AGARD-CP-582

AGARD-CP-582

AGARD

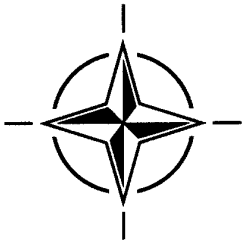
ADVISORY GROUP FOR AEROSPACE RESEARCH & DEVELOPMENT
7 RUE ANCELLE, 92200 NEUILLY-SUR-SEINE, FRANCE

AGARD CONFERENCE PROCEEDINGS 582

Remote Sensing:
A Valuable Source of Information
(la Télédétection — source précieuse de renseignements)

Papers presented at the Sensor and Propagation Panel Symposium, held in Toulouse, France 22-25 April, 1996.

19970107 039



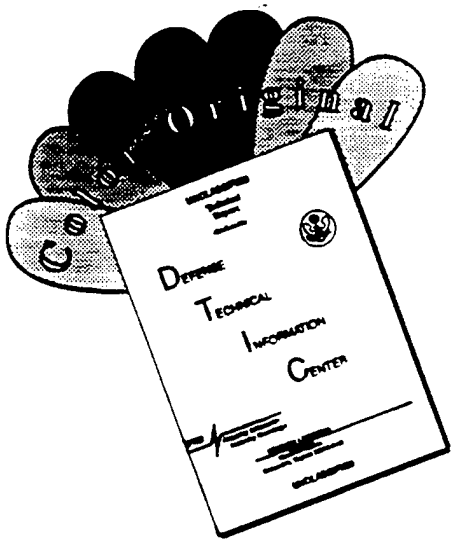
NORTH ATLANTIC TREATY ORGANIZATION

DISTRIBUTION STATEMENT A
Approved for public release

Published October 1996

Distribution and Availability on Back Cover

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF COLOR PAGES WHICH DO NOT REPRODUCE LEGIBLY ON BLACK AND WHITE MICROFICHE.

AGARD-CP-582

AGARD

ADVISORY GROUP FOR AEROSPACE RESEARCH & DEVELOPMENT
7 RUE ANCELLE, 92200 NEUILLY-SUR-SEINE, FRANCE

AGARD CONFERENCE PROCEEDINGS 582

Remote Sensing: A Valuable Source of Information

(la Télédétection — source précieuse de renseignements)

Papers presented at the Sensor and Propagation Panel Symposium, held in Toulouse, France 22-25 April, 1996.



North Atlantic Treaty Organization
Organisation du Traité de l'Atlantique Nord

DTIC QUALITY INSPECTED 1

The Mission of AGARD

According to its Charter, the mission of AGARD is to bring together the leading personalities of the NATO nations in the fields of science and technology relating to aerospace for the following purposes:

- Recommending effective ways for the member nations to use their research and development capabilities for the common benefit of the NATO community;
- Providing scientific and technical advice and assistance to the Military Committee in the field of aerospace research and development (with particular regard to its military application);
- Continuously stimulating advances in the aerospace sciences relevant to strengthening the common defence posture;
- Improving the co-operation among member nations in aerospace research and development;
- Exchange of scientific and technical information;
- Providing assistance to member nations for the purpose of increasing their scientific and technical potential;
- Rendering scientific and technical assistance, as requested, to other NATO bodies and to member nations in connection with research and development problems in the aerospace field.

The highest authority within AGARD is the National Delegates Board consisting of officially appointed senior representatives from each member nation. The mission of AGARD is carried out through the Panels which are composed of experts appointed by the National Delegates, the Consultant and Exchange Programme and the Aerospace Applications Studies Programme. The results of AGARD work are reported to the member nations and the NATO Authorities through the AGARD series of publications of which this is one.

Participation in AGARD activities is by invitation only and is normally limited to citizens of the NATO nations.

The content of this publication has been reproduced
directly from material supplied by AGARD or the authors.

Published October 1996

Copyright © AGARD 1996
All Rights Reserved

ISBN 92-836-0032-0



Printed by Canada Communication Group
45 Sacré-Cœur Blvd., Hull (Québec), Canada K1A 0S7

Remote Sensing: A Valuable Source of Information

(AGARD CP-582)

Executive Summary

Remote sensing systems can be broadly grouped in two categories: land remote sensing systems (designed to look at natural and man-made features at the land surface) and environmental systems, which can be further sub-divided into ocean/ice observing and meteorological systems. Land remote sensing and ocean/ice observing systems are often imaging systems, while meteorological systems can be both imaging and non-imaging. Imaging sensors can be passive (optical and infra-red) or active (radar, especially SAR: Synthetic Aperture Radar). The objectives of the symposium were to describe the state-of-the-art in remote sensing and to bring the experts in the field together to chart new directions. This was achieved by devoting the four-day program of sessions to sensors systems, propagation effects, signal processing and modeling, special techniques and observations, military benefits of civil systems, and military aspects of remote sensing.

Most satellite systems are civil systems. In 1995, the AGARD AAS-42 Study Team performed a study of the military applications of commercial satellites for remote sensing, communication and navigation. In an invited paper, a summary was given of remote sensing study results indicating that well-organized use of non-military remote sensing satellites among NATO countries can be a valuable supplement to current NATO surveillance data sources. Also, in a variety of situations, fast access to commercial remote sensing images can have a high value for a potential NATO adversary.

Examples of civilian systems that may have potential military application include the use of RADARSAT satellite data for global multilateral operations, change detection and geographical information systems. This satellite and future (optical) high resolution satellites provide basic data for digital military maps (topographic mapping).

Three invited overview papers addressed the areas of spaceborne optical remote sensing and SAR systems, including signal processing aspects and interferometry. The potential values of different techniques in the different parts of the electromagnetic spectrum were described. SAR interferometry can be usefully exploited to generate Digital Elevation Maps, to monitor terrain changes and to improve range resolution. The various techniques and the signal processing involved are not always straightforward and simple. High spatial resolution, including the vertical direction, can be obtained from interferometry using airplanes; however, due to the instability of the platform the single pass technique has to be used in which two antennas are placed on board the aircraft.

In addition to SAR interferometry and the basic SAR imaging mode, systems can be made to operate in other modes such as multi-frequency imaging, polarimetry, moving target indication and the spotlight mode. Application of complex and imaginative signal processing techniques were shown to significantly enhance the information content derived from the data.

Results from the EOPACE project (Electro-Optical Propagation Assessment in a Coastal Environment) showed the importance of large, near-surface aerosols on sea skimmer missile detection. The first phase of this multi-national project is being executed near the west coast of the US. Within the program, experiments and modeling efforts focus on refractivity profiles, propagation, target and (sea) background signatures, weather forecast, ducting and aerosol extinction. This project addresses the ever-increasing importance of predicting the propagation environment in coastal zones.

One outcome of the symposium is the recognition that, because of the enormous cost of space-based platforms, civilian-military sharing of such platforms and international cooperation will become increasingly necessary. Evolutionary changes to existing systems and platforms are favored over revolutionary changes, again because of cost constraints.

La télédétection — source précieuse de renseignements

(AGARD CP-582)

Synthèse

Les systèmes de télédétection terrestres sont composés de deux grandes catégories : les systèmes terrestres (conçus pour l'observation des caractéristiques naturelles et artificielles de la surface de la terre) et les systèmes d'observation de l'environnement, lesquels peuvent être subdivisés à leur tour en systèmes d'observation des océans et des glaces océaniques et des systèmes météorologiques. Les systèmes de télédétection terrestres et les systèmes d'observation des océans/des glaces sont souvent des systèmes à imagerie, tandis que les systèmes météorologiques sont indifféremment des systèmes à imagerie ou non. Les senseurs d'images sont soit passifs (optiques et infrarouge), soit actifs (radar, et, en particulier, les radars à ouverture synthétique SAR). Ce symposium a eu pour objectif de présenter l'état actuel des connaissances dans le domaine de la télédétection et de rassembler les spécialistes en la matière pour établir les nouvelles directions. Cet objectif a été réalisé par le biais d'un programme de quatre jours couvrant les systèmes de télédétection, les effets de propagation, le traitement et la modélisation du traitement du signal, les techniques et les observations spéciales, l'intérêt militaire présenté par les systèmes civils, et les aspects militaires de la télédétecion.

La plupart des systèmes à satellites sont des systèmes civils. En 1995, le groupe d'étude AAS-42 de l'AGARD a réalisé une étude sur les applications militaires des satellites commerciaux pour la télédétection, les télécommunications et la navigation. L'une des communications présentée sur invitation, a fait la synthèse des résultats d'un certain nombre d'études sur la télédétection. Ces études indiquent que l'utilisation planifiée des satellites de télédétection civils par les pays membres de l'OTAN peut apporter un supplément appréciable aux sources de données de surveillance de l'OTAN. Aussi, dans diverses situations, l'accès rapide aux images de télédétection commerciales peut être d'une grande valeur pour les adversaires potentiels de l'OTAN.

Parmi les exemples de systèmes civils ayant des applications militaires potentielles, l'on peut citer les données satellite de RADARSAT pour les opérations globales multilatérales, la détection des changements et les systèmes d'informations géographiques. Ces satellites, ainsi que les futurs satellites à haute résolution (optiques), fournissent les données de base pour l'élaboration des cartes militaires numériques (cartographie topographique).

