time dependence of sea ice concentration and multi-year fraction in the Arctic basin", Boundary Layer Meteorology, 13, pp 339-359, 1978.

20. Gray A.L., Ramseier, R.O. and W.J. Campbell, "Scatterometer and SLAR results obtained over Arctic sea ice and their relevance to problems of Arctic reconnaisance", Proc. 4th Canadian Symposium on Remote Sensing, Quebec City, pp 424-442, 1977.

21. Hawkins, R.K., Livingstone, C.E., Gray, A.L., Okomoto, K., Drapier-Arsenault, L. and D.E. Pearson, "Single and multiple parameter microwave signatures of sea ice", Proc. 6th Canadian Symposium on Remote Sensing, Halifax, pp 217-248, 1980.

22. Livingstone, C.E., Hawkins, R.K., Gray, A.L., Okomoto, K., Wilkinson, T.L., Young, S., Drapier-Arsenault, L. and D. Pearson, "Classification of Beaufort Sea ice using active and passive microwave sensors", Proc. COSPAR/SCOR/IUCRM Symposium on Oceanography from Space, Plenum NY, pp 813-821, 1981.

23. Hawkins, R.K., Gray, A.L., Livingstone, C.E. and L. Drapier-Arsenault, "Seasonal effects on the microwave signatures of Beaufort Sea ice", Proc. 15th International Symposium on Remote Sensing of the Environment, Ann Arbor, pp 239-257, May 1981.

24. Gray, A.L., Hawkins, R.K., Livingstone, C.E., Drapier-Arsenault, L. and W.M. Johnstone, "Simultaneous scatterometer and radiometer measurements of sea ice microwave signatures", IEEE Trans. Ocean. Eng., OE-7, pp 20-33, 1982.

25. Livingstone, C.E., Onstott, R.G., Drapier-Arsenault, L., Gray, A.L. and K.P. Singh, "Microwave signatures at the onset of melt", IEEE Trans. Geosci. and Rem. Sensing, GE25, #2, pp 174-187, 1987.

26. Livingstone, C.E., Singh, K.P. and A.L. Gray, "Seasonal and regional variations of active/passive microwave signatures of sea ice", IEEE Trans. Geosci., Rem. Sensing VGE25, #2, pp 159-173, 1987.

27. Cavalieri, D.J., Burns, B.A. and R.G. Onstott, "Investigation of multi-dimensional algorithms using active and passive microwave data for ice concentration determination", Proc. IGARSS '87, Ann Arbor, pp 1151-1153, 1987.

28. Livingstone, C.E., Brown, P., Collins, M., Asmus, K. and S. Argus, "LIMEX '89 active/passive microwave signatures of Newfoundland pack ice", presented at IGARSS '89, Vancouver, 1989.

29. Campbell, W.J., Iversen, P.G., Zwally, H.J., Ramseier, R.O. and C. Elachi, "Simultaneous passive and active microwave observations of near shore Beaufort Sea ice", J. Petrol-Technol., 21, pp 1105-1112, 1980.

30. Troy, B.E., Hollinger, J.P., Lerner, R.M. and M.M. Wisler, "Measurements of the microwave properties of sea ice at 90 GHZ and lower frequencies", J. Geophys. Res. 86, pp 4283-4289, 1981.

31. Hollinger, J.I., Troy, B.E., Ramseier, R.O., Asmus, K.W., Hortman, M.F. and C.A. Luther, "Microwave emission from high Arctic sea ice during freeze-up", J. Geophys. Res. 89, pp 8104-8122, 1984.

32. Gloersen, P., Wilheit, T.T., Chang, T.C., Nordberg, W. and W.J. Campbell, "Microwave maps of the polar ice of the Earth", Bulletin Amer. Met. Soc., 55, #12, pp 373-389, 1974.

33. Comiso, J.C. and H.J. Zwally, "Antarctic Sea ice concentrations inferred from Nimbus 5 ESMR and Landsat imagery, J. Geophys. Res., 88. #C8, pp 5836-5844, 1983.

34. Gloersen, P., Cavalieri, D. and W.J. Campbell, "Derivation of sea ice concentration, age and surface temperature from multi-spectral radiance obtained with the Nimbus 7 scanning multi-channel microwave radiometer", Oceanography from Space, pp 823-829, 1981.

35. Svendsen, E., Kloster, K., Farrelly, B., Johannessen, O.M., Johannessen, J.A., Campbell, W.J., Gloersen, P., Cavalieri, D. and C. Matzler, "Norwegian remote sensing experiment: evaluation of the Nimbus 7 scanning multi-channel microwave radiometer for sea ice research", J. Geophys. Res., 88, #C5, pp 2781-2791, 1983.

36. Comiso, J.C., "Characteristics of Artic winter sea ice from satellite multi-spectral observation", J. Geophys. Res., 91, #C1, pp 975-994, 1986.

37. Ramseier, R.O., "Early results with the AIMR System", Private Communication, 1989.

38. Onstott, R.G., Kim, Y.S. and R.K. Moore, "Active microwave measurements of sea ice under fall conditions: the RADARSAT/FIREX fall experiment", U. Kansas Remote Sensing Laboratory Report TR-331-30/578, 1984.

39. Onstott, R.G. and R.K. Moore, "Active microwave measurements of sea ice in the marginal ice zone under summer conditions", Proc. IGARSS '84, Strasbourg, pp 359-363, 1984.

40. Onstott, R.G. and R.K. Moore, "Radar backscatter of sea ice during winter", Proc. IGARSS '88, Edinburgh, pp 1115-1118, 1988.

41. Kim, W.S., Moore, R.K. and R.G. Onstott, "Theoretical and experimental study of radar backscatter from sea ice", U. Kansas Remote Sensing Laboratory report RSL TR 331-37, 168 pp, 1984.

42. Onstott, R.G., "Theoretical and experimental study of the radar backscatter from Arctic sea ice", Proc. IGARSS '87, Ann Arbor, pp 1127-1129, 1987.

43. Onstott, R.G., Grenfell, T.C., Matzler, C., Luther, C.A., and E.A. Svendsen, "Evolution of microwave sea ice signatures during early summer and mid-summer in the marginal ice zone", J. Geophys. Res., 92, #C7, pp 6825-6835, 1987.

44. Crawford, J., "Three band: P, L, C SAR images of sea ice", Personal Communication, 1989.

45. Carsey, F.D., "Summer Arctic sea ice character from satellite microwave data", J. Geophys. Res., 89. #C3, pp 5015-5034, 1984.

THE DISCRIMINATION OF SEA ICE TYPES USING SAR BACKSCATTER STATISTICS

Robert A. Shuchman, Christopher C. Wackerman, Andrew L. Maffett*,
Robert G. Onstott and Laura L. Sutherland,

Radar Science Laboratory
Environmental Research Institute of Michigan
Ann Arbor, MI 48107 USA

*ERIM Consultant

ABSTRACT

X-band (HH) synthetic aperture radar (SAR) data of sea ice collected in March and April 1987 during the Marginal Ice Zone Experiment (MIZEX) was statistically analyzed with respect to discriminating open water, first-year ice, multiyear ice and Odden. Odden is large expanses of nilas ice that rapidly form in the Greenland Sea and transforms into pancake ice. A first order statistical analysis indicated that mean versus variance can segment out open water and first-year ice, and skewness versus modified skewness can segment the Odden and multiyear catagories. In addition to first order statistics a model has been generated for the distribution function of the SAR ice data. Segmentation of ice types was also attempted using textural measurements. In this case, the general co-occurrency matrix was evaluated. The textural method did not generate better results than the first order statistical approach.

1. INTRODUCTION

Active microwave measurements were made of various sea ice forms in March and April 1987 during the Marginal Ice Zone Experiment (MIZEX). The microwave measurements were made at 1, 5, 10, 18, and 35 GHz using a ship-based scatterometer and aircraft mounted synthetic aperture radar (SAR). The SAR measurements were made at 9.8 GHz, horizontal transmit and receive polarization.

The sea ice forms present in the Greenland Sea MIZEX operations area included: open water; open water with grease ice streamers; new ice (5-8 cm thick); first-year ice (20-40 cm thick); first-year ice with rubble (.60-1.5 m thick); and multiyear ice (2-4 m thick). Large expanses (200,000 km^2) of new nilias sea ice (5 cm thick) forms in the Greenland Sea as a result of oceanographic upwelling of cold water interacting with cold ($< -10°c$) northerly polar winds. This rapid ice formation is referred to as the Odden and is discussed in Ref. [1]. The nilas transitions into pancake floes (10-15 cm thick) due to continued growth and wave action.

The scatterometer data were used to validate the SAR backscatter values obtained at 10 GHz. Additionally, the scatterometer data extended the SAR sea ice type classification to C- and L-band frequencies in order to predict the performance of the SAR instruments to be flown on the European, Canadian, and Japanese SAR satellites.

NASA is presently building a satellite receiving station in Fairbanks, Alaska that will collect and process in near real-time SAR data from these satellites that will be launched in the next 5-7 years. The SAR facility will contain a geophysical processor, a dedicated computer that will utilize the processed SAR data to provide sea ice concentration, ice type, and kinematics information. The MIZEX '87 data although it was collected at X-band (HH) offers a unique opportunity to develop and evaluate algorithms that can be implemented on the geophysical processor. As reported in ref. [2], the radar backscatter response between X-band (HH) and C-band (VV) is very similar.

This SAR analysis included the generation of standard statistics (i.e., mean, standard deviation, variance, skewness, and kurtosis), within areas that were intensively "sea truthed" by scientists operating from the M/V POLAR CIRCLE. In addition to the standard statistics generated from the SAR data several probability distributions were evaluated to describe the various ice types present within the SAR scene. These distributions include: uniform, gamma, Gaussian, inverse Guassian, lognormal, and modified Beta. It is postulated that such distributions can be used to further differentiate ice types (particularly first-year ice with rubble). The distributional analysis has suggested that the SAR sea ice data are best fitted by gamma and lognormal (and sometimes inverse Gaussian) distributions and that these distributions may prove useful in differentiating all ice types present within the MIZEX SAR scenes.

The use of textural methods (i.e., higher order statistics) were also evaluated with respect to differentiating the ice sea forms present in the MIZEX SAR data set. The general co-occurrence matrix was used and found to generate very similar results to the mean and standard deviation analysis.

In this paper, we will first describe the MIZEX '87 SAR data set and then discuss four cases of sea ice forms that were selected for the statistical analysis. The statistics results will then be presented.

2. DATA SETS

MIZEX '87 in the Greenland and Barents Seas combined observations from both remote sensing and *in situ* data collections to provide an integrated approach to the study of winter marginal ice zone (MIZ) conditions. Favorable weather permitted 18 consecutive days of SAR coverage and field operations. The SAR system, with its high resolution (15 x 15 m), clarity of image and real-time availability, proved to be a powerful and efficient tool to aid in the planning and carrying out of field experiments.

MIZEX was the first international experiment having daily SAR coverage with real-time imagery down-linked

to the ships in the field. This imagery was used on-board POLAR CIRCLE to identify areas of interest such as the location of the ice edge, eddies, and ocean fronts. The ship would then proceed to the SAR identified areas to collect sea truth. The data was also used to select sites for detailed active and passive microwave measurements and characterization of physical and electrical properties of the ice and snow. In addition to being down-linked, the SAR data was recorded on-board the aircraft on high density digital tapes.

During MIZEX '87, two Intera SAR equipped aircrafts; STAR-1 and STAR-2 were deployed to collect ice edge imagery. These missions are described in refs. [3-4]. The Intera STAR-1 and STAR-2 systems are X-band (9.8 GHz) radars that transmit and receive with horizontal polarization. Table 1 summarizes the parameters of both systems while Figure 1 indicates the imaging geometries. Figure 2 is a representative mosaic covering a 445 x 195 km area of the MIZ. On the X-band SAR data, bright tones on the image represent multiyear ice while the darker tones are various stages on young ice. The blackest signatures on the image are open water. The SAR mosaic was interpreted using a hybrid manually assisted digital technique to provide the interpretation key shown in Figure 3. Note that the ice type and concentration information is obtainable from this data based on tonal signatures.