La télédétection optique spatiale et les systèmes SAR, y compris le traitement du signal et l'interférométrie, ont été traités dans trois communications de synthèse présentées sur invitation. L'intérêt potentiel de l'emploi de différentes techniques dans les différents secteurs du spectre électromagnétique a été décrit. L'interférométrie SAR peut être exploitée pour l'élaboration de cartes de niveaux numériques, pour le suivi de l'évolution du terrain et pour l'amélioration de la résolution en distance. Les différentes techniques utilisées, comme le traitement du signal, ne sont pas toujours claires et simples. La haute résolution spatiale, y compris verticalement, peut être obtenue à l'aide de l'interférométrie aéroportée; cependant, en raison de l'instabilité de la plate-forme la technique par passage unique, à deux antennes aéroportées, est nécessaire.

En plus de l'interférométrie SAR et du mode imagerie SAR de base, il est possible de faire fonctionner les systèmes en d'autres modes tels que l'imagerie multifréquence, la polarimétrie, le MTI et le mode téléscope. Il a été démontré que l'application de techniques de traitement du signal complexes et novatrices permet d'accroître considérablement le contenu informationnel obtenu à partir des données brutes.

Les résultats obtenus du projet EOPACE (Electro-Optical Propagation Assessment in a Coastal Environment) montrent l'importance des grandes nappes à proximité de la surface de la mer dans la détection des missiles à vol rasant. La première phase de ce projet multinational est en cours de réalisation sur un site situé près de la côte ouest des Etats-Unis. Les différentes expériences et modélisations entreprises dans le cadre de ce programme concernent principalement les profils de réfringence, la propagation, les signatures de fond de la cible (et de la mer), la météo, la propagation guidée et l'extinction des aérosols. Ce projet examine l'importance toujours croissante de la prédiction de l'environnement de propagation en zone côtière.

L'une des conclusions tirées du symposium est que, en raison du coût énorme des plates-formes spatiales, le partage civil-militaire de tels équipements, ainsi que la coopération internationale seront de plus en plus nécessaires. Au niveau des plates-formes et systèmes existants, la préférence est accordée aux changements évolutifs plutôt qu'aux développements révolutionnaires, d'ailleurs en raison des contraintes budgétaires.

Contents

	Page
Executive Summary	iii
Synthèse	iv
Theme/Thème	ix
Sensor and Propagation Panel	x
Keynote Address/Note dominante de la Conférence **by Ingénieur Général de l'Armement G. Bonnevalle** **(AGARD French National Delegate)**	K
SESSION I: OVERVIEW **Chairman: Prof J. Rogge**	
Spaceborne Optical Remote Sensing — Abstract Only by F. Lanzl	1
Possibilities and Limits of SAR Interferometry by F. Rocca, C. Prati and A. Monti Guarnieri	2
SAR Systems and Related Signal Processing by P. Hoogeboom, R.K. Dekker and M.P.G. Otten	3
SESSION II: SENSOR SYSTEMS **Chairman: Dr. Ph. Raisonville**	
Spotlight Synthetic Aperture Radar: System Overview and Application to the Maritime and Land Environments by A. Damini and G.E. Haslam	4
MM-Wave Airborne Cloud Radars by J. Galloway, A. Pazmany, R. McIntosh, R. Kelly and G. Vali	5
Imaging Tools for Spaceborne SAR Programme by J.M. Hermer, E. Normant, P. Martynerie, J.P. Hardange, J.M. Boutry and U. Erich	6
SARAPE: Un outil de mesures radar du fouillis **(A Tool for Radar Clutter Measurement)** by J.-C. Motet, B. Marechal, C. Cochin, R. Guillerey and A. Berges	7
Remote Sensing by Active and Passive Optical Techniques by C.R. Philbrick, M.D. O'Brien, D.B. Lysak, T.D. Stevens and F. Balsiger	8
SESSION III: PROPAGATION ASPECTS **Chairman: Dr. W. Keydel**	
Diffusion des micro-ondes par des hydrométéores non sphériques et inhomogènes. Simulations, observations et restitutions durant l'expérience TOGA-COARE. by E. Obligis and C. Klapisz	9

An Overview of EOPACE (Electrooptical Propagation Assessment in Coastal Environments), Including In Situ and Remote Sensing Techniques 10
by K.M. Littfin and D.R. Jensen

Lateral Resolution Considerations of Refractivity Profiles Associated with Remote Sensing Measurements of Over-Water Coastal Regions 11
by J. Goldhirsch and G.D. Dockery

SESSION IV: PROPAGATION EFFECTS
Chairman: Dr. F. Barbaliscia

Direct Measurement of Microwave Propagation Effects 12
by J. Stapleton and S. Kang

Modelling of Rough Surface Effects with the Matched Transform PE 13
by M.F. Levy

Topex Poseidon Follow-On Radar Altimeter Design and Development Status — Abstract Only 14
by P. Escudier, P. Raizonville, G. Carayon and P. de Chateau Thierry

Paper deleted 15

Using Remote Refractivity Sensing to Predict Tropospheric Refractivity from Measurements of Microwave Propagation 16
by D. Boyer, G. Gentry, J. Stapleton, S. Burk and J. Cook

SESSION V: SIGNAL PROCESSING
Chairman: Dr. E. Krogager

On the Importance of Utilizing Polarimetric Information in Radar Imaging and Classification 17
by E. Krogager and W.-M. Boerner

Problème de modélisation lié à la limitation temporelle de l'impulsion émise par le radar-altimètre POSEIDON 18
by J. Barckicke

Characterization of Natural Targets Using Multiparametric SIR-C/X-SAR Imagery 19
by L. Castellano, P. Murino and A. Siciliano

Microwave Remote Sensing of Natural Targets: Theory and Experiment 20
by J.-C. Souyris, T. Le Toan, N. Floury, C.-C. Hsu, L.-F. Wang and J.A. Kong

SESSION VI: SIGNAL PROCESSING AND MODELLING
Chairman: Prof. M. Acheroy

Autoadaptive Scene Identification in Multi-Spectral Satellite Data 21
by P. Boeckaerts, E. Nyssen and J. Cornelis

Paper not available for publication 22

Paper deleted 23

Modeling and Wave-Oriented Processing of Scattering Data from Complex Environments 24
by L.B. Felsen

A Generic Image Processing Tool: Top Down Bayesian Image Analysis 25
by S. Dewitte, J.M. Mangen, J. Cornelis and D. Crommelynck

Closed Form Expressions of Algorithmic Performance and Complexity in Signal Processing 26
by A. Marguinaud and P. Bertrand

A Gradient Based Line Detector 27
by V. Lacroix and M. Acheroy

Eco-Climatic Classifications by Use of NOAA-AVHRR Data. A Case Study: Tuscany (Central Italy) 28
by L. Petkov, F. Maselli, C. Conese and G. Tacconi

SESSION VIII: SPECIAL TECHNIQUES AND OBSERVATIONS
Chairman: Prof. A.K. Hyder

Target Detection Via Measurements Taken by a Transmitter — Independent Receiver Network 29
by N.J. Farsaris and P.P. Stavroulakis

Short-Term Persistence of Low Cloud Features 30
by R.A. Helvey

Electromagnetic Properties of Sea Ice — Abstract Only 31
by K.C. Jezek

A Global Positioning System (GPS) Tide Gauge 32
by K.D. Anderson

Assessment of NAM, the Navy Aerosol Model, for Visibility Determinations in Inland Seas such as the Persian Gulf 33
by S.G. Gathman

Sensing Zero-Range Surface Ship Infrared Signatures 34
by C.P. McGrath and C.R. Zeisse

Radiance of the Wind-Ruffled Sea 35
by C.R. Zeisse, C.P. McGrath, K.M. Littfin and H.G. Hughes

An Algorithm for Ship Temperature Retrieval from Airborne Infrared Radiometric Measurements 36
by H.G. Hughes, C.R. Zeisse and C.P. McGrath

SESSION X: MILITARY BENEFIT OF CIVIL SYSTEMS
Chairman: Dr. R.B. Gomez

Some Results from AGARD Study AAS-42: Potential Use of Commercial Satellite Systems for Military Applications 37
by M. Dinguirard, T. Wahl, J.M. McDonald, G. Carrin, D. Levy, U. Erich, K.H. Zeller and X. Briottet

RADARSAT: A Dual Use Earth Observation System 38
by M. McKean, M. Rey, H. Edel, C. Bjerkelund, J.L.P. Dionne and D. Nazarenko

Automatic Change Detection in Space Borne SAR Imagery 39
by D.G. Corr and S.W. Whitehouse

La spatiocarte numérique Défense 40
by M.N. Sclafer and J.P. Cantou

SESSION XI: MILITARY ASPECTS
Chairman: Prof. P. Murino

Electronic Countermeasures Against Synthetic Aperture Radars 41
by G. Krämer

SAR Measurements of Extended Targets at 94 GHz 42
by H. Schimpf and H. Essen

Airborne System Detection and Localization of Radio Interference Sources 43
by B. Audone, F. Bresciani and F. Novarese

VLSI Implementation of a Chip for a Systolic Array Architecture for Direction Finding Systems 44
by B. Audone, F. Bresciani and D. Zerlottin

SESSION XII: SPECIAL APPLICATIONS
Chairman: Dr. R.B. Gomez

Multispectral Remote Sensing of the Coastal Atmospheric Boundary Layer 45
by C.H. Wash, K.L. Davidson and M.S. Jordan

Use of Remote Sensing for Monitoring the Lagoon Köyceğiz and its Environment in Turkey 46
by D. Maktav, F. Sunar, S. Kapdaşli and N. Musaoğlu

Synoptic Variability Revealed by Satellite and Equivalent Altitude 47
by P. Greiman, J. Rosenthal and R.A. Helvey

Remote Sensing of Evaporation Ducts Using SHF Propagation Measurements 48
by L.T. Rogers

Review of Millimeter Wave Radiometry Potentialities. Activities in CERT-ONERA and DLR-HF 49
(Potentialités de la Radiométrie en Ondes Millimétriques. Activités au CERT-ONERA et au DLR-HF)
by J. Lemorton and H. Suess

Expectations from EMET and OESEE Ground Listening of Possible E.M. Events Generated by the TSS 1R 50
by R.D. Estes, J. Sanmartin and G. Tacconi

Theme

Remote sensing, either from satellites, airplanes or RPVs, can be used to obtain information from virtually all areas in the world. The information is related to such diverse topics as the position of military installations, the location of fixed natural objects (rivers) and man-made objects (infrastructure, roads, bridges, the vegetation, the topography, the ground-state, the meteorological conditions, the oceanographic conditions, the environmental situation, etc. Different sensor systems use different parts of the electromagnetic spectrum. Each of these systems has its own (dis)advantages, amongst others related to geometrical resolution and allweather/day-and-night capabilities. Multispectral systems and data fusion can to a large extent combine the advantages and overcome the disadvantages. Civil remote sensing systems provide mainly general, non-real time data, useful in the planning phase of operations, whereas sensors aboard military platforms can contribute to the actual awareness of the situation. Increased technological capabilities and new methods of signal processing techniques will lead to new military applications.

Military operations in which NATO forces are involved currently take place in areas outside the NATO theatre. Areas to which (for political and/or military reasons) ground or short range airplane surveillance equipment may have no access, neither prior to the military operation, nor during the actual conduct. Remote sensing offers a potential source of information for all areas of the world, independant of the approval of local authorities. A wealth of information is relatively easily (and relatively cheaply) available from sensor systems in civil satellites, but this availability could decrease during times of increasing political tensions. Cooperative programs must be encouraged to extract the required military information from unclassified civil sources. This symposium provides a platform for discussion between civil institutes, active in the area of remote sensing equipment, data handling and processing on the one hand, and those who are involved in the military applications on the other hand.

The topics to be covered include:

- sensor systems;
- platforms;
- propagation effects;
- signal processing;
- military aspects (requirements, applications);
- military benefit of civil systems;
- (special) civil applications;
- special techniques;
- future developments.

Thème

La télédétection, réalisée par des satellites, des aéronefs et des véhicules téléguidés, permet de recevoir des informations de pratiquement toutes les régions du monde. Ces informations concernent les sujets les plus divers, tel que l'emplacement d'installations militaires, le tracé d'éléments naturels fixes (cours d'eau) et artificiels (infrastructures, routes), la végétation, la topographie, l'état du sol, les conditions météorologiques, les conditions océanographiques, la situation de l'environnement etc. Les différents systèmes de détection utilisent des parties différentes du spectre électromagnétique. Chaque système a ses propres avantages et inconvénients, en particulier en ce qui concerne la résolution géométrique et les caractéristiques d'emploi tous temps et jour/nuit. Les systèmes multispectre et de fusionnement des données permettent, dans une large mesure, de cumuler les avantages et de s'affranchir des inconvénients inhérents en ces systèmes. Les systèmes de télédétection civils fournissent pour la plupart des données générales, non temps-réel, exploitables lors de la phase de planification des opérations, où les senseurs installés sur les plates-formes militaires peuvent apporter une contribution à la perception de la situation. Les capacités technologiques accrues et les nouvelles techniques du traitement du signal conduiront à de nouvelles applications militaires.

A l'heure actuelle, les opérations militaires nécessitant l'intervention des forces de l'OTAN ont lieu en dehors du théâtre des opérations habituel de l'OTAN, c'est-à-dire dans des régions où, pour des raisons politiques et/ou militaires, l'accès nécessaire à l'exploitation d'équipements de surveillance basés au sol ou aéroportés par avion léger n'est disponible ni avant, ni pendant les opérations. La télédétection représente une source possible d'informations sur toutes les régions du globe, indépendamment de l'approbation des autorités locales. Un volume de données énorme est à présent disponible grâce aux systèmes de détection installés sur les satellites civils, et ceci de façon relativement rapide et peu coûteuse, mais cette disponibilité risque d'être restreinte en période de tension internationale. Il y a lieu de promouvoir des programmes de coopération afin de pouvoir extraire des sources civiles non classifiées de renseignement militaire demandé. Ce symposium fournit une plate-forme pour la discussion de ces questions par les représentants des instituts civils, qui travaillent activement dans le domaine de la conception des équipements de télédétection et de la collecte et du traitement des données d'une part, et par ceux qui sont responsables des applications militaires d'autre part.

Les sujets à examiner comprennent :

- les systèmes de détection;
- les plates-formes;
- les effets de propagation;
- le traitement du signal;
- les aspects militaires (les besoins, les applications);
- l'intérêt militaire des systèmes civils;
- les applications civiles (spéciales);
- les techniques spéciales;
- les développements futurs.

Sensor and Propagation Panel

Chairman: Prof. D.H. HÖHN
FGAN
Forschungsinstitut für Optik – FfO
Schloss Kressbach
72072 Tübingen
Germany

Deputy Chairman: Mr. F. CHRISTOPHE
Dept Micro-Ondes
ONERA-CERT Toulouse
BP 4025
2, Avenue E. Belin
31055 Toulouse Cedex
France

SYMPOSIUM
TECHNICAL PROGRAMME COMMITTEE

Co-Chairmen: Dr. J. RICHTER (US)
Prof. J. ROGGE (NE)

Members: Prof. M. ACHEROY (BE)
Dr. F. BARBALISCIA (IT)
Dr. R.B. GOMEZ (US)
Prof. A.K. HYDER (US)
Dr. W. KEYDEL (GE)
Dr. E. KROGAGER (DE)
Prof. P. MURINO (IT)
Dr. Ph. RAISONVILLE (FR)
Dr. T. ROBINSON (UK)

PANEL EXECUTIVE

Lt-Col. G. DEL DUCA (Italy)

from Europe
AGARD-NATO
ATTN: SPP Executive
7, rue Ancelle
92200 Neuilly-sur-Seine
France

Phone: 33-1-47.38.57.68

from North America
AGARD NATO / SPP
PSC 116
APO AE 09777

Fax: 33-1-47.38.57.99

Telex: 610176F

Keynote Address

Ingénieur Général de l'Armement G. BONNEVALLE
Inspecteur de l'Armement pour l'Aéronautique et l'Espace
B.P. 24
92211 Saint Cloud Cedex
France

Ladies and Gentlemen,

I am very happy to welcome you to Toulouse on behalf of the French National Delegates to AGARD for the inaugural session of this 5th Symposium organised by the SPP Panel on the military aspects of remote sensing.

Like all AGARD Panels, the SPP is particularly active and effective; its meetings are attended by a high number of participants and your presence here today serves to confirm this once again.

Since its creation in 1952 by Professor von Kármán, with the aim of encouraging the development of aeronautical techniques in the Alliance countries, AGARD has endeavoured to enhance the state of the art in the field of aerospace and to solve certain specific problems submitted to it by the NATO Military Committee.

The advancement of this knowledge is achieved by the exchange of information between representatives of government institutions, manufacturers and academics in the member countries, by Panel meetings, by working groups, specialists meetings, symposia, lecture series and scientific and technical publications of high quality and considerable notoriety.

The particular problems raised by the NATO Military Committee, which are of a technico-operational nature, are examined by means of summary reports.

Thanks to the contacts set up by these different activities, a real community spirit has developed within AGARD.

This AGARD spirit is a living reality.

Among the different NATO Organisations, it seems to me that AGARD has emerged as the one which has been the most successful in the field of cooperation, thanks to the links which have been forged between scientists, research workers and specialists in the aerospace sciences.

This cooperation is more than ever necessary today, when R&D costs are increasing and the financial and human resources available are assigned with increasing parsimony.

The SPP Panel is concerned with a field in which scientific and technical progress is decisive for the development of equipment and of working conditions: electromagnetic and electrooptical propagation, electromagnetic and infrared signature, communications, surveillance and detection techniques, electronic warfare and directed energy.

The field of activity includes in particular, sensors, targets, the external environment and propagation through the atmosphere and in space.

A brief look at the programme for your symposium shows the variety, interest and relevance of the papers presented.