TABLE 1. STAR SPECIFICATIONS AS USED IN MIZEX

PROPERTY	STAR-2	STAR-1
Operating Altitude	29,000 ft.	
Wave length	X-band	
Polarization	HH	
Viewing Direction	Left or Right	
Processing	Real time	
Recordirg	8 bit data, full bandwidth data recording on parallel HDDR	4 bit data, either 12 x 12m or 24 x 24m pixels on serial HDDR
Swath width		
Narrow (HI-Res)	17 km	23 km
Wide (Lo-Res)	63 km	45 km
Pixel size	Along track/ cross track	Along t.ack/ cross track
HI-Res	4 x 4 m	Not used
Lo-Res	5.2 x 16m	12 x 12m or 24 x 24m
Downlink	4 bits	4 bits
Azimuth Looks	7	7
Lo-Res	16 x 16m or 32 x 32m	12 x 12m or 24 x 24m

For this study, the generic ice types present within the test site were divided into four catagories. These included: open water; first-year ice; multiyear ice; and the areas termed Odden (i.e., nilas and pancake). These four ice types represent ice thickness intervals that are accepted product outputs of the Alaska geophysical processor.

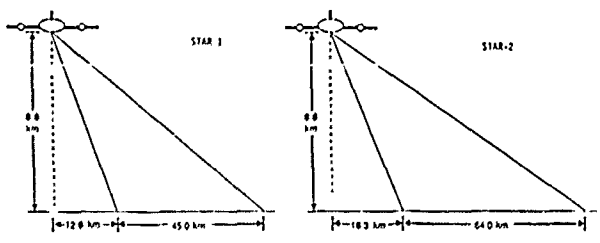


Figure 1. Viewing Geometry for STAR-1 and -2 in Wide Swath Mode

Fifty-four areas from twelve SAR flights during the 31 March to 8 April 1987 time period were used in the statistical analysis. Each area was approximately 100 x 100 pixels which corresponds to a ground area of approximately 700 x 700 meters. The areas selected included incident angles of 30° to 70°. Typical open water (w), first-year ice (f), multiyear ice (m), and Odden (O) areas used in the study are shown on Figure 2.

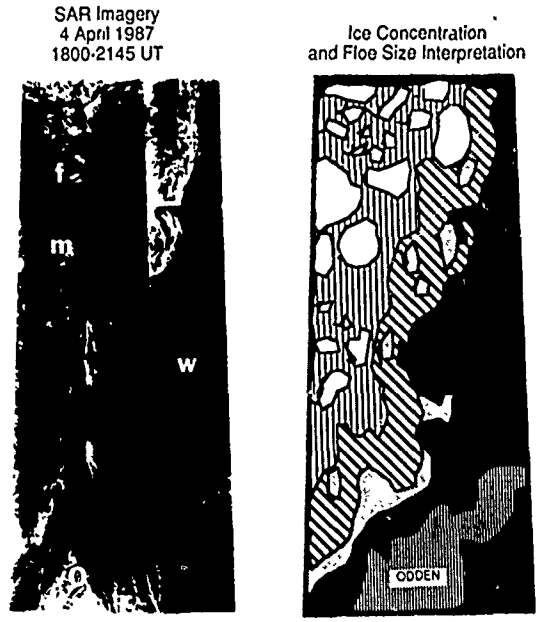


Figure 2.

Figure 3.

50% Multi-Year Fragments, (1.5 - 3.5 m Thick), 50-250 m in Diameter in a First Year Ice Framework, (< 1 m Thick) 20-45% Total Ice Concentration

50-60% Multi-Year Floes, (1.5 - 4 m Thick) 50-250 m in Diameter Surrounded by First Year Ice, (< 1 m Thick) 70-90% Total Ice Concentration, Some Open Water Leads Exist

40-50% Multi-Year Floes, 1-30 km in Diameter Surrounded by First-Year Floes, (< 1 m Thick) 95% Total Ice Concentration, Some Open Water Leads Exist

Ice Free Ocean and Polynyas

New Nilas Ice, 3-15 cm Thick Transitioning to Pancake Floes, 10-100 cm in Diameter 75-90% Total Ice Concentration

Large Individual Floes

3. DISCRIMINATION USING FIRST-ORDER STATISTICS

The ability of first-order statistics (i.e., statistics based on individual pixel values, not spatial correlations) to segment the SAR images into their appropriate four classes; open water, first-year ice, multiyear ice and Odden was analyzed. Since the results of an initial distributional fitting analysis [Ref. 2] indicated that two parameter functions did a good job of fitting the data, segmenting the data with

pairs of statistics was tried. The two best pairs are shown in Figures 4 and 5. Figure 4 shows mean versus standard deviation (all of these statistics are intensity values). Note that although Odden ice and multiyear ice are very mixed, the other two classes segment out rather nicely; almost linearly along the curve we have open water, then first-year ice, then the combination of multiyear ice and Odden ice.

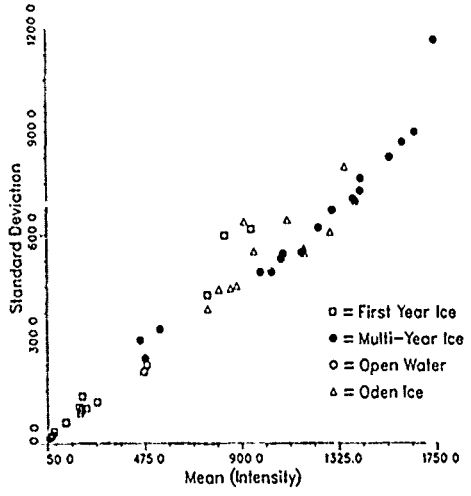


Figure 4. Segmentation Using Modified Skewness and Skewness

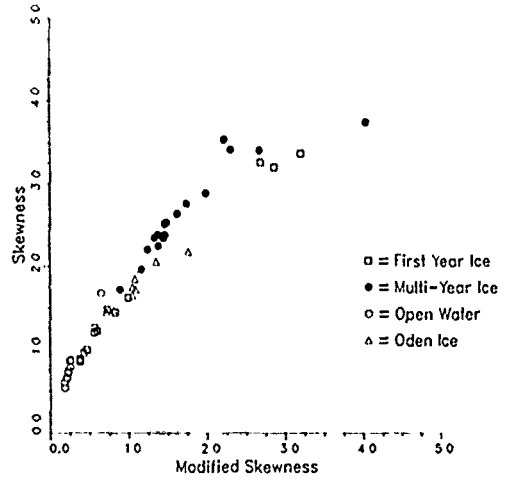


Figure 5. Standard Deviation vs Mean Segmentation of Intensity Ice Data

The Odden is representative of new first-year ice that has undergone extensive deformation due to wave action. The nilas ice that transitions into pancake floes have very rough edges (see Figures 2 and 4 in ref [1]). These rough edges reflect radar energy and as such the scattering coefficient for pancake ice is large [Ref. 2]. Thus, it is not surprising that multiyear and Odden ice types overlap on the mean versus standard deviation curve.

Figure 5 shows skewness versus modified skewness. Note that this also does a nice job of segmenting, although the distinction of first-year ice from Odden ice is somewhat less sharp then on Figure 4. However, multiyear ice and Odden ice are much better separated then on [illegible] Again, almost linearly along the curve, we [illegible] water, first-year ice, Odden ice and multiye[illegible].

In conclus[illegible] appears that mean versus variance can segment ou[illegible] water and first-year ice, then skewness versus [illegible]ified skewness can segment out Odden ice and multiyear ice. In general, first-order statistics appear to do a good job.

Currently under progress is an attempt to optimize the combination of first order statistics for segmentation. Eigenvector analysis is being performed on the covariance matrices for each ice class to determine which linear combination of parameters generates the most variation within the class and the least variation within the class. In addition, the same analysis is being performed on the entire set of SAR images. Optimal open water (for example) classification can then be performed by finding the linear combination that minimizes the variation within the class but maximizes the variation between classes.

4. DISTRIBUTION FUNCTION ANALYSIS

An attempt has been made to generate a model for the distribution function of the SAR ice data that would fit all of the MIZEX '87 images. If this is possible, then the parameters of the distribution function would suffice to characterize the image, and we could segment the images based on their parameters alone. Unfortunately, we have found no model that passes the statistical tests for all of the images; this search is still in progress. We have tested a number of simple analytical forms but the results indicate that the actual distribution function is more complicated than this. Currently, work is being performed on a model that deals more with the physical scattering phenomena [Ref. 5] and the results look encouraging. Unfortunately, with complexity comes computational cost so no definite results for the ice data are available yet.

For the analytical forms, we tested uniform, gamma, inverse gaussian, gaussian, lognormal and modified beta. Figure 6 shows the Kolmogorov statistic (a statistical measure of how different the analytical distribution function is from the actual SAR image distribution function) for the ice data set for all the types. Lognormal, inverse gaussian and modified beta all produce similar results with a preference to the modified beta. It should be noted that the modified beta is the only model that allows the distribution function to be fit to higher order moments (i.e., width and skewness) while the other models fit only the mean and variance. This appears to support the results in section 3 where these higher order moments are necessary to separate Odden ice from multiyear ice, and in addition did a fair job of segmenting the other types.

Although these three models do a good job of fitting to the data, they do not pass the statistical tests (i.e. it can not be said with statistical certainty that they are a good fit, we can only say that the fit looks good to the eye). Figure 7 shows the same Kolmogorov statistic but subtracted by the threshold that it has to be less than in order to pass the test. Thus values less than zero on Figure 7 mean it passed and values greater than zero indicate failure. Note the very large amount of failures; although there are a few passes. In addition, it appears that the modified beta advantage goes away for this plot; a result that is not explainable at the present time.

It is interesting to note how the amount of passes and failures change with ice type. Figure 8-11 shows the same numbers as Figure 7, but separated according to ice type. Note that open water fails miserably; an

interesting result which indicates that the scattering going on is not simple. First-year ice and multiyear ice do about the same; much more successful (for the three candidate models) then open water. This seems to indicate that scattering from these types is more "traditional". Finally, Odden does the best of all; the points are very clustered at zero. Perhaps this is the most "traditional"? Also note that the most successes appears to be for multiyear ice with first-year ice next.

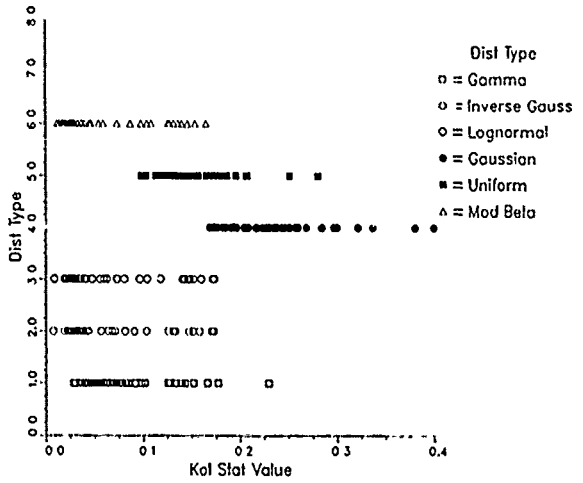


Figure 6. Ice Measurements Kolmogorov Statistics

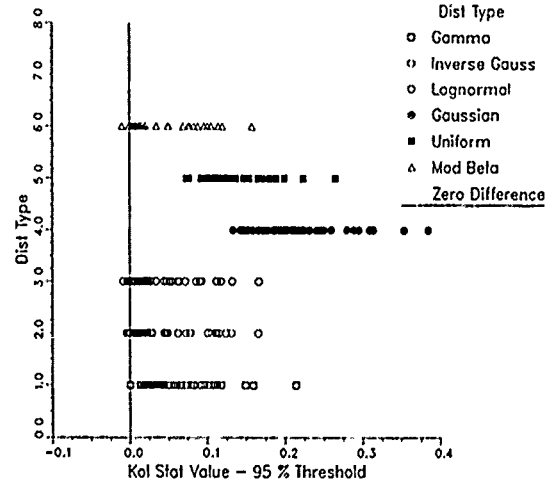


Figure 7. Ice Measurements Kolomgorov Statistics Subtracted by the Threshold Value

All of the above distribution models had their parameters generated from the data statistics. We also investigated generating optimal parameters by iteratively finding the ones that minimized a mean square error metric between the model and the data. Figure 12 shows the result for optimizing the parameters of a gamma model, then generating a cluster plot of the equivalent mean and standard deviations from these optimal parameters. Comparing to Figure 4, we get much better segmentation this way then simply taking the data statistics; although Odden and multiyear are still mixed. We are currently trying this for other models, specifically the modified beta and the more complicated model mentioned above.