This unclassified Spring symposium, concerns long range sensing systems carried by satellites, aircraft and drones, which are a source of high value operational intelligence and which can be used to obtain information about any region of the globe.

The subjects dealt with include sensor systems, platforms, propagation effects, signal processing, military aspects, the use of civil systems for military purposes, special civil applications, special techniques and future developments.

At a time when the French defence community is reaffirming its priorities in the field of intelligence, when European military cooperation is establishing itself in satellite observation, and when NATO forces are liable to intervene in distant and little known theatres of operation, the new observation technologies have an essential role to play.

Optimised by skilful use of the physical laws, these technologies can effectively enable remote, real-time assessment of ground, sea and atmosphere states and can

Paper presented at the AGARD SPP Symposium on "Remote Sensing: A Valuable Source of Information", held in Toulouse, France from 22-25 April 1996, and published in CP-582.

also be used to quantify concentrations of troops or vehicles, and to detect, identify and locate tactical and strategic targets.

But high performance remote sensors are not only useful for military requirements; observation of the earth's resources involves related technologies which have been tried and tested for several years now. This dual approach is fundamental in the present period of military budget reductions.

Toulouse, which already holds a prime place in the field of civil observation satellites, thanks to the presence of the National Space Studies Centre, to educational establishments such as the one we are in today, to a number of major government laboratories and to leading manufacturers in this sector, is involved in a significant way in the development of military observation satellites, in close collaboration with our European partners in the Helios and Horus programmes.

With such an appropriate setting, I am sure the presentations will be of excellent quality and will be followed by lively discussions.

I would particularly like to thank all those who have willingly undertaken the considerable task of preparing a presentation.

I should also like to thank the organisers of these meetings.

Finally, I wish to thank the Director of the National Aeronautics and Space Academy for kindly placing these premises at our disposal.

I am convinced that the results of your work will also be useful to the decision makers in the Alliance nations and I wish you every success.

Note dominante de la Conférence

Ingénieur Général de l'Armement G. BONNEVALLE
Inspecteur de l'Armement pour l'Aéronautique et l'Espace
B.P. 24
92211 Saint Cloud Cedex
France

Mesdames, Messieurs,

C'est pour moi une très agréable mission de vous souhaiter la bienvenue à Toulouse au nom des Délégués Nationaux Français de l'AGARD lors de la séance inaugurale de ce cinquième symposium organisé par le Panel sur les Senseurs et la Propagation, sur les aspects militaires de la télédétection.

Comme tous les panels de l'AGARD, le Panel sur les Senseurs et la Propagation est particulièrement actif et efficace; ses réunions sont suivies par un grand nombre de participants; votre présence aujourd'hui dans cette salle en atteste une fois encore.

Depuis sa création en 1952 par le Professeur von Kármán dans le but de favoriser le développement des techniques aéronautiques dans les pays de l'Alliance, l'AGARD s'efforce de faire progresser l'état des connaissances techniques dans le domaine aérospatial et de résoudre certains problèmes spécifiques posés par le Comité Militaire de l'OTAN.

La progression des connaissances passe par des échanges d'informations entre les représentants des institutions gouvernementales, industrielles et universitaires des pays membres par les réunions de panels, par les groupes de travail, les réunions de spécialistes, les séries de conférences, les symposia, les publications techniques et scientifiques de haut niveau et de grande notorieté.

Les problèmes spécifiques posés par le Comité Militaire de l'OTAN, de nature technico opérationnelle, font l'objet d'études de synthèse.

Grâce aux contacts initiés par ces différentes activités, un véritable esprit communautaire s'est developpé au sein de l'AGARD.

L'esprit AGARD est une vivante réalité.

Parmi les différentes organisations OTAN, l'AGARD s'est imposée me semble-t-il comme celle qui a le mieux réussi dans le domaine de la coopération grâce aux liens forgés entre les savants, les chercheurs et les spécialistes des sciences aérospatiales.

Aujourd'hui même cette coopération est toujours et plus que jamais nécessaire, d'autant que les coûts de recherche et des développements sont de plus en plus élevés et que les resources financières et humaines à y affecter sont de plus en plus comptés avec parcimonie.

Le panel SPP a pour vocation de s'intéresser à un domaine ou les progrès scientifiques et technologiques sont déterminants pour le développement des materiels et les conditions d'emploi : propagation électromagnétique et électro optique, signature électromagnétique et infrarouge, technique de communication, de surveillance et de détection, guerre électronique, énergies dirigées.

Son domaine d'activité comprend notamment les senseurs, les cibles, l'environnement extérieur et la propagation à travers l'atmosphère et l'espace.

La lecture du programme de votre symposium montre la varieté, l'intérêt et la pertinence des communications présentées.

Ce symposium de printemps, non classifié, concerne les systèmes de détection lointains à bord de satellites, d'avions, de drônes qui constituent une source de renseignement de très grande valeur opérationnelle et qui peuvent etre utilisés pour obtenir des informations sur n'importe quelle zone géographique dans le monde.

Les sujets traités incluent les systèmes de senseurs, les plateformes, l'effet des propagations, le traitement des signaux, les aspects militaires, l'emploi des systèmes civils pour des utilisations militaires, les applications civiles spéciales, les techniques spéciales, les développements futurs.

A l'heure où la Défense française réaffirme ses priorités dans le domaine du renseignement, où la coopération militaire européenne se met en place pour l'observation satellitaire, et ou les forces de l'OTAN sont susceptibles

Paper presented at the AGARD SPP Symposium on "Remote Sensing: A Valuable Source of Information", held in Toulouse, France from 22-25 April 1996, and published in CP-582.

d'intervenir sur les theatres d'operations eloignes et mal connus, les nouvelles technologies d'observations sont appelées à un rôle essentiel.

Optimisées grâce à la maîtrise des lois physique, ces technologies peuvent en effet permettre d'apprécier à distance et un temps réel du sol, de la mer et de l'atmosphère mais aussi de quantifier les concentrations de troupes ou de véhicules, de détecter, d'identifier et de localiser des cibles tactiques ou stratégiques.

Mais des capteurs de télédétection performants ne répondent pas seulement aux besoins militaires, puisque l'observation des ressources terrestres fait appel à des technologies voisines, éprouvées depuis plusieurs années maintenant. Cette dualité est fondamentale dans la période de réduction des budgets militaires que nous connaissons tous actuellement.

Toulouse qui occupe déjà dans le domaine des satellites civils d'observation une position privilégiée grâce à l'implantation du Centre National d'Études Spatiales, d'établissements d'enseignement comme celui ou nous sommes, de plusieurs grands laboratoires publics et d'industriels majeurs de ce secteur est appelée à intervenir de façon significative dans le développement des satellites militaires d'observation, et cela en liaison étroite avec nos partenaires européens des programmes Hélios et Horus.

Dans un lieu aussi approprié, je ne doute pas que les exposés seront d'une excellente qualité, suivis de discussions très animées.

Je remercie tout particulièrement tous ceux qui ont bien voulu se charger du travail considérable que repésente la préparation d'un exposé.

J'adresse également mes remerciements aux organisateurs de ces réunions.

Je tiens également à remercier le Directeur de l'Ecole Nationale Supériere de l'Aéronautique et de l'Espace qui a aimablement mis ces installations à notre disposition.

Je suis convaincu que les résultats de vos travaux seront utiles égalements aux décideurs dans les nations de l'Alliance et je vous adresse tous mes vœux de succès.

ABSTRACT

of

Paper 1

"Spaceborne Optical Remote Sensing"

Prof. Dr. Franz Lanzl
Director of the Institute of Optoelectronics
DLR Germany

The combined high-resolution stereo-multispectral optoelectronic camera MOMS was flown on the Shuttle D2 Mission in 1993 and will be operating for about two years in the framework of the PRIRODA/MIR mission to be launched in April 1996.
Mission objectives, data evaluation, application examples including digital terrain model derivation are presented.

Plans for a near-term operational high-resolution optical Earth observation system based on the MOMS development are discussed.

Paper presented at the AGARD SPP Symposium on "Remote Sensing: A Valuable Source of Information", held in Toulouse, France from 22-25 April 1996, and published in CP-582.

DISCUSSION

Paper 1

E. Schweicher (BE)

Why did you restrict the thermal image of your last slide to the band 10400-1500 nm? Why not starting the band at 8000 nm?

Author's reply:

The LANDSAT data is available for a 10400-12500 nm band and not at 8000 nm.

M. McKean (CA)

The sheer amount of high resolution optical data and access to it is an issue of concern. How is this, or can it be dealt with?

Author's reply:

This is an important issue, and one that requires careful consideration when designing the remote sensing ground segment. In the MOMS case we are working with 10 minutes of data taken per day, and this translates to 30 Gbit per day corresponding to approximately 200,000 sq.km.

However, the data processing chain is designed to handle this amount of data.

R. Philbrick (US)

The addition of the MOMS instrument to the MIR station appears to be a good opportunity to extend the imaging data. What problems are there in flying the MOMS package on MIR, particularly due to the manned activity influence on pointing?

Author's reply:

Since the operation of MOMS is done automatically by operating from the ground there should be no influence by the astronauts. In addition, the position and the attitude of MOMS are measured with a frequency of 10Hz. This additional information is regularly used for data correction.

POSSIBILITIES AND LIMITS OF SAR INTERFEROMETRY

Fabio Rocca, Claudio Prati, Andrea Monti Guarnieri
Dipartimento di Elettronica e Informazione- Politecnico di Milano
Pzza. L. da Vinci, 32, 20133 Milano. Italy
Tel: +39-2-23993573 Fax: +39-2-23993413
E-mail: rocca@elet.polimi.it

March 6, 1996

1. Summary

In this presentation we summarize the impact of interferometry on the design of SAR surveys, its possibilities and limits. From the fringes a very good DEM is obtainable, with a vertical resolution that could be in the meter range or even less from airborne platforms. Millimetric motion of large areas of the terrain or of corner reflectors can be measured with good reliability and therefore the possibility of using the system to measure subsidence, landslides, coseismic motions has been demonstrated. The coherence of the interferometric pair is an important clue, that combined with the more usual backscatter amplitude, leads to high quality images segmentation. The wavenumber shift concept is finally introduced that may lead to a 3 dimensional view of the terrain. The same principle can be invoked to improve the resolution of objects that do not change with time.

2. Introduction

Synthetic Aperture Radar (SAR) systems record both amplitude and phase of the backscattered echoes. The phase of each pixel (ground resolution cell) of a focused SAR image is the sum of three distinct contributions:

1 - the two-ways travel path (sensor-target-sensor: hundreds of kilometers in the satellite case) that, divided by the used wavelength (a few centimeters), corresponds to millions of cycles;

2 - the interaction between the incident e.m. waves and the scatterers within the ground resolution cell;

3 - the phase shift induced by the processing system used to focus the image.

Therefore, the phase of a single SAR image is of no practical use. On the contrary, if two SAR images from slightly different viewing angles are considered (*interferometric pair*) their phase difference (*interferometric fringes*) can be usefully exploited to generate Digital Elevation Maps (DEMs), to monitor terrain changes and to improve the range resolution. The interferometric fringes image is derived as the phase of the SAR *interferogram*, that is the complex image formed by cross-multiplying the two SAR images. The relation between the interferometric fringes and ground elevation is usually explained by means of the *monochromatic approach* [1, 2]. It is based on the assumption that the RF bandwidth is so small (and this is the case of most satellite systems including SEASAT, ERS-1, JERS-1, ERS-2 and RADARSAT) to be negligible. Thus the system is considered monochromatic. However, if the finite bandwidth of the system is considered (*wavenumber shift* approach), a relative shift of the ground wavenumber spectra dependent on the baseline and the local slope is found. A few important consequences come out from this result [3, 4, 5]. Using the simpler monochromatic approach we show the relationship between the relative terrain elevation and the interferometric fringes. Then we will determine the quality of the Digital Elevation Model derived from interferometry. Two sections 6 and 9.1 will be dedicated to the description of other applications of SAR interferometry: small terrain motion detection and uses of the coherence for image segmentation. In section 9 a deeper insight into SAR interferometry will be given by exploiting the wavenumber shift approach and the slant range resolution enhancement that can be obtained by using it. Symbols that will be used through the text and the main parameters values of the ERS-1 SAR system (that will be assumed as typical parameters of SAR from satellites) are shown in table 3.1.

3. The monochromatic approach

Let us consider two complex SAR images from two slightly different viewing angles (see figure 3.1), v_1 and v_2. Even if non simultaneous acquisitions are considered, we shall suppose, for now, that the terrain backscatter did not change.

Let us now exploit the "monochromatic approximation": the relative system bandwidth is so small (i.e. in the ERS-1 case its value is $3 \cdot 10^{-3}$) to be neglected. Thus, the phase difference ϕ between correspondent complex pixels in v_1 and v_2 is proportional to the travel

Paper presented at the AGARD SPP Symposium on "Remote Sensing: A Valuable Source of Information", held in Toulouse, France from 22-25 April 1996, and published in CP-582.

Symbol	Meaning
λ	wavelength
$f_o = c/\lambda$	central frequency
W	system bandwidth
f_s	sampling frequency
$f_{Ny} = f_s/2$	Nyquist frequency
θ	off-nadir angle
α	local terrain slope (range)
H	platform altitude
ρ_s	slant range resolution
Δr	slant range sampling interval
ρ_g	ground range resolution
B	baseline
B_n	normal baseline
B_r	radial baseline
B_{nc}	critical baseline
r	slant range axis
r_o	sensor-target distance
y	ground range axis
z	elevation axis
k_y	ground range wave number
ϕ	interferometric phase
Symbol	**ERS-1 value**
λ	5.66cm
$f_o = c/\lambda$	5.3GHz
W	16MHz
f_s	18.96MHz
θ	23deg.
H	780km
ρ_s	9 meters
B_{nc}	1100 meters

Table 3.1: List of symbols and values for ERS-1

path difference $2\Delta r_o$ (the factor 2 accounts for the two ways travel path):

$$\phi = \frac{4\pi}{\lambda}\Delta r_o \tag{3.1}$$

(λ is a constant). The interferometric phase ϕ of a single pixel is still of no practical use. The travel path difference $2\Delta r_o$ is usually much greater than the wavelength λ (in most of the practical cases, the travel path difference from the satellite can be as large as a few hundred meters whereas the used wavelength is of the order of a few centimeters) and the measured phase ϕ shows an ambiguity of many cycles. On the other hand, passing from one pixel to its neighbor (only a few meters apart in the slant range direction), the variation of the travel path difference ($\Delta(\Delta r_o)$) may be much smaller than λ and the variation of the interferometric phase $\Delta\phi$ is not ambiguous. Moreover, a simple relation between $\Delta\phi$ and the relative terrain elevation can be derived.

In figure 3.1 we have indicated the position of the

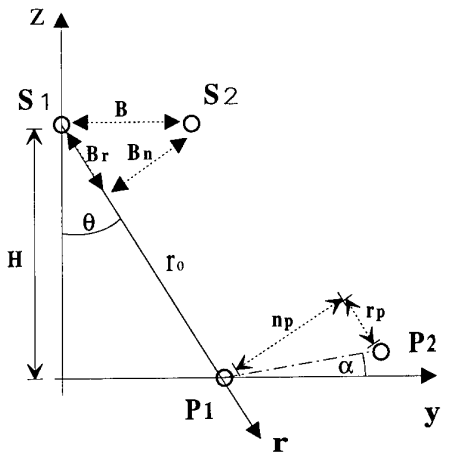


Figure 3.1: Interferometric SAR geometry in a plane orthogonal to the rectilinear platforms (S_1 and S_2) trajectory.

two SAR sensors (S_1 and S_2) and their relative displacements parallel (B_r) and normal (B_n) to the slant range direction. We have also indicated the position of two point scatterers and their relative displacements parallel (r_p) and normal (n_p) to the slant range direction. Let us assume as a reference the positions of S_1 and P_1 with their relative distance r_o. By changing the sensor and target position by (B_n, B_r) and (r_p, n_p) respectively the sensor-target distance becomes:

$$r = \sqrt{(r_o + r_p - B_r)^2 + (n_p - B_n)^2} \tag{3.2}$$

Since the distance between the two SAR sensors is generally much smaller than the sensor-target distance r_o (a few hundred meters compared with 800km, in the ERS-1 case), the following expression of the interferometric phase variation holds:

$$\Delta\phi = \frac{4\pi}{\lambda}\Delta(\Delta r_o) = \frac{4\pi \cdot B_n \cdot n_p}{\lambda \cdot r_o} \tag{3.3}$$

This result shows that if we know the relative displacement of the two orbits normal to the slant range direction B_n, the distance r_o and the value of the SAR wavelength λ, then the phase difference $\Delta\phi$ depends only on the n_p value, i.e. the elevation difference between the points displayed in figure 3.1, measured in the direction normal to the slant range axis. Thus the interferometric phase image represents a map of the relative terrain elevation with respect to the slant range direction. A linear term can then be subtracted from the interferometric phase so that the phase corresponding to the flat horizontal terrain is a constant. After some algebra, the version of equation 3.3 that refers to the relative elevation $q = \Delta z$ with respect to a flat terrain becomes:

$$\Delta\phi = \frac{4\pi \cdot B_n \cdot q}{\lambda \cdot r_0 \cdot \sin\theta} = K \cdot q \qquad (3.4)$$

4. Phase unwrapping

Up to now it has been assumed that the phase difference $\Delta\phi$ of equation 3.4 would be directly obtained from the interferometric fringes. However such an assumption is in some way misleading. In fact, from the interferogram, the complex values $e^{j\Delta\phi}$ can be determined but not the $\Delta\phi$ values themselves. What is measured is not the phase difference value $\Delta\phi$, but its principal value $\Delta\phi_p$, limited between $-\pi$ and $+\pi$. The interferometrical fringes show typical discontinuities like those shown in figure 4.1.



Figure 4.1: ERS-1 - Mt. Vesuvius (Italy) August 27th and September 5th, 1991. - Interferometric fringes. The estimated normal baseline is $B_n = 193$ meters.