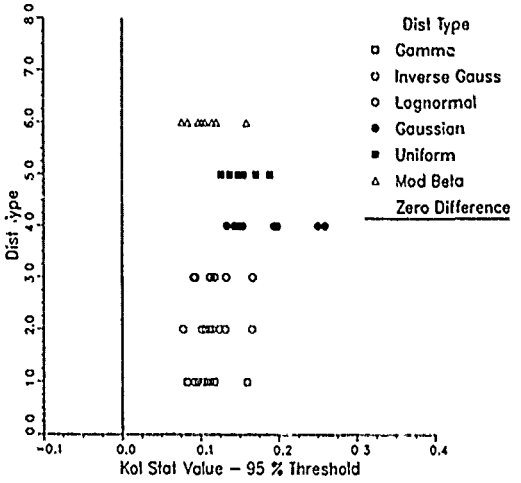


Figure 8. Ice Measurements Kolmogorov Statistics (Open Water)

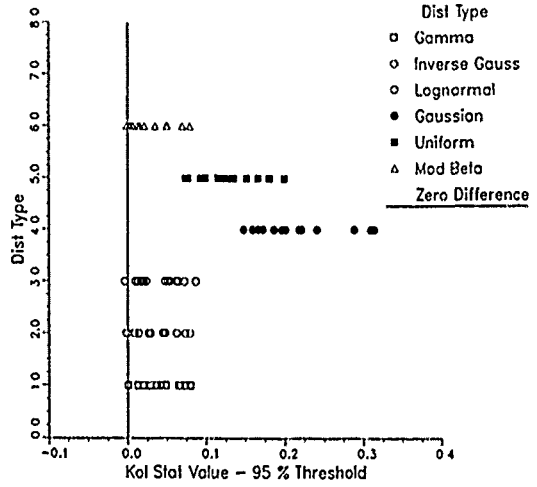


Figure 9. Ice Measurements Kolmogorov Statistics (First-Year Ice)

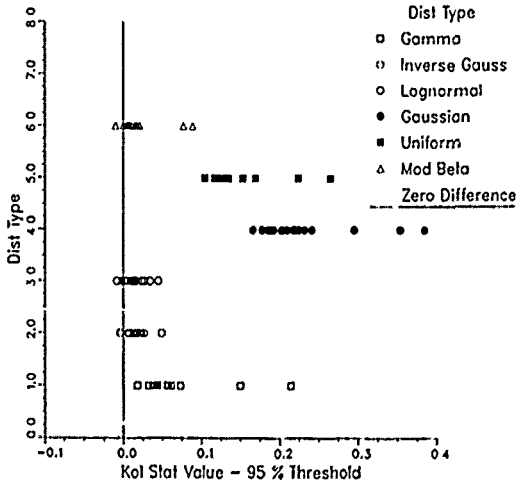


Figure 10. Ice Measurements Kolmogorov Statistics (Multiyear Ice)

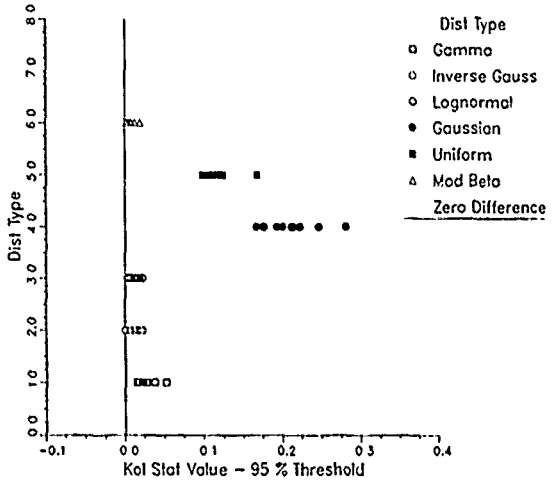


Figure 11. Ice Measurements Kolmogorov Statistics (Odden Ice)

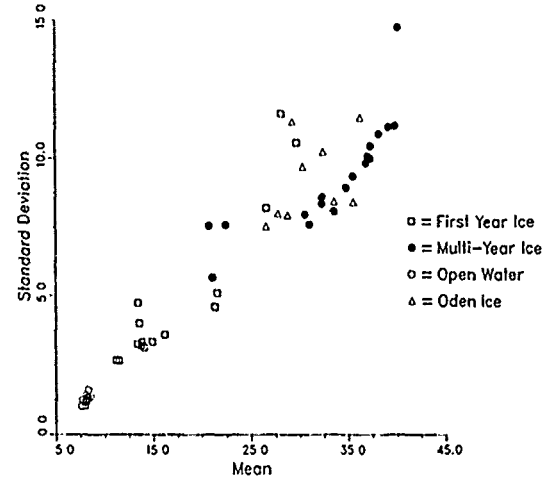


Figure 12. Standard Deviation vs Mean Segmentation From Simple Gamma Model

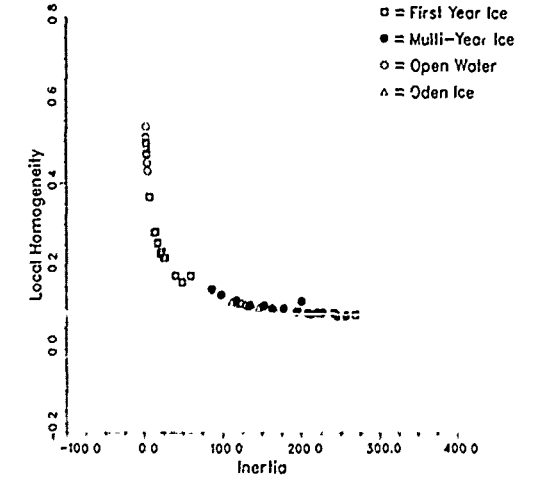


Figure 13. Segmentation Using Texture Measures

5. SEGMENTATION USING TEXTURE MEASURES

Segmentation of the four ice types was also evaluated using textural methods. We used the most favorite texture measure among image processors; the general co-occurency matrix (GCM). In general, the GCM measure the spatial correlation structure of the data and generates estimates of the direction of correlation, the strength of correlation, etc. We again generated cluster plots of pairs of various measures; the best pair is shown in Figure 13. The results appear very similar to the mean and standard deviation plots; open water and first-year are well segmented from the rest, but Odden and multiyear are mixed. We are currently applying the eigenvector analysis to this data also to generate optimal combinations; however we do not anticipate any better results than the first order statistics have given.

6. SUMMARY

The MIZEX '87 X-band SAR data was divided into four sea ice classes (open water, first-year, multiyear, and Odden) to evaluate first-order statistics, higher order statistics (texture), and distribution analysis to segment the SAR image into the required classes. The first order statistical analysis (i.e., statistics based on individual pixel values, not statistical correlations) indicated that mean versus variance can segment out open water and first-year ice, and skewness versus modified skewness can segment out Odden and multiyear ice. The use of texture techniques, in this case the use of the general co-occurency matrix, did not yield results superior to the first older statistical analysis. In general, the first older statistics appear to do a good job and considering computation efficiency is the recommended algorithm approach for the Alaska geophysical processor.

The search for a distribution function of the SAR ice data that will fit the majority of MIZEX '87 imagery is continuing. The distribution analysis suggests the data are best fitted by gamma and lognormal (and sometimes inverse Gaussian) distributions.

7. ACKNOWLEDGEMENTS

This work was supported under Office of Naval Research (ONR) Contracts N00014-81-C-0295 and N00014-88-C-0680. The ONR technical monitors are Mr. Charles A. Luther and Dr. Thomas Curtin. ERIM's ONR contract N00014-88-C-0680 is jointly funded by NASA HQTRS/Ocean Processes Branch. The NASA technical monitor is Dr. Robert Thomas.

8. REFERENCES

1. Sutherland, L.L. and R.A. Shuchman, "SAR and Passive Microwave Observations of the Odden During MIZEX '87", these proceedings.

2. Shuchman, R.A., R.G. Onstott, L.L. Sutherland, and C.C. Wackerman, "Intercomparison of Synthetic and Real-Aperture Radar Observations of Arctic Sea Ice During Winter MIZEX '87, IGARSS '88, Edinburgh, Scotland, September 1988.

3. Shuchman, R.A., L.L. Sutherland, O.M. Johannessen, and E.D. Leavitt, "Geophysical Information on the Winter Marginal Ice Zone Obtained From SAR", IGARSS '88, Edinburgh, Scotland, September 1988.

4. Shuchman, R.A., L.L. Sutherland, B.A. Burns and E.D. Leavitt, "MIZEX 1987 SAR Data Summary" ERIM Report #154600-34-T, 1988.

5. Wackerman, C.C., "Calculation of the Spatial Distribution of Scatterers in a Diffuse Scene from SAR Data", these proceedings.

DETECTION AND CHARACTERIZATION OF ICE RIDGES IN THE BALTIC SEA USING CV-580 SAR IMAGERY

R. JOHANSSON

Dept. of Radio and Space Science
Chalmers University of Technology
Göteborg, Sweden

ABSTRACT

In this paper we report on a study of the azimuth dependence of ice ridges in SAR imagery. The ridge line is divided into two classes, one consisting of ridges with a triangular geometry and one consisting of rubble. It is shown that ridges show a strong azimuth dependence whereas rubble do not. The contrast between ridges viewed broad-side and end-on was determined to 2.5 dB at an incidence angle of 68°. However, the variation is expected to be large and changes with resolution, ridge properties and incidence angle are discussed. Furthermore a ridge detectability index is defined and from this it is concluded that ridges viewed end-on will hardly be detectable in a first-year sea ice background.

Keywords: ice ridges, SAR, azimuth angle, detection.

1. INTRODUCTION

The Baltic Sea, located between Sweden and Finland, is partly ice covered most of the time in the winter season and a costly ice breaker service is required to break the ice and penetrate ridges. In order to assist and guide the ice breakers monitoring of the ice and ridges is very important and one of the most promising tools in this respect is the synthetic aperture radar (SAR). However, few attempts have been made to actually study ridge signatures and to quantify the detectability of ridges from SAR imagery.

One of the main objectives of the Bothnian Experiment in Preparation for ERS-1, BEPERS-88, was to study ice ridge detection for different radar parameters and ridge properties but also to confirm or reject model calculations. In this paper we pay special attention to the azimuth dependence of ice ridges which has been predicted to be strong by Askne and Johansson (1988). A general discussion of ridge detection and modelling is also given and a detectability index defined.

2. THEORY

The detectability of sea-ice ridges is dependent upon the contrast between the ridge and its background as well as the radar resolution. For a multi-look amplitude image we define the detectability as:

$$\text{detectability}=\frac{\overline{DN_r^2}}{\overline{DN_b^2}+2\sigma_{sb}} \qquad (1)$$

where $\overline{DN_r^2}$ is the mean squared ridge DN (Digital Number) value, $\overline{DN_b^2}$ the background mean squared DN value and σ_{sb} the standard deviation of DN_b^2. Furthermore, for a homogeneous background $\sigma_{sb}/\overline{DN_b^2}=k$ is a constant depending upon the number of looks (resolution) and decreases with increasing number of looks (or lower resolution). By chosing the background value plus two standard deviation we ensure to be above 95% of the speckle variation. The radar contrast is defined as $R=\overline{DN_r^2}/\overline{DN_b^2}$ and rewriting Eq. 1 gives:

$$\text{detectability}=\frac{R}{1+k} \qquad (2)$$

We note that the backscatter coefficient is directly proportional to the mean squared DN value minus noise so that the radar contrast can be expressed in terms of the ridge and background backscatter coefficients: $R=\sigma^o_r/\sigma^o_b$. The detection index will vary with the resolution and the ridge and background backscatter coefficients which will be discussed in the next sections.

2.1 Resolution

The resolution effects the detectability in two ways. Firstly, a decrease in resolution, which is the same as increasing the resolution cell area from A_{ridge} to A, averages the ridge backscatter coefficient according to

$$\sigma^o_{ridge,low}=\sigma^o_{ridge,high}\frac{A_{ridge}}{A}+\sigma^o_{bg}\left(1-\frac{A_{ridge}}{A}\right) \qquad (3)$$

The radar contrast then becomes:

$$R_{low\ res.} = R/N + (1 - 1/N) \quad (4)$$

where it is assumed that the ridge pass across the resolution cell and that $N = A/A_{ridge}$. The contrast decreases with N and k with $\sqrt{N}$. Hence, according to Eq.2, the detectability is reduced.

Secondly, ridges closer than a resolution cell will be merged which means that a high resolution is required to detect ridges unambiguously. However, such a high resolution is often limited due to practical considerations such as telemetry rate, data storage and speckle reduction.