Two adjacent discontinuities separated by a constant elevation increment, corresponding to a 2π phase drop, represent the height contours of the terrain elevation map. However, since the real phase values are " wrapped" around a 2π interval, to obtain the correct map (and, thus, the correct labels of each contour line), a map of the multiples of 2π to be added to the "wrapped" phase should be carried out (*phase unwrapping*). At the beginning of the studies on SAR interferometry (1988-89), the phase unwrapping problem was investigated assuming that only two SAR images were available. That was the actual situation at that time since the only few interferometric SAR images were from the SEASAT and the JPL single pass airborne systems. Many techniques of phase unwrapping of interferometric SAR images have been proposed [6, 7, 8]. A summary of these techniques is out of the scope of this papers. Nonetheless the following conclusion came out from these works: phase unwrapping can be either trivial or practically impossible depending on the coherence of the images (see section 8) and on the presence of foreshortening and layover (see section 5). More recently, however, a different strategy to solve the phase unwrapping problem has been considered. With the huge amount of SAR images from the European satellites ERS-1 and ERS-2 together with the availability of Digital Elevation Models (typically from the optical stereo SPOT) the basic assumption changed: all the available information should be exploited to get a DEM [9]. Then, an available Digital Elevation Model (DEM) can be transformed into the SAR geometry and subtracted from the interferogram. Depending on the DEM accuracy and on the normal baseline B_n the phase of the residue can be so small to show a single fringe with no need of phase unwrapping. On the other hand, if more interferograms are available the linear relation between the baseline and the interferometric phase (see equation 3.4) can be exploited together with the theorem of the "chinese reminder" to reduce the wrapped phase ambiguity. As a conclusion, even if a fully automatic phase unwrapper that works in every condition is not available yet, phase unwrapping does not appear to represent a serious problem in SAR interferometry anymore.

5. Digital Elevation Map preparation

The elevation map derived from SAR interferometry lies on a plane where the reference axes correspond to the azimuth and slant range directions.

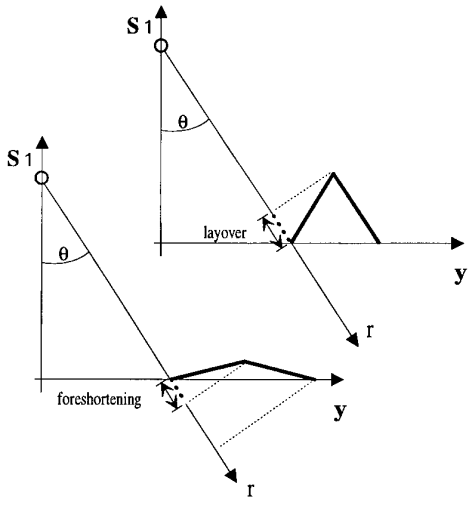


Figure 5.1: Typical geometric deformations introduced by SAR (foreshortening and layover).

Such a coordinates system is completely different from the reference systems in the azimuth-ground range plane used in the usual el-

evation maps. Thus it is necessary to transfer the unwrapped phase from the slant range coordinates system to the ground range system; the obtained values must be interpolated and resampled in terms of uniform ground range cells.

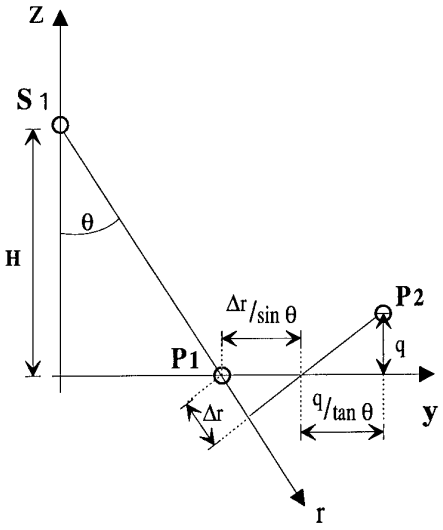


Figure 5.2: Cross-section of the SAR system geometry normal to the azimuth direction. The ground range coordinate depends on both the range position and the point elevation.

5.1. Rectification of the elevation map

From Figure 5.2 it is evident that the horizontal position of a backscatterer depends on both the slant range coordinate and the elevation. Through simple geometrical expressions the relation between these three parameters can be found. The ground range coordinate with respect to the initial point, indicated by y, is the sum of two components: the first is the horizontal displacement in the case of flat terrain, the second caused by a non-zero elevation drop.

$$y(r,q) = \frac{\Delta r}{\sin\theta} + \frac{q(r)}{\tan\theta} \qquad (5.1)$$

Since the position of the points depends on its elevation, the correspondence between ground range and slant range is quite irregular. In fact, the well known foreshortening effect causes a compression of the areas with ascending slope and a spread of the descending areas. As a consequence the ground difference, corresponding to a constant slant range displacement, will be much larger in the case of ascending slopes. Furthermore, when layover effects occur several areas of the earth's surface can disappear from the SAR image (see figure 5.1).

In a ground range reference system the obtained elevation map will have a quite uneven sampling interval. Thus to obtain a regular sampled map the elevation values must be interpolated. For our purposes a linear interpolation is quite adequate. In fact, in flat or descending areas the interpolating points are fairly close, whereas with an ascending slope the foreshortening effect produces such a large slant range compression that the interpolating points lie much further away and no interpolator would operate correctly. For example, for geometrical parameters like the ERS-1 mission, a slant range sampling $\delta r \simeq 8$m corresponds to a ground range step $\delta y = \delta r/\sin\theta \simeq 20$m for flat terrain; on the contrary for a terrain slope $\alpha = 15°$, the ground range step is $\delta y = \delta r \cos\alpha/\sin(\theta-\alpha) \simeq 60$m. The results of the rectification process, performed for the SAR image of the area near the Vesuvius is shown in figure 5.3. Contour lines correspondent to descending orbits are shown.

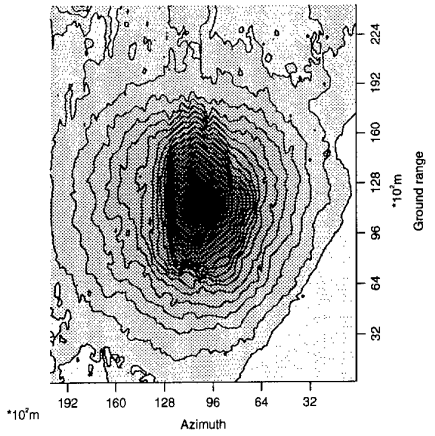


Figure 5.3: A topographic map of the area of Vesuvius obtained from the unwrapped and rectified phase, shown in the previous picture.

5.2. Combination of ascending and descending orbits data

Any current DEM estimate is very useful for unwrapping. In fact, data from different takes and therefore with different baselines can be combined, but only if properly positioned in space; in the case of a flat region, uninteresting for the unwrapping problem, the coregistration can be carried out using ephemerides. Not so in complex topography situation, where the registration is DEM dependent. We arrive to an iterative procedure where the available data are combined in such a way to achieve a progressive improvement of the DEM. In this operation, the combination of data from ascending and descending orbits could be very helpful. First, to unravel the layover areas, that, if not in shadow, may appear well behaved with the other orbit. Second, and equally relevant, the combination of the two views is only possible if the DEM is correct; hence, it is a pow-

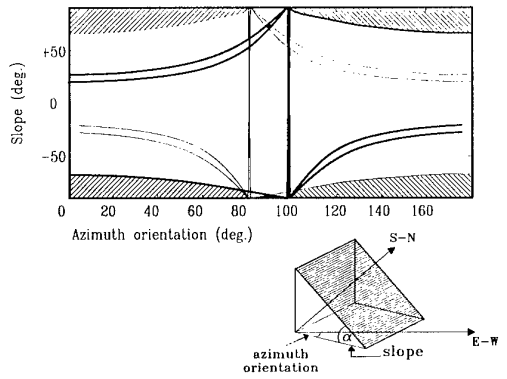


Figure 5.4: Combinations of slope and azimuth orientation that cannot be recovered from ERS-1 data. With ascending orbits (gray and shaded areas). With both ascending and descending orbits (shaded areas only). With layover separation (black areas).

erful check of its quality, to remove blunders, reduce the elevation dispersion, etc.

5.2.1. Slopes coverage improvement

Ascending and descending orbits combination is essential for slopes coverage in SAR interferometry. As an example, the different curves of figure 5.4 identify terrain slopes α recoverable from ERS-1 as a function of their azimuth orientation (see the lower part of figure 5.4). Thick and thin lines refer to ascending and descending passes respectively (here a baseline of 100 meters has been assumed both for ascending and descending passes). If just ascending passes are used, slopes that can be recovered from ERS-1 (at the equator in this example) are contained in the white area. To understand how that plot has been generated, let us consider a 9 degrees azimuth orientation (i.e. the orbit orientation) and let us consider slopes ranging from -90 to $+90$ degrees (i.e. from the lower to the upper limits of the plot). It is clear that, if we consider an incident angle of 23 degrees for flat terrain, slopes from -90 to -67 degrees will not be observed since in shadow (lower shaded areas). On the other hand, slopes from 20 to 26 degrees will not be observable since the interferometric images are uncorrelated [5] (this is the region between the two thick lines). Finally, slopes from 26 to 90 degrees are in layover (upper gray and shaded area). These simple considerations can be repeated for every azimuth orientation and for descending ERS-1 orbits. Then, if both ascending and descending passes are used, the unrecoverable areas are represented by the shaded areas only. Those areas affected by foreshortening and layover in one image are well covered (if not in shadow) in the other image. Finally, if layover areas could be separated from the rest of the image (an operation not always possible), the irrecoverable slopes are further reduced to the small black areas shown in figure 5.4.