2.2 Level ice backscatter

The backscatter coefficient for level sea-ice has been modeled by several authors (Ulaby et al 1982) and show good agreement with measurements. For off-nadir angles the backscatter coefficient is purely non-coherent and only considering surface scattering, which seems reasonable for the conditions met during BEPERS-88, it is a function of frequency, polarization, angle of incidence and the small scale roughness of the surface. Solutions such as the small-perturbation and the Kirchhoff theory gives explicit expressions and shows that the backscatter curve decays with incidence angle and is very sensitive to the surface roughness. Experimental data over thin first-year ice from BEPERS-88 (Ulander 1989) show a decrease of 0.25dB/° between 20° and 60° and 0.5 dB/° between 60° and 70°. As long as the small scale roughness of the surface is isotropic the scattering will also be isotropic and there should be no difference in viewing the level ice from one or the other azimuth angle.

2.3 Ridge backscatter

The backscatter coefficient for ice ridges is far more complicated to model and has been payed little attention during the past years. However, Johansson and Askne (1987) made an attempt to model ridges viewed broad-side by assuming a triangular shape of the ridge see figure 1.

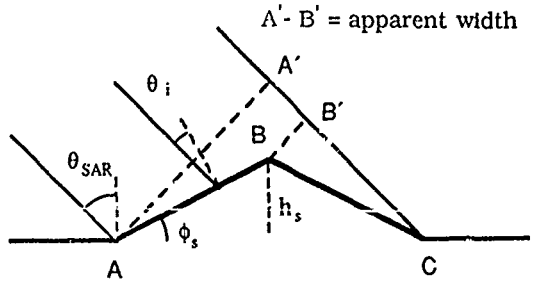


Figure 1. Ridge geometry.

The model includes two scattering components, one diffuse from distributed ice blocks within the ridge and one coherent from specularly oriented blocks which also includes the contribution of blocks oriented as corner reflectors. The model predicts an increase in the backscatter coefficient from the frontside and a decrease from the backside and slower decay with incidence angle than the level ice. Because of the geometry there is a strong azimuth dependence and ridges viewed broad-side should have a higher backscatter coefficient than those viewed end-on. This means that the detectability will increase for an increasing angle of incidence and for ridges viewed broad-side.

The triangular geometry also leads to the definition of the apparent width see figure 1. Ridges viewed broad-side should appear, not only brighter, but also narrower than ridges viewed end-on. However, to study this effect requires that the resolution cell is smaller than the apparent width which was not the case during BEPERS-88.

From inspection of aerial photography it was concluded that along a ridge line ice blocks simply distributed on the ice appeared in places. Hence, two classes were defined, the one considered above with a triangular geometry and a second one being the ice blocks distributed on the ice which we call rubble. In principle rubble could be modeled as above with zero height but with a different and probably more isotropic distribution function of the ice blocks. This leads to the conclusion that there should be no azimuth dependence of the rubble part and the concept of apparent width has no meaning.

3. EXPERIMENT SETUP

The SAR imagery were acquired by the CCRS CV-580 system and several tracks were flown with different frequencies and polarizations. To study the effect of azimuth angle crossing tracks were flown covering the same area. Two images obtained from crossing tracks have been chosen in this study and are shown in Figure 2. The imagery were acquired at C-band with VV polarization and at a nominal slant range resolution of 6m×6m. However, we have observed that the resolution in azimuth rather is 15m. The imagery shown in Figure 2 have been averaged to a resolution of 40*40 m and covers the incidence angle range 10° to 74° degrees. In the figure the line of same incidence angle from the two azimuth angles has been marked and goes across each image.

At the same time as the radar overflight aerial photography were taken at a height of 2000 m and at a scale of 1:13000. The area covered is also indicated in Figure 2. Unfortunately, this area does not perfectly match the constant incidence angle line and falls rather on one side of it.

4. ANALYSIS AND METHODOLOGY

From Figure 2 two full resolution images have been chosen as examples and illustrations of the azimuth angle dependence and are shown in Figures 3 and 4. In Figure 3 the ridge in the middle of the picture is viewed at the same angle of incidence from the two azimuth angles and the azimuth angle effect is indeed strong. In Figure 4 the incidence angle is approximately 8° larger in the left image than in the right one and the effect is again evident. The large ridge seen in the picture is a good example of a ridge line consisting of the two previously defined classes which appear interchangeable.

To quantify the azimuth angle dependence it is necessary to keep fairly close to the crossing line in Figure 2 to eliminate the effect of incidence angle variations. Because of this and the fact that we only have two azimuth angles for each ridge several problems arises in the analysis. Firstly, to keep the contrast between the two directions independent of ridge properties we have to deal with similar ridges in some way. In this paper ridges with approximately the same height and width are considered to be similar Secondly, the ridge may be invisible from one direction or hidden in the background clutter and thus impossible to find. Therefore the analysis will only consider ridges seen from both directions. Thirdly, because of the limited area it is impossible to find enough similar ridges to do a rigorous analysis.

To determine the ridge, rubble and level ice backscatter coefficients, DN values from the imagery have been picked manually to get the statistics. The statistics of the level ice was easily determined whereas it was more complicated to obtain it for ridges and rubble. A border had to be defined within which the DN values should be picked and we chose the mean background level plus two standard deviations.

Once the statistics have been determined an calibration was performed and the level, ridge and rubble backscatter coefficients obtained The accuracy of the relative calibration is estimated to be ±0.6 dB at constant incidence angle and ±2 dB otherwise (Ulander 1989).

5. RESULTS

In Table 1 the azimuth angle dependence for ridges, rubble and thin first-year ice are summarized at an incidence angle of 68°. The result is presented as the magnitude of the contrast for the two azimuth angles 0° and 90° respectively. These were the only azimuth angles where ridges could be identified simultaneously. The detectability index for the two directions are also given.

Table 1 Contrast between azimuth angles 0° and 90° and the detectability index

	$\lvert R_{90°-0°} \rvert$	$D_{90°}$	$D_{0°}$
ridge	2.5 dB	1.4	0.8
rubble	0.6 dB	1.1	1.3
level	1.3 dB		

It is evident from the table that there exist a pretty strong azimuth angle dependence for ridges but not for rubble. The azimuth angle dependence of the background may be due to locally anisotropic effects in the small scale surface roughness and/or calibration uncertainties. We should keep in mind that the ridge contrast value is dependent on the resolution and angle of incidence and is anticipated to increase for higher resolution or larger angle of incidence. Note also that the presented values are mean values and that along a ridge line the variability may be large.

The detectability index is above 1 for all cases but the ridges viewed end on. Even though the index for rubble is smaller than for ridges viewed broad-side, rubble is more easily detected because of its larger width.

In Table 2 the result of ridges viewed from the two azimuth angles but with differing angle of incidence are shown. The ridges viewed broad-side had an incidence angle of 67° and those viewed end on 74.5°.

Table 1 contrast between azimuth angles 0° and 90° with a difference in incidence angle of appr. 8°

	$R_{90°-0°}$	$D_{90°}$	$D_{0°}$
ridge	5.6 dB	1.5	1.1
rubble	1.5 dB	1.5	1.9
level	4.3 dB		

The contrast is much larger than for the previous case where the incidence angle were kept constant. The increase in the azimuth angle dependence may be caused by the increase in incidence angle but may also be due to the fact that we are not looking at the same ridges as before. We observe that the difference of 4.3 dB for the level ice between 67° and 74° gives 0.6 dB/° which agrees pretty well with the value given in section 2 of 0.5 dB/°. The rubble does not show any significant azimuth angle dependence especially remembering the inaccuracy of ±2dB in the calibration procedure.

The detection index show as before that ridges viewed broad-side are much easier to detect but also that the detectability increases with angle of incidence. The rubble does as well show an increase in detectability and assuming no azimuth angle dependence it can be explained by a larger decrease of the background backscatter than the rubble backscatter with incidence angle.

6. CONCLUSIONS

We have shown that there exist an azimuth angle dependence of ice ridges in SAR imagery. As an example the contrast between ridges viewed broad-side and end-on was 2.5 dB at an angle of incidence of 68°. However, this value may change significantly as resolution, ridge properties and incidence angle varies. By defining a detectability index it is shown that ridges viewed end-on will hardly be detectable in a thin first-year background at least for large angles of incidence. It is also concluded that the rubble does not show any significant azimuth angle dependence and that the detectability increases with incidence angle.

ACKNOWLEDGEMENTS

Financial support for this study was provided by the Swedish Board for Space Activities.

REFERENCES

1. Askne, J., Johansson, R., "Ice Ridge Observations by means of SAR", Proceedings of IGARSS'88 Symposium held at Edinburgh, ESA SP-284, pp817-818, 1988.
2. Johansson, R. and Askne, J., "Modelling of radar backscattering from low-salinity ice with ice ridges", Int. J. Remote Sensing, Vol. 8, No.11, pp1667-1677, 1987.
3. Ulaby, F.T., Moore, R.K., Fung, A.K., "Microwave Remote Sensing", Vol.2 (Reading, Massachusetts: Addison-Wesley), 1982.
4. Ulander, L., "Absolute calibration of the CCRS C-band SAR during BEPERS-88", To be presented in Proceedings of IGARSS'89 Symposium held at Vancouver, 1989.

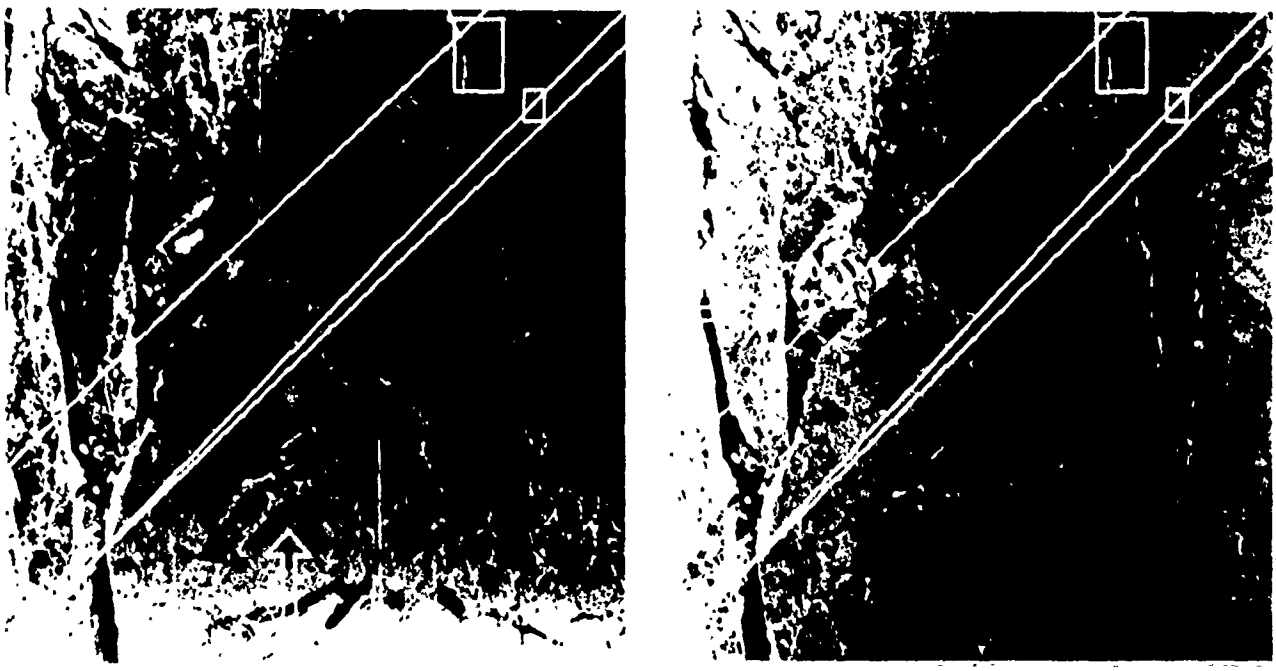


Figure 2. C-band, VV polarized SAR imagery obtained from two crossing tracks.