The results of the rectification process, performed for the SAR image of the area near the Vesuvius using both ascending and descending orbit data are shown in Figure 5.5.

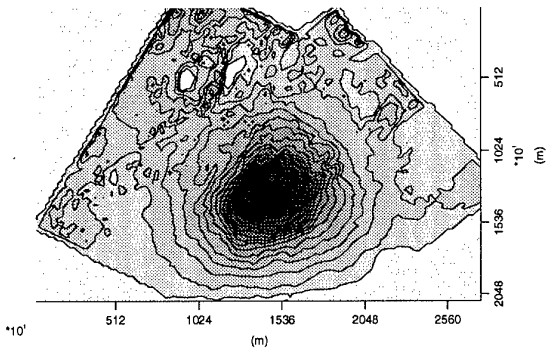


Figure 5.5: A topographic map of the area of Vesuvius obtained from the unwrapped and rectified phase using both ascending and descending orbits.

6. Monitoring terrain changes

If the SAR acquisitions are not simultaneous, the interferogram phases (*fringes*) are also affected by the possible terrain changes Two types of temporal terrain evolution can be identified by means of multitemporal interferometric SAR images: decorrelation changes and small terrain motions.

Decorrelation changes. A random change of the position and physical properties of the scatterers within the resolution cell can be detected as low coherence of the images and will be dealt with in section 8.

Small terrain motions. Centimetric slant range relative motions of local areas that occur between two SAR acquisitions, generate large local phase shifts in the interferogram. The phase difference $\Delta\phi_s$ generated by small terrain relative motions is governed by a mathematical relation completely different from that of the interferometric fringes (that, in order to avoid any confusion we shall indicate with $\Delta\phi_T$) described by equation 3.3. It is, in fact, proportional to the ratio between the relative motion Δs along the slant range direction and the transmitted wavelength λ. Thus, if we have a non simultaneous interferometric SAR pair with a given baseline B_n and a small terrain relative motion occurs between the two acquisitions, the following expression of the interferometric phase difference holds:

$$\Delta\phi = \Delta\phi_T + \Delta\phi_s = \frac{4\pi B_n n_p}{\lambda r_0} + \frac{2\Delta s}{\lambda} \qquad (6.1)$$

From equation 6.1, it is clear that the two terms $\Delta\phi_T$ and $\Delta\phi_s$ should be separated in order to recover the terrain relative motion. The simplest way to estimate small motions consists of choosing an image pair with a very small baseline (zero or a few meters would be sufficient in the case of ERS-1) so that the first term is much smaller than the second. An interesting example of such a solution is shown in figure 6.1. Two ERS-1 images of the area of Nice on the border between France and Italy, with a time interval of 9 days and a baseline B_n = 6m, have been used to generate an interferogram. In figure 6.1 the interferometric fringes are shown. Due to the very small baseline the fringes do not show rapid variations (even if the topography of that area is not flat at all) but for a small area close to the center of the map. In that area a very active landslide has being monitored by the group of the Institute de Physique du Globe (IPG) in Paris. From the ERS-1 interferometric data, an average landslide velocity of about 1cm per day has been estimated. The result is in good agreement with the data provided by the IPG group. Moreover, the ERS-1 interferometric SAR image provides a set of measurements of the landslide relative motion on a very dense grid (4×20 meters) at a cost that is much smaller than that of any other traditional techniques. The accuracy of the motion measurement offered by such a technique has been recently tested with an ERS-1 experiment (the "Bonn experiment"), where elevation change as small as 9 mm of a few artificial radar targets (corner reflectors) has been detected with no ambiguity [10, 11, 12].

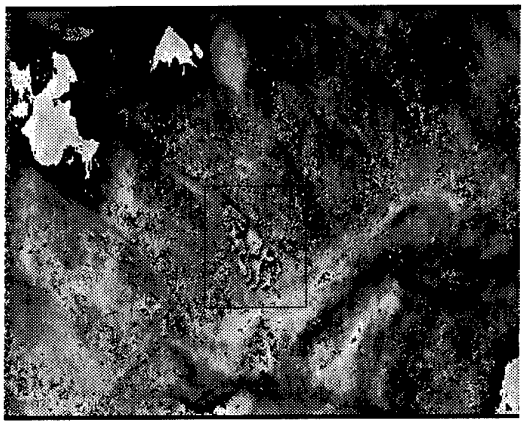


Figure 6.1: ERS-1 SAR interferometry. Fringes generated from two images of the area of Nice on the border between France and Italy, with a time interval of 9 days and a baseline B_n = 6m.

If a stereo SAR pair with a very small baseline is not available, the topography contribution to the interferometric phase ($\Delta\phi_T$) must be subtracted from the fringes. It can be accomplished in two different ways. If a topographic map of the area of interest is available, it has to be transformed in the azimuth slant range coordinates and scaled proportionally to the baseline of the interferometric pair in order to have an estimate of the phase component $\Delta\phi_T$. Then, it should be subtracted from the fringes (an impressive example of this technique has been shown at CNES [13] for the Landers earthquake). As an alternative, an additional SAR image can be exploited in order to have an interferometric SAR pair with no terrain changes [14]. For the sake of simplicity, let us label the three SAR images with 1, 2 and 3 with no regard to their time consecution. Let us also assume that during the first two images no terrain changes occur. Thus, the fringes generated from the first two images will be proportional to the phase $\Delta\phi_t$ to be subtracted from the fringes generated either from the couple $1-3$ or $2-3$. From equation 3.3, it is clear that the proportionality coefficient is given by the ratio of the two baselines B_{n12} and B_{n13} (or B_{n23}). However, since the proportionality holds on the phases and not on their principal values[1], the fringes obtained from the couple $1-2$ must be first unwrapped and then scaled.

7. Atmospheric effects

In repeated pass SAR interferometry from satellite, different propagation velocities along the scene due to atmospheric changes (at the time of the two surveys) could be responsible of interferometric phase variations that cannot be related neither to the topography nor to relative terrain motions. The above said effects can be clearly seen on the interferometric phase shown in figure 7.1.

This phase image has been generated from two ERS-1 images of the black forest near Freiburg (a quite stable area) taken on August 17 and 20, 1991. The normal baseline is 9 meters (i.e. a 2π phase change corresponds to more than 1000 meters relative elevation). The phase variations with very low spatial frequency visible in figure cannot be associated either to the topography or to terrain deformations and might be generated by evapotranspiration phenomena, atmosphere or soil moisture changes within the two surveys. These effects appear to be the major limitation to the use of SAR interferometry as a technique for generating highly accurate Digital Elevation Models and for detecting small surface deformations. However some considerations should be made at this regard showing that in both applications these parasitic effects can be strongly reduced if many independent interferograms are available. Let us first consider the DEM generation case. The interferometric phase depends on the relative elevation through a coefficient that is directly propor-

[1] Apart from the special case of an integer ratio between B_{n12} and B_{n13} (or B_{n23})

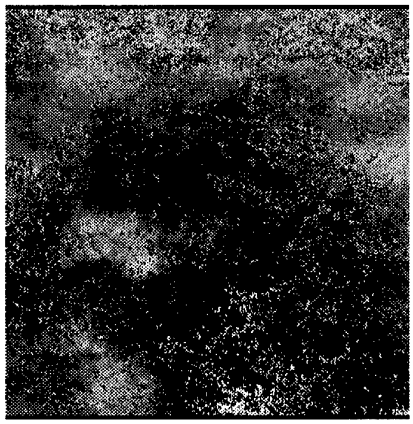


Figure 7.1: Phase image generated from two ERS-1 images of the black forest near Freiburg (a quite stable area) taken on August 17 and 20, 1991. The normal baseline is 9 meters.

tional to the baseline (see equation 3.3): the higher the baseline, the higher the phase variation correspondent to the same topography. On the other hand, the phase variations due to atmospheric changes are independent of the baseline. Thus, the higher is the baseline of the interferometric pair the smaller is the topographic error due to parasitic effects. Moreover, if many interferometric pairs of the same area are available (as in the case of ERS-1/ERS-2 Tandem mission) "outliers" can be identified and eliminated from the database. The remaining results can be combined to "filter out" the effects of atmospheric changes. The availability of an interferograms database can be also exploited to improve the measurements accuracy of surface deformations. In this case the interferometric phase variations due to surface deformation is independent of the baseline. However, if a model of the deformation development with time is available, the interferometric images can be used to identify the parameters that fit the model to the time series. Clearly, the larger the available interferograms time series the more accurate the parameters estimation. Also in this case outliers can be identified and eliminated. As a conclusive remark of this section, it should be pointed out that an attempt of modeling the phase variations due atmospheric changes is being carried out at CNES [13]. A simplified model has been supposed where the interferometric phase variation is the sum of two terms: the phase variation due to the different travel paths in two atmospheres with a different refraction index (there assumed to vary linearly with the altitude) and phase variations due to localized atmospheric changes (i.e. clouds, fog ...). Due to the first contribution, a moderate phase variation among interferograms (linearly dependent on the elevation) is generated. This effect can be strongly reduced if many interferograms are available. On the contrary, the second contribution is unpredictable and can be so strong to make the interferometric pair useless. In this case such an interferometric pair should be eliminated from the database as an outlier.