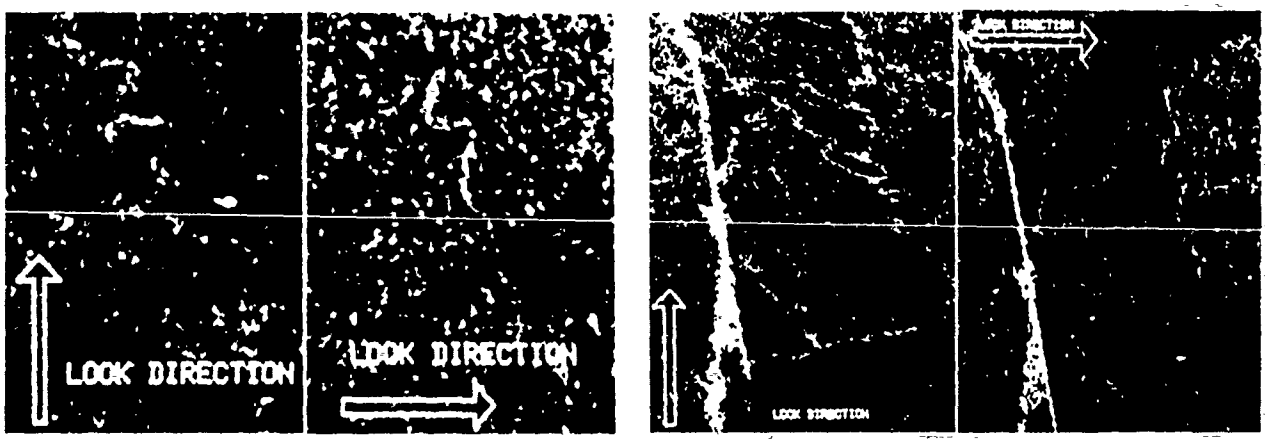


Figure 3 and 4. Subimages illustrating the azimuth angle dependence of ice ridges.

LIMEX'87 ICE SURFACE CHARACTERISTICS AND THEIR EFFECT UPON C-BAND SAR SIGNATURES

Mark R. Drinkwater
Jet Propulsion Laboratory
California Institute of Technology
4800 Oak Grove Drive
Pasadena CA 91109

ABSTRACT

Ice surface characterization data were recorded during March 1987 in the Labrador Sea marginal ice zone, at the onset of spring melt. Measured data are used as input parameters in a simple scattering model to simulate the effects of temporal variations in material properties upon C-band scattering signatures. Snow moisture and surface roughness have a significant effect upon σ^{o}_{HH} and large differences are predicted between undeformed floe surfaces and deformed or rubbled ice areas. The model reproduces a calibrated SAR-derived signature obtained during the experiment with a reasonable degree of certainty. Predictions also simulate a trend observed in SAR images of increasing backscatter contrast between deformed and undeformed ice over period of surface warming.

Keywords: LIMEX, 'early melt', dielectric, autocorrelation function, physical optics model, backscatter signature, radar contrast.

1. INTRODUCTION

The 'early melt' season marks the transition from winter conditions to warmer spring conditions, and the onset of melt is characterized by significant free water in floe surface snow. Backscatter signatures are as a result dominated by liquid water in this surface layer. In March 1987, during the pilot Labrador Ice Margin Experiment (LIMEX'87) [1,2], ice floe surface snow and sea ice were sampled at various locations as ground support data for several C-band synthetic aperture radar (SAR) overflights. These flights were conducted by the Canada Centre for Remote Sensing (CCRS) Convair-580 over the experiment area off the east Newfoundland coast. Measurements of snow and ice properties enabled temporal records of surface roughness, salinity, wetness, depth and density, and photographs of microscopic snow structure to be obtained during the onset of melt.

Other than during the Marginal Ice Zone Experiment [3], few C-band SAR data have been acquired in the marginal ice zone during spring melt. Experience gained from other experiments has been utilised to a degree in development of models to simulate microwave signatures of combined media [4,5,6]. Sea ice and 5.3 GHz SAR data provided an opportunity to apply simple theoretical models to marginal ice examined in the Grand Banks region of the Labrador Sea under warming conditions. Models are driven using physical properties measured *in situ* and are used to explain the range of scattering signatures of targets observed in the C-band radar images.

2. SURFACE MEASUREMENTS

Ice conditions off the east coast of Newfoundland were monitored throughout LIMEX'87, between 15 and 26 March 1987 [1,2]. Detailed periodic measurements were made, often concurrently with remote sensing overflights, in order to characterize surface conditions and identify changes in material properties. Geographic locations of experiment surface sites are detailed in Fig. 1, along with the position of the ice edge.

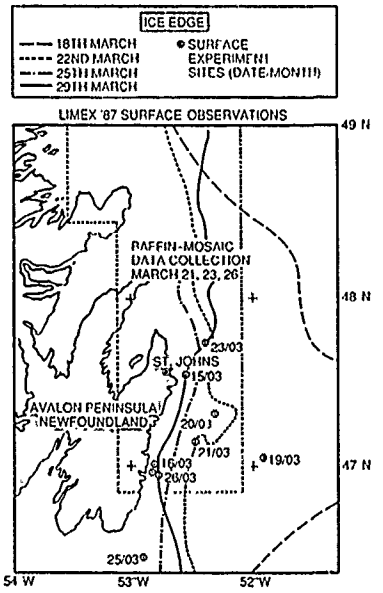


Figure 1. LIMEX '87 surface experiment sites and extent of the marginal ice zone. The dotted rectangle delineates the area within which coincident aircraft SAR coverage was obtained in the CSS Baffin locality.

Parameters recorded in the field by the surface data team included snow and ice depths and salinity profiles, and snow density, wetness, and temperature. Surface roughness was recorded with the aid of a profiling instrument, along with macro-photographs of snow grain morphology (x10 magnification). Local meteorological conditions were monitored at St. John's, and by an automated weather station located on the vessel CSS *Baffin*. Melting conditions are induced by warm moist south-easterly winds while sub-zero refreezing conditions occur during bouts of predominantly north-easterlies.

Fig. 2 summarises ice surface conditions measured on several days, and documents changes occurring over the experiment period between 16 and 26 March. Surface snow decreased rapidly in depth, and snow crystal macro-photographs illustrate that the collective processes of destructive metamorphism changed precipitated snow into a variety of stages of firn. The processes of equi-temperature and melt-freeze metamorphism caused crystals to become rounded, forming clusters of touching spheres. A continuum of states was observed, between the beginning of snow metamorphosis and the formation of higher density polycrystalline ice. By 26 March, the top few centimetres of the surface had become a continuous layer of dense firn (or snow ice).

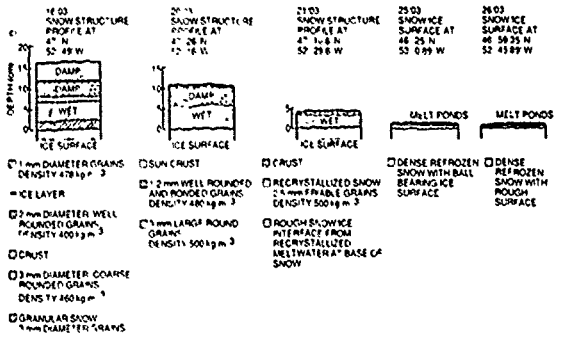


Figure 2. Typical ice surface conditions during the LIMEX '87 experiment period, in terms of the snow structure, crystal morphology, density and wetness.

3. ICE SURFACE MELT AND DESALINATION

Surface characterisation data indicate that melt-freeze cycles and melt drainage reduced the firn to near zero depth in most locations. Ice cores extracted in several locations on 23 and 26 March indicate that significant desalination had occurred in the upper ice layer (0-30cm) probably due to brine drainage channel development and flushing by snow melt water. Upper 10cm salinities varied from 0 - 2.0‰. On 26 March, a strong salinity gradient was observed between 25 and 35cm, and a maximum salinity of 8‰ was measured at 35cm depth in one core. Below 35cm the mean salinity remains reasonably constant in both profiles, remaining between 3 and 5‰ down to 115cm.

Flushing occurs under gravity drainage as the pressure head of surface meltwater overcomes capillary retention of brine in pockets within the ice. Providing the ice is permeable, meltwater is able to drain downwards. Mean ice thicknesses in the core localities were 2.7 and 1.6m, on 23 and 26 March respectively, indicating that rafting was of particular importance in augmenting ice thickness and initiating desalination by uplifting and increasing the freeboard of the ice and the resulting pressure head. This factor explains lower bulk salinities observed on 23 March.

4. SURFACE DIELECTRIC PROPERTIES

4.1 Wet snow

Wet snow permittivity (ε') and dielectric loss (ε") are frequency dependent, showing large increases with liquid water content in the GHz range. At microwave frequencies close to the relaxation frequency of water, such as 5.3 GHz, ε' and ε" become almost independent of the snow structure, and water dominates the behaviour of the complex dielectric constant ε^* (where $\varepsilon^* = \varepsilon' - j\varepsilon''$).

A dielectric mixture formula is used to model a combination of air and water between ice particles, and water inclusions are assumed to have a shape between randomly oriented needle-like shapes and spheroids. Fig. 3 shows the resulting relationship between ε' and ε" and volume fraction of water W_v. The value ε" tends to zero for snow with no water, but increases rapidly as water is included. ε' responds, increasing by 1.9 with the addition of 15% water by volume fraction.

Fig. 3 applies to old coarse snow in the 'pendular regime' (ie. less than 14% of pore volume filled by water). This state occurs when liquid water is held at the necks between snow grain boundaries. Variability in ε'_{wet} and ε''_{wet} increases when snow crosses the transition from the pendular to funicular regime (where drainage begins), and so these two regimes are delineated by shading for ε'_{wet} at each density. For snow of 400 kg m^{-3}, in transition between the two regimes, ε' and ε" are estimated at 2.65 and 0.5.

Snow wetness calculations were made by NORDA on 20 March using the dye dilution technique. Measurements obtained (at temperatures close to freezing) from the damp surface snow layer (Fig. 2) gave an estimate of a liquid mass volume fraction of about 0.22. When corrected for locally measured snow densities ranging between 400 and 500kg m^{-3}, these data indicate an approximate range of W_v

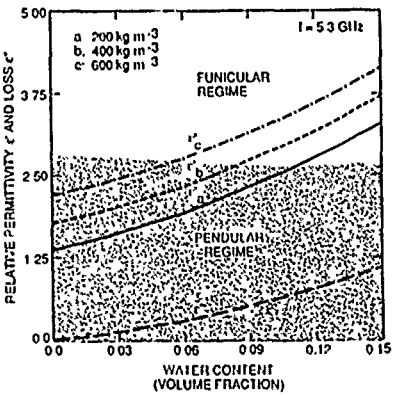


Figure 3. The dielectric properties of wet snow at a frequency of 5.3 GHz. The permittivity and dielectric loss are plotted for varying snow water content and densities of 200, 400, and 600kg m^{-3}. The shaded region indicates the values of ε' for snow of given density and wetness in the pendular regime.

between 0.08 and 0.11 (ie. 8 to 11% by volume). Thus, on 20 March the snow was in transition between pendular and funicular regimes.

4.2 Labrador Sea Ice

As the complex dielectric constant ε^* is closely linked to the volume fraction of liquid brine, it is important to know the effects of desalination upon the dielectric properties of the ice medium. As low salinity ice is warmed to temperatures between -1 and 0°C, the volume fraction of brine in the ice rapidly reaches a maximum. For ice in the salinity range 0.7-0.9‰, ε' and ε" increase from levels more typically associated with pure ice ($\varepsilon^* = 3.1 - j\,0.01$) at -2°C to values around $\varepsilon^* = 4.3 - j1.0$ at a temperature of -0.2°C [7].

4.3 C-band Penetration Depth

Snow liquid water content increases loss, making the underlying ice less significant in the backscattering process. With the addition of 10% water by volume fraction, the penetration depth δ_p is reduced from the order of several metres to between 0.02 and 0.03m, as in [3]. Late in the experiment when the firn is shallow and drained, or completely melted, the underlying low salinity ice surface is able to contribute in the backscattering process. Throughout LIMEX'87 the upper ice interface temperature was never colder than -3°C, remaining close to 0°C throughout the experiment. δ_p tends rapidly to zero for ice of any salinity in the temperature range -5 to 0°C and consequently volume scattering within the sea ice may effectively be disregarded.

5 SURFACE ROUGHNESS STATISTICS

Several roughness profiles were measured on snow and ice surfaces using a technique to determine the height of discrete elements along a transect. Sampling was undertaken at 1cm horizontal intervals (Δx) in 1m sections, and statistics derived using this technique have been analysed numerically. The degree of roughness of these surface sites is quantified in terms of the standard deviation σ of surface height elements (or rms height), the auto-correlation function-derived correlation length l and the rms slope s.

Statistics from a 22m long co-registered transect measured on 20 March indicate height elements normally distributed around the mean, with Gaussian distribution. The standard deviation σ=2.72cm, the rms slope s =0.33rad (~19°), and the form of the distribution and high rms slope are influenced by the overall degree of deformation of this marginal ice. Rubbled floe edges, and rafting caused by extreme pressure conditions, are responsible for the undulations of the order of 10 or 15cm amplitude. It is evident from photography and the properties and statistics extracted from surface profiles that the MIZ was composed largely of rubbled or deformed ice separated by areas of undeformed, relatively flat, desalinated first year ice.

Fig. 4 illustrates autocorrelation statistics most representative of the two distinct roughness classes identified in field data. Solid curves in

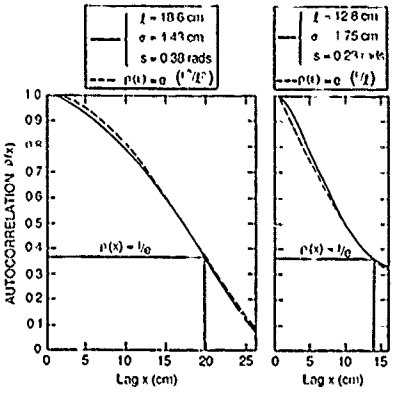


Figure 4. Autocorrelation functions for ice in a) deformed areas, and b) undeformed, areas Solid lines indicate measured curves along with their respective surface correlation length (l), standard deviation (σ) of surface roughness, and rms slope (s). Dashed lines indicate closely corresponding curves for ideal surfaces with a) Gaussian, and b) Exponential autocorrelation as functions of height (ε) and l.

Fig. 4a and 4b are representative of the autocorrelation functions of deformed, rafted or rubbled areas, and flat floe surfaces, respectively. The maximum observed rms slope is 0.5rad, and, because measurements were biased towards the lower end of the roughness spectrum, it is suggested that locally rough areas such as pressure ridge ice piles had values as great as s =0.8 rad (or 45°).

6. SCATTERING MODELS FOR A COMPOSITE MARGINAL ICE ZONE

From examination of the roughness statistics, it appears that no single scattering approximation can be used to simulate backscattering from this composite MIZ surface. Various scattering formulations can be used to predict the backscatter response in given circumstances. Plotted in Fig. 4a and 4b are dashed Gaussian and Exponential autocorrelation functions, which indicate that the solid curves can be represented by these simple functions to a first approximation. Providing criteria for applicability are met, surface scattering from these two distinct surface classes may be predicted using simple Physical optics formulations [8, 9] with similar approximations.

6.1 Surface Scattering

Physical optics formulations are widely used in modelling surface backscatter; they are applicable to undulating surfaces whose roughness is large with respect to the incident wavelength. Given the statistics in Fig. 4a and b for rough and relatively smooth surfaces, the criteria for its application are met. In these cases, the Kirchhoff surface integral is further simplified using two approximations, one for surfaces with $s < 0.25$rad and the other for rougher surfaces with $s > 0.25$rad. The former is a 'scalar approximation' and is applicable when σ is a fraction of λ, as is the case of the surface segment used for Fig. 4b. The alternative for a rough surface, described by the autocorrelation function in Fig. 4a, is a simple geometric optics formulation with a 'stationary-phase approximation'.

A scalar approximation is used to calculate the backscatter coefficient (σ^o) of undeformed ice. Previously, this technique has been used to reproduce surface scattering signatures of first year ice with an Exponential autocorrelation function similar to Fig. 4b [5, 10] and the formulae are not given here. For rough or deformed ice surfaces, a geometric optics model is more appropriate. Similarly, this formulation is widely used and discussed in [9, 10] for a Gaussian autocorrelation function.

6.2 Volume Scattering

Volume scattering from the snow layer depends primarily upon the amount of energy transmitted into the medium, and the effective value of δ_p. Even for wet snow, over 90% of the incident energy is transmitted across a smooth air/snow interface (when $\Gamma_{HH}(0) < 0.28$). At the start of the experiment, snow wetness values estimated at W_v=0.05 or less mean δ_p is sufficiently large that several centimetres of the snow volume is energised. Volume scatter takes place from ice grains and free water inclusions, and the intensity of backscatter is governed by their size and dielectric properties, and the extinction within the layer. The total volume scattering coefficient is estimated by using the Rayleigh approximation, and the formulae are outlined in [6, 10].

6.3 Combined Backscatter Coefficient

Both the surface backscatter σ^o_s and the volume backscatter σ^o_v are combined to calculate the total backscatter coefficient from ice floe surfaces. Clearly, calculations of δ_p for wet snow indicate that a negligible portion of energy penetrates deep enough for the sea ice medium beneath to contribute to overall backscatter totals. Additional attenuation within the wet snow layer means that sea ice volume backscatter can effectively be ignored under these circumstances.

7. COMBINED MODEL RESULTS

7.1 Undeformed Ice Floe Surfaces

Five sets of field parameters are used to drive the model, giving a scattering signature for each surface described in Fig. 2. Predicted signatures in Fig. 5, for floes without changes in surface roughness, indicate that as snow density and volume fraction of liquid water increase, the gradient of σ^o in the incidence range $40° < \theta < 80°$ increases. The level of the curves does not vary by more than 3dB, yet it is evident from radar images that certain floes appear much darker than their surrounding counterparts [11] taking on an image intensity almost as low as open water areas. None of the combinations of surface parameters account for standing water or a saturated upper layer upon ice floes (Fig. 2) and an additional curve is plotted in Fig. 5 to represent saturated floes observed on 25 and 26 March. This signature, in contrast, shows a marked reduction in σ^o throughout the

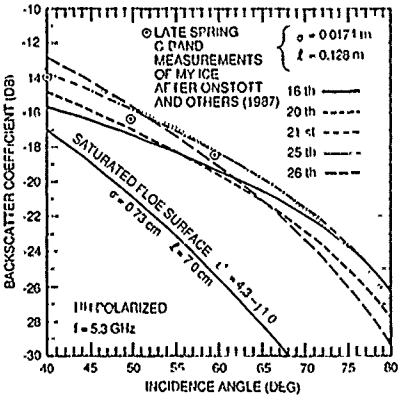


Figure 5. Simulated C-band HH polarized scattering signatures for each of the five days, using parameters listed in Figure 2 and roughness statistics for an undeformed exponential surface Scatterometer data points recorded over similar low salinity multiyear (MY) ice are indicated, and a signature predicted for smooth saturated floe surfaces is included for comparison.

incidence angle range, of between 4 and 12 dB. It demonstrates that flooded floes have relatively lower backscatter values over the whole angular range, with maximum contrast at the largest incidence angles.

7.2 Deformed Ice

A C-band scattering signature from 13 March was obtained prior to surface observations recorded in Fig. 2. It is a mean calibrated scattering signature from a deformed ice region. A model curve is fitted to these data in Fig. 6 and the snow parameters necessary to generate this inverse model response recorded. The curve fit is excellent, given actual surface conditions for that period. A 16cm deep, low density snow layer (ρ=300 kgm^{-3}) with small grains (r=0.8mm radius), and low volume fraction of liquid water (W_v=3%) are predicted, and are realistic given that cold weather and fresh snowfall preceded 16 March measurements in Fig. 2. The lack of change in snow properties over the three days is consistent with temperatures predominantly below zero between 13 and 15 March,

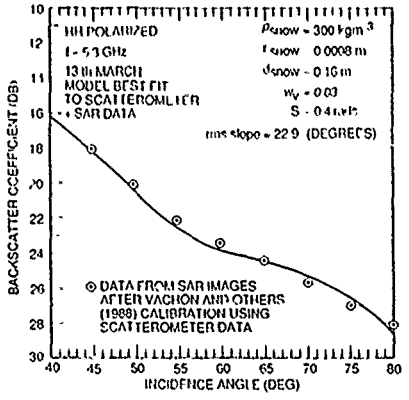


Figure 6. Modelling fit to calibrated SAR data points extracted from a LIMEX '87 SAR image acquired on 13 March, with snow parameters indicated.

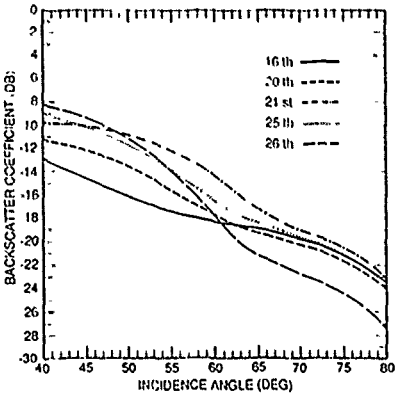


Figure 7. Simulated C-band HH polarized scattering signatures for a deformed, Gaussian distributed sea ice surface on each of the five days, using parameters listed in Figure 2 and recorded roughness statistics.

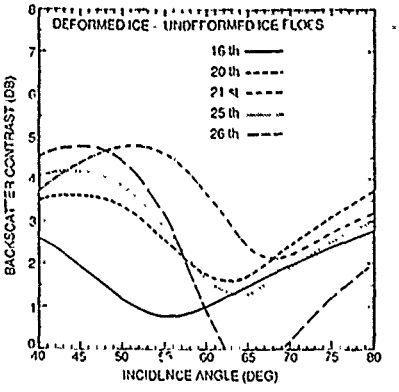


Figure 8. Predicted radar backscatter contrast between deformed and undeformed ice signatures on each of the experiment days for Gaussian and exponential surface correlation assumptions, respectively.

although some metamorphosis and increase in density is natural. A prediction of s =0.4 rad (~23°) is also consistent with surface roughness conditions plotted in Fig. 4a.

Observations of surface roughness were limited only to areas of shipborne surface activity and the most deformed regions of ice apparent on SAR images were not sampled. The MIZ suffered a severe ice compression event between 16 and 21 March. Rapid compaction took place, resulting in a high degree of uplifting and pulverisation of ice floes under the extreme pressure. Limited samples from uplifted and rafted blocks of ice suggested that they are better drained and more desalinated than floe surfaces. Such materials are more efficient volume scatterers than wet snow-covered floe surfaces and this factor is expected to be the major cause of increasing σ^o values. Tilted blocks and floe edges exposed by rafting and pressure ridging contribute to higher rms slopes, and values of $0.4 < s < 0.6$ are common. By 25 March, under the influence of swell penetration, only an extremely small proportion of the ice floes remained undeformed. Winds also changed, allowing the ice pressure to be relieved, and as waves fractured the ice into a discrete pack the discontinuous floe field no longer supported large piles of ice in rubble or ridges.

Fig. 7 shows the results if W_v in ridged ice is consistently lower by 0.02-0.04, and when s increases from 0.4 to 0.6rad between 16 and 21 March (during ice compaction) and falling to 0.5rad by 25 March. Using this approach, it is predicted that the highest σ^o values occur on the 21st March. Only on the 25 and 26 March does σ^o exceed that on the 21st, between 40 and 50° incidence, due to increasing water content and snow density, and reduction in snow depth.

Ice pressure conditions reached a peak on 21st March and regions of deformed ice in SAR images are expected to increase in brightness over the preceding days. Digby-Argus and Carsey [11] extracted DN (digital number) values from images of regions of deformed and undeformed ice. They confirm that the radar contrast between floes and their surrounding material does increase between 16 and 21 March. In Fig. 8 the relative contrast between deformed and undeformed ice has been predicted using the pairs of curves for each date in Figs. 5 and 7, to see if the trend is reproduced. Fig. 8 shows that the mean angular contrast does in fact reach a maximum on 21 March of 3.5dB. Peak radar contrast is approximately 5dB and occurs for $45° < \theta < 60°$. The predicted contrast between the deformed ice on 21 March and the saturated surface in Fig. 5 is also calculated for comparison ranging from 7dB at 40° to 19dB at 80°.

8. DISCUSSION

LIMEX '87 saw a range of surface melt conditions which lend themselves well to the use of this simplified backscatter model. The main factors modulating scattering characteristics are snow metamorphosis, surface wetness and large scale surface roughness. The result is a high degree of contrast between smooth or wet floes and their rough deformed ice surroundings, which is readily observed in the SAR images. The model predicts contrasts in σ^o between specific ground targets in order to make comparisons with image properties. Results indicate the ability to identify areas of thick, deformed, and thin undeformed ice at a time of year when signature variations due to ice typology are suppressed by snow moisture.

9. ACKNOWLEDGEMENT

The author thanks all LIMEX '87 participants who have freely circulated data, and in particular CCRS who supplied images. This work was performed at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration, while funded by a National Research Council Research Associateship.

10. REFERENCES

[1] McNutt, L., S. Argus, F. Carsey, B. Holt, J. Crawford, C. Tang, A.L. Gray, and C. Livingstone, "LIMEX '87: The Labrador Sea Ice Margin Experiment, March 1987 - A pilot experiment in anticipation of RADARSAT and ERS 1 data," *EOS* vol. 69, no. 23, p. 634, 1988.

[2] Drinkwater, M.R., and S. Digby-Argus, "LIMEX '87: An international experiment in the Labrador sea marginal ice zone," *Polar Record*, in press.

[3] Onstott, R.G., T.C. Grenfell, C. Matzler, C.A. Luther, and E.A. Svendsen, "Evolution of microwave sea ice signatures during early summer and midsummer in the marginal ice zone," *J. Geophys. Res.*, Vol. 92, no. C7, pp. 6825-6835, 1987.

[4] Fung, A.K., and H.J. Eom, "Application of a combined rough surface and volume scattering theory to sea ice and snow backscatter," *IEEE Trans. Geosci. and Remote Sensing*, vol. GE-20, no. 4, pp. 528-536, 1982.

[5] Kim, Y.S., R.K. Moore, R.G. Onstott, and S. Gogineni, "Towards identification of optimum radar parameters for sea ice monitoring," *J Glaciology*, vol. 31, no. 109, pp. 214-219, 1985.

[6] Drinkwater, M.R., and G.B. Crocker, "Modelling changes in the dielectric and scattering properties of young snow-covered sea ice at GHz frequencies," *J Glaciology*, vol 34, no 118, pp. 274-282, 1988.

[7] Hallikainen, M.T., M.V.O. Toikka, J.M. Hyyppä, "Microwave dielectric properties of low-salinity sea ice," *Proc. IGARRS '88 Symp.* (Edinburgh, 13-16 Sept, 1988), pp. 419-420, 1988.

[8] Eom, H.J., "Theoretical scatter and emission models for microwave remote sensing," Ph.D. thesis, University of Kansas, 1982, 212pp, 1982.

[9] Barrick, D.E. "Rough surface scattering based on the specular point theory," *IEEE Trans. Antennas Propagat.*, vol. AP-16, no. 4, pp. 449-454, 1968.

[10] Drinkwater, M.R., "LIMEX'87 ice surface characteristics: implications for C-band SAR backscatter signatures," accepted for publication IEEE Trans. Geosci. and Remote Sens, Sept. 1989.

[11] Digby-Argus, S., and F.D. Carsey, "SAR imagery of the Grand Banks ice pack and its relationship to surface features," *Proc. IGARSS '88 Symp.* (Edinburgh, 12-16 Sept. 1988), pp. 1425-1428, 1988.

EXTRACTION OF SEA ICE INFORMATION FROM SAR IMAGERY

Tor I. Olaussen, Ola M. Johannessen, Ridvan Karpuz

Nansen Remote Sensing Center
Edv. Griegsvei 3a, N-5037 Solheimsvik, Bergen, Norway
Telephone: 47-5297288

1. ABSTRACT

Textural and contextual image processing algorithms have been used to automatically estimate sea ice concentration, floe statistics and ice type classification from Synthetic Aperture Radar (SAR) imagery. In order to prevent speckle induced noisy classes and concentration estimates, SAR images were filtered using a median filter which preserves edges while reducing high frequency speckle noise. Areas covering different ice types in the imagery were analysed digitally to obtain information about their textural characteristics. Classification methods used in the study are based on segmentation of images, supervised and unsupervised classification algorithms. Ice floe shape descriptions and floe statistics were extracted using binary operations. Using texture information in SAR images provided a better understanding of lead properties and their spatial distributions.

The result of this work is a generation of thematic sea-ice maps and quantitative sea-ice information.

Key Words:

SAR, Texture, Segmentation, Classification, Speckle.

2. INTRODUCTION

Understanding the characteristics of sea ice is of vital importance for the Arctic regions both for detecting climatic changes caused by variations in spatial and temporal distribution of sea ice and solving operational and logistic problems of industrial actvities in the region. Airborne SAR with its weather independent operational capability provides data for studying sea ice. Therefore much effort has been put into developing sea-ice classification algorithms using satellite and airborne SAR data. However, one of the major problems, faced in these studies is associated with identifying representative signatures of different ice types in non-homogeneous images. Classification algorithms solely based on tonal-grey level pixel statistics do not provide satisfactory results in an image consisting of strongly deformed first year ice and multi year ice. Both ice types give very similar backscatter values in this situation because of the SAR's sensitivity to surface rougness characteristics. Therefore a need arises to utilize not only grey level differences but also textural features for sea ice classification as well. Wackerman et al. (1988) also points out possible complications in sea ice type classification caused by seasonal differences during data acquisition. A fully automatic classification algorithm developed by Wackerman et al. (1988) is based on multiple segmentation of images containing multiple ice types without any in-situ data input. Holt et al. (1988) employed a Bayesian maximum likelihood algorithm to classify the seasonal sea ice SAR imagery off Labrador. The classification algorithm used in their work employs tonal and textural information from SAR. A generalized flow diagram of the sea-ice information algorithms developed in our study is shown in figure 1.

3. REMOTE SENSING DATA ANALYSIS

SAR images used in this study were acquired during the Marginal Ice Zone Experiment (MIZEX'87) in the Greenland Sea in March-April 1987 (MIZEX Group, 1989). During the experiment both remote sensing and in-situ data were collected. Daily SAR coverage was obtained from an aircraft while in-situ measurements were carried out using the research vessel "Polar Circle" and helicopters. The SAR data were recorded in digital form as a 7-look imagery by an X-band SAR system with a resolution of 15m. Later processing of the data was performed at the Nansen Remote Sensing Center (NRSC) using a ContextVision GOP-300 image processing system. This included geometric correction of the SAR data with slant to ground range conversion algorithms. The sub-scenes used in this study were extracted from the representative locations from the image acquired on April 6 1987.

3.1 SAR Post-Processing and Speckle Reduction

Speckles in SAR data prevent discriminating image characteristics due to their so called high frequency multiplicative noise originating from the coherent registration of pixels by SAR sensors. More detailed description of speckle properties of SAR imagery can be found in Elachi (1987). In order to use SAR images for classification, the speckles

should be reduced prior to classification operations. This is of vital importance if SAR images contain high frequency textural information other than speckles such as structural details of congealed ice floes consisting of various ice types with rough surface topography due to ridging and shearing processes. These areas are characterized by a variance representing both the speckle and texture variables. The speckles can be reduced to some extent during SAR processing with averaging the independent number of looks (N) recorded from the same scene at the cost of reduced resolution. When linear detection is used in a SAR sensor, the fading variance (due to speckles) can be reduced by a factor of N as it is seen from the following formula (Ulaby et al., 1986):

$$SD/M = (0.273/N)^{1/2}$$

where SD is the standard deviation of a homogeneous area, M is the mean intensity of a homogenous area, and N is number of looks. The mean value calculated from a homogeneous field can be used to find out the variance due to speckle (fading variance).

Another approach is to filter SAR images with filters in an image processing system. To date numerous filtering algorithms have been tested on SAR images Frost (1981), Lee (1981), Li (1988), Pratt (1975). The present trend is to develop texture preserving adaptive filters. A simple and comparatively efficient way to reduce speckles while preserving edges is the median filtering technique. This filter preserves structural detail in images while reducing the speckle noise. One should be very careful with the selection of kernel size during filtering. Fig. 2 shows our test results from median filtering of the images indicating that increased kernel size from 3x3 pixels to 7x7 pixels result in considerable degradation of fine textural details exhibited over the area covered with first year and new ice. Li (1988) solved this problem with implementing a moving average filter and a combined moving average and median filter where local variance ratio is used to control window size. If the images contain only low spatial variations in texture, it is more suitable to use a larger kernel size .

3.2 Texture Signature of Ice-types and their implications on classification.

Physical properties, surface geometry and surface roughness of ice types govern backscattering coefficient values recorded by SAR sensors. The most pronounced backscattering differences result from surface roughness properties of ice. This is a pitfall when SAR sea ice images are interpreted solely on the bases of tonal variations. This can be clearly observed i. fig. 3 where both multi year ice (MY) and sheared first year ice (FY) floe boundaries are characterized by bright tones. Fragmented floe parts, rubble, strongly up turned pancake edges can be falsely classified as multi year ice floe fragments because of the high backscattering coefficient resulting from their increased surface roughness. Another example is seen in fig. 4 where swell propogation through the ice results in systematic variations in backscattering . Bright linear features in the swell region are more likely associated with broken-up floes with increased surface roughness regardless of their ice type.

3.3 Classification Algorithms

a) Segmentation: Image segmentation is particularly found to be suitable to stretch SAR images when the image histogram is unimodal. Our segmentation algorithm is based on iteration technique. The speckle reduced SAR image is further filtered with low pass filters in order to enhance subtle variations in the image histogram (Fig. 5). As it is seen from the final image histogram of the segmented image, the unimodal histogram is now converted to a multimodal histogram with four major tops. The image can be segmented into water and ice classes which otherwise was impossible using the unimodal distribution. Statistical information extracted from the homogeneuos marked area shows that the standard deviation was reduced from 4.4 to 1.1 after the segmentation process was completed. This is due to reduction of in-class variations during segmentation. On the other hand overall standard deviation is increased from 20.2 to 46.3 as a consequence of segmentation, reflecting a more suitable class discrimination due to enhanced subtle differences between the ice types. An edge based segmentation algorithm was used to define ice floe boundaries.

b) Supervised Algorithms: Supervised classification algorithms uses in-situ data in order to classify images. For more detail information, see the paper "Speckle Reduction and Maximumlikelyhood Classification of SAR images from Sea Ice Recorded During MIZEX'87" (Hansson, Tjelmeland, Johannessen, Olaussen, Karpuz, 1989).

c) Unsupervised Algorithms: The 'clustering' classification method has been used to estimate distribution of ice types where in-situ data are not available. The algorithm is based on the Minimum Distance classification procedure (Mather 1987). The results of this algorithm are suitable to obtain a rough estimation of ice classes where in-situ data are not available. It is also used prior to running a supervised algorithm in order to gain beforehand information on ice types.

4. QUANTITATIVE ICE-FLOE INFORMATION

Quantitative ice information is obtained through a series of binary operations performed on segmented images as well as from classified images. We used a series of neighbourhood operations to remove defects from the floe boundaries before applying statistical information extraction routines. This is accomplished using shrinking-expanding and filling operations. We employed feature detection operations to extract information on floes. Feature detection operations were performed in order to define each floe in terms of the following parameters; floe perimeter, diameter, convex perimeter, circularity, compactness, area, width and length.

4.1 Lead Statistics and Orientation

Leads are studied employing various enhancement operations. Their areal coverage and geometric characteristics are automatically obtained with binary feature detection algorithms. Edges adjacent to leads and ice gives a very high variance value when the percentage of water is higher compare to leads covered with new ice. This is achieved using a filter which calculates the local variance for each neighbourhood in the image. The higher the variance, the higher the brightness on edges separating leads from ice floes.

4.2 Ice Concentration

Total ice concentration is automatically obtained from segmented images with setting a single threshold to the segmented image. When this operation carried out through the whole imagery, the results can be displayed as contoured ice concentration values. Multiple thresholding is also used to extract the concentration of different ice types. When SAR images classified with either supervised or unsupervised classification, areal coverage of ice types are automatically found from the number of classified pixels of each class.

5. CONCLUSION

1) Selection of a proper speckle reduction algorithm is crucial in SAR processing. Important failures may occur in classification results due to either too large or too small kernel size.

2) Some areas in SAR images may exhibit the same backscatter value regardless of ice type. This is clearly evident in deformed regions where floes adjacent to the shear/deformation zones display lower backscatter values compare to their fragmented, broken-up parts with higher radar returns because of their increased surface roughness. Similar phenomena can be seen along some ice-edges where very bright signatures of ice bands are characterized with small fragmented floes consisting of different ice types. If these problems are not recognised prior to classification operations, the results obtained from classification will not be representative of true ice classes but only spectral classes.

3) Segmentation is the simplest and most efficient way to map different ice types when using one frequency imagery. Iterative low-pass filtering enhances gross variations in backscatter while supressing high frequency variations due to ice structures, swell propogation, etc. Structural and textural details can be preserved using adaptive filters and enhancement algorithms. Therefore analysis of SAR images should contain both image components.

4) Supervised (maximum likelyhood, mean distance) and unsupervised (clustering) classification algorithms have proven useful when suitable texture images were incorporated in the classification algorithms. Seasonal differences should be considered carefully when classifying images. Ablational processes occuring on ice types reduce the backscatter coefficient. Characteristics of snow cover, especially interbedded ice layers within a snow pack overlying the floes, sea water flooded ice floes, wind drag over open water areas create ambiguities in classification results. These may be avoided with a contextual approach to classification. More work remains to be done in this field. When available, coincident multi-sensor approach should be used in conjunction with SAR images to correlate ice signatures observed on SAR with other sensor data.

5) Floe characterization and estimate of total ice concentration is done on segmented images with binary operations. These operations allow to define each floe with its shape and size.

6) Leads are analyzed using image enhancement techniques. Their areal distribution is calculated with binary operations. Orientation data is statistically represented as histograms and rose diagrams.

Acknowledgement

This work was supported by the Office of Naval Research and Geophysical Institute at the University of Bergen. The SAR images were provided by R. A Shuchman at ERIM.

REFERENCES

1. Wackerman C., C., Jentz R., R., "Sea Ice Classification of SAR Imagery", Proceedings of IGARSS '88 Symposium, Edinburg, Scotland, p. 425-427, 1988.

2. Holt B., Lee M., Yang W., " Mapping of Seasonal Sea Ice Off Labrador with C-Band SAR Imagery", Proceedings of IGARSS '88 Symposium, Edinburg, Scotland, p. 1423, 1988.

3. MIZEX Group, "MIZEX East 1987: The Winter Marginal Ice Zone Program in the Fram Strait/ Greenland Sea", to be pulished in EOS, 1989.

4. Elachi C., "Introduction to the Physics and Techniques of Remote Sensing", p.196-205, 1987.

5. Ulaby F. T., Moore R. K., Fung A. K., "Microwave Remote Sensing", Vol. 3, p. 1919-1920, 1986.

6. Frost, V.S., J.A. Stilles, K.S. Shanmugan, and J.C. Holtzman, "Model for Radar Images and Its Applicatioon to Adaptive Digital Filtering of Multiplicative Noise", IEEE Trans. Patt. Anal. Mach. Intell., Vol. PAMI-4, p. 157-166, 1982.

7. Lee, J. S., "Speckle Analysis and Smoothing of Synthetic Aperture Radar Images", Comp. Graphics and Image Proc., Vol. 17, p. 85-89, 1981.

8. Li, C., "Two Adaptive Filters for Speckle Reduction in SAR Images by Using the Variance Ratios", Int. J. Remote Sensing, Vol. 9, No.4, p. 641-653, 1988.

9. Pratt, W. K., "Median Filtering", Semiannual Report of Image Proc., Inst., Univ. of Southern California, 1975.

10. Begin, D., Q.H. Gwyn, F. Bonn, "Radiometric Correction of SAR images, a new algorithm", Int. J. Remote Sensing, Vol. 8, no. 3, p. 385-398, 1987.

11. Hansson, Tjelmeland, Johannessen, Olaussen, Karpuz, " Speckle Reduction and Maximumlikelyhood Classification of SAR Images from Sea Ice Recorded During MIZEX'87" Proceedings of IGARSS '89 Symposium.

12. Mather, P.M., "Computer Processing of Remotely Sensed Images, Suffolk, p. 283-289, 1987.

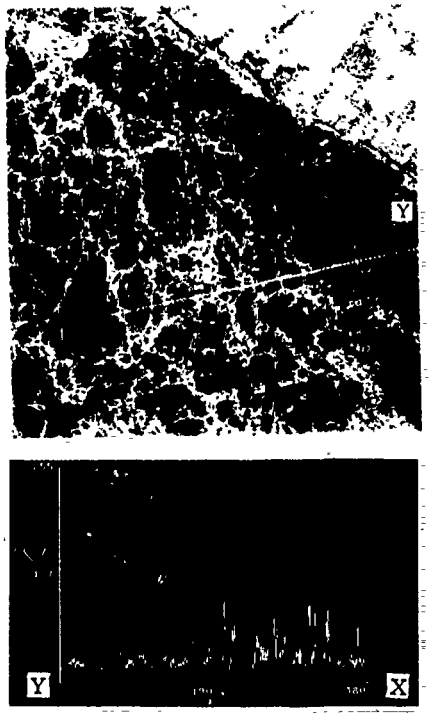


Fig. 2a

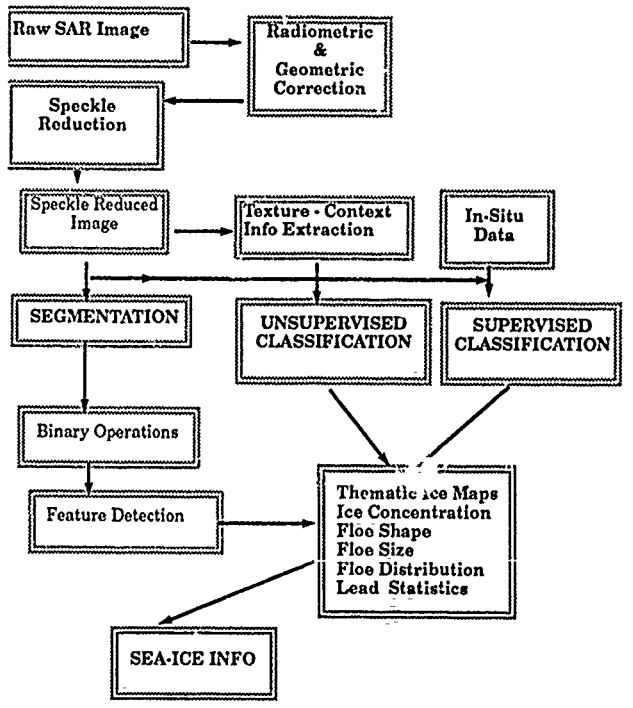


Fig.1 General outline of sea-ice information extraction algorithms.

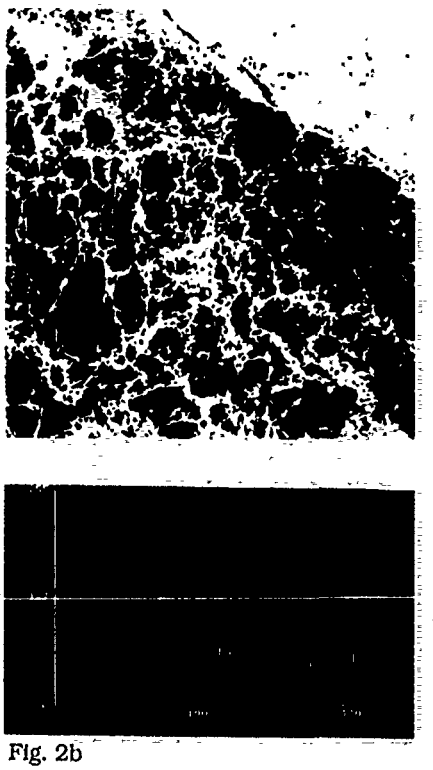


Fig. 2b

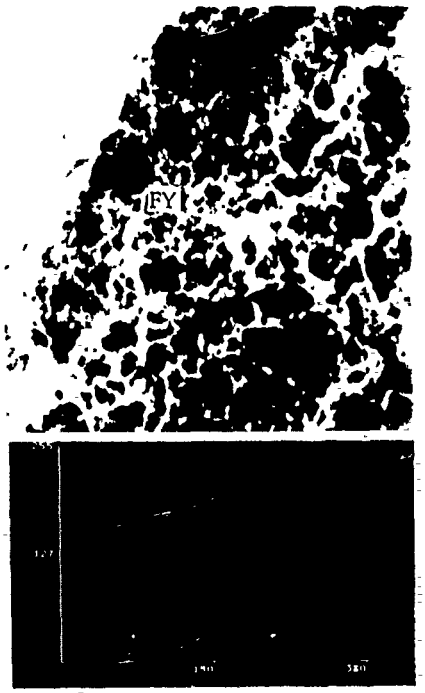


Fig. 2c

Fig. 2 Results of median filtering with various kernel sizes and their image profiles through X-Y. Note degraded textural details in first yerar ice (FY) covered areas. Ground resolution : 15m Scale : 1/100000

a) Raw SAR sub-scene

b) Filtered with a kernel size 3x3 pixels

c) Filtered with a kernel size 7x7 pixels

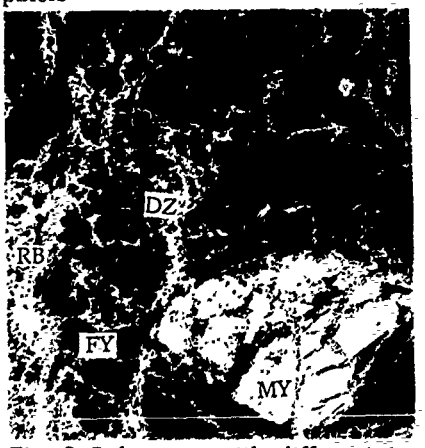


Fig. 3 Sub-scene with different ice types. Deformation zones (DZ) between first year floes (FY) and multi year ice floe fragments (MY) as well as rubble fields (RB) have the same backscatter values. Ground resolution: 15m Scale : 1/100000



Fig. 4 Swell propogation in ice creates systematic bright bands consisting of broken floe parts, mixture of different ice types. High backscatter values of these bands can not be correlated with ice types. Ground resolution : 15m Scale : 1/100000

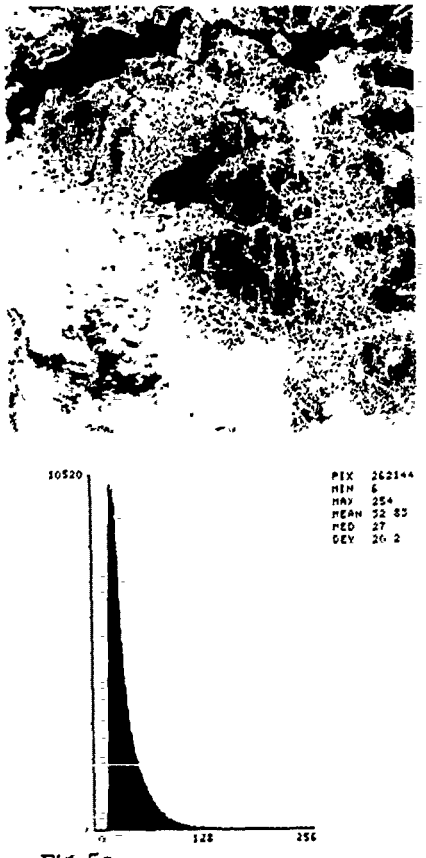


Fig. 5a

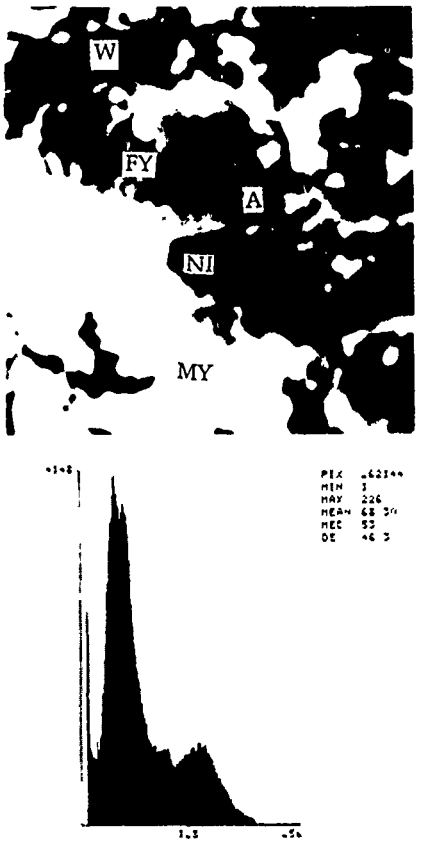


Fig. 5b

Fig. 5 Segmentation of images with iterative filtering. Sea water/thin new ice can be separated from other ice types with a single threshold in the multimodal image histogram. The area marked with "A" is used to obtain local statistics. Ground resolution : 15m Scale : 1/100000

a) Raw image and its unimodal image histogram

b) Segmented image and its multimodal histogram.

Water /thin ice : W

New ice : NI

First year ice : FY

Multi year ice : MY