8. Coherence

8.1. Generalities

The quality of the interferometric phase depends on the amount of noise that, in general, comes from distinct sources [2, 15]: i- system noise; ii- terrain change (non simultaneous acquisitions); iii- images misregistration; iv- approximate and unequal focusing of the two passes; v- decorrelation due to the baseline ("geometric" decorrelation). It is obvious that there is no way to avoid the first two sources of noise. On the other hand, as far as the last three sources are concerned, they can be taken under control. In other words, since in most cases the system noise is quite small compared with the usually sensed signals, and the processor noise is well under control if it is designed to be **phase preserving** [2], it will be shown that the fringes quality is degraded by scattering change in time and volumetric effects (see section 9.2) only. The coherence γ of two complex SAR images v_1 and v_2, is defined as follows [3]:

$$\gamma = \frac{E[v_1 v_2^*]}{\sqrt{E[\mid v_1 \mid^2] E[\mid v_2 \mid^2]}} \tag{8.1}$$

where E[.] means the expected value (that in practice will be approximated with a sampled average) and * the complex conjugate. The absolute value of γ is a fundamental information on the exploitability of SAR interferograms. The signal (usable fringes) to noise ratio Ξ can be usefully expressed as a function of the coherence:

$$\Xi = \frac{\mid \gamma \mid}{1 - \mid \gamma \mid} \tag{8.2}$$

Thus, it is clear that every effort should be dedicated to avoid coherence loss during the interferogram generation process (section 8.1).

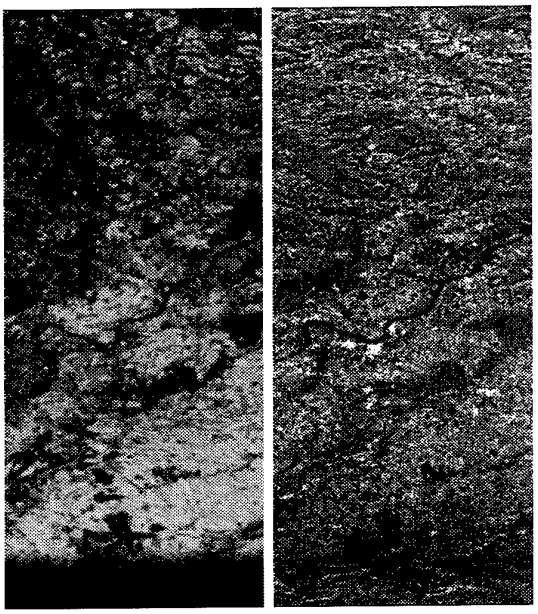
The statistical confidence of the estimated coherence (sampled coherence) and of the derived measurements, depends on the number of independent samples (n) that can be combined for the computation. As a first approximation, the standard deviation of the estimator is proportional to $1/\sqrt{n}$. Thus, whenever uniform areas (in the statistical sense) are identified, the sampled coherence can be computed as:

$$\gamma = \frac{\sum_n v_1 v_2^* e^{-j\phi}}{\sqrt{\sum_n \mid v_1 \mid^2 \sum_n \mid v_2 \mid^2}} \tag{8.3}$$

In fact, since the coherence is estimated from the combination of the phases of a few pixels at the very least, the topography effects on the interferometric phase ϕ proportional to the known terrain changes have



Figure 8.1: Coherence map of Mt. Vesuvius.



a b

Figure 8.2: **a** - Coherence map of the Bonn area. The Rhein river is clearly visible crossing from left to right the image. **b** - ERS-1 detected image of the same area.

to be removed from the result. Thus, in order to compensate this unwanted effect, the vectors at the numerator of equation 8.3 must be *deskewed* before summing. It is also clear that, in order to generate an interferogram, the pixels of the images gathered in the two different images must be registered accurately, so that the random variates corresponding to the reflectivity are properly aligned. A single pixel shift, if the focusing processor is a good one, is enough to practically zero the correlation. In the following we will not consider the effects due to misregistration and system noise, since they can be avoided with a good system or with a proper processing. The elevation error σ_n of maps generated by means of SAR interferometry will follow the value of Ξ as:

$$\sigma_n = \frac{\lambda r_0 \sin\theta}{2\pi n_s}\sqrt{\frac{1}{2\Xi}}$$

As an example, the coherence map of the area of Mt. Vesuvius in Italy observed from ERS-1 on August 27th and September 5th 1991 is shown in figure 8.1 (see also the fringes obtained with the same passes that are shown in figure 4.1). The sampled coherence has been computed on small rectangles 16×4 (azimuth, slant range) pixels large ($n = 64$) that reasonably belong to uniform areas. The coherence map can be converted into an elevation error map (apart from systematic errors). As expected, areas covered by thick vegetation or in foreshortening or layover show an almost zero coherence and are not usable for SAR interferometric applications (e.g. DEM generation and super-resolution).

8.2. Applications to image segmentation

SAR coherence is an additional source of information with noticeable diagnostic power. In the following we shall enumerate some of the most relevant applications. In [12] it was first observed that forests, that appeared with variable reflectivity in the ERS - 1 detected images, appeared almost black in the coherence images: this effect is due to the scarce penetration of C band radiation in the vegetated canopy, so that small variations of the positions of leaves and smaller branches were enough to change the disposition of the scatterers and therefore practically annihilate coherence; likewise happens for water bodies, that appear always with negligible coherence. In [16, 17, 18] it was also observed that cultivated field changed their coherence after plowing, harvesting etc., so that it was possible to detect anthropogenic effects in multitemporal sequences of takes of the same area, by looking at sudden coherence losses. In general, the combination of multitemporal observation both of detected images and coherence allows a very good segmentation of agricultural areas; it is thus possible to identify cultures (potatoes are harvested in that month, whereas corn matures in that other ..). The coherence map, achieved from interferometric quick-look processing, generated from two ERS-1 images of the Bonn area taken 3 days apart in time (March 8th and March 11th 1992) is shown in figure 8.2. Notice the black curve (low coherence) that corresponds to Rhein river crossing from left to right figure 8.2. For comparison, also the detected image is shown in figure 8.2. Other authors [19] observe that from the phase of the interferometric takes the height of the trees and therefore the biomass can be estimated.

8.2.1. Identification of exposed rocks

The coherence image of Mt. Vesuvius 8.1 shows interesting features that are not visible in the detected image of the same area. Apart from the already mentioned zero-coherence areas, a very bright stripe (high coherence) is visible on the left hand side of Mt. Vesuvius. This area that is not identified on the absolute value image, has a geological peculiarity: it is an old lava flow (year 1944). Since exposed rocks are coherent even after years, such areas can be easily identified by using SAR coherence.

9. The wavenumber shift approach

The above described monochromatic approximation, notwithstanding its simplicity, hides an important aspect of the interferogram generation mechanism that will be discussed in this section. Let us consider the approximated relation between the frequency f and the ground wavenumber k_y (along the range direction y as shown in figure 9.1). It can be easily derived by projecting the transmitted wavelength λ, measured along the slant range direction, on the terrain. For sake of simplicity we shall suppose that the terrain has a constant slope α. Thus, as a first approximation, the wavelength λ projected on the terrain is scaled by the factor $1/\sin(\theta-\alpha)$. Then, the factor 2 accounting for the two ways travel path, the following expression of the ground wavenumber holds:

$$k_y = \frac{4\pi}{\lambda}\sin(\theta-\alpha) = \frac{4\pi f}{c}\sin(\theta-\alpha) \qquad (9.1)$$

Thus, in general, a looking angle difference $\Delta\theta$ generates a shift and a stretch of the imaged terrain spectra. However, if the relative system bandwidth is small, the frequency f in the second term of equation 9.1 can be substituted with the central frequency f_o. The stretch can be neglected and the following equation holds.

$$\Delta k_y = \frac{4\pi f_o \Delta\theta}{c}\cos(\theta-\alpha) \qquad (9.2)$$

Finally, since the radar is not monochromatic (we have a bandwidth W centered around the central frequency f_o), we can conclude that by changing the looking angle of the SAR acquisition, we get a different band of the ground reflectivity spectrum. Now, in order to compare the shift of the ground reflectivity spectrum with the SAR bandwidth W, it is worth to express the ground wavenumber shift of equation 9.2 as an equivalent frequency shift Δf. The following expression of Δf for an angular separation $\Delta\theta$ can be obtained by differentiation of equation 9.1 directly,

$$\Delta f = -\frac{f_o \Delta\theta}{\tan(\theta-\alpha)} = -\frac{cB_n}{r_o \lambda \tan(\theta-\alpha)} \qquad (9.3)$$

We would like to stress again that equation 9.3 does not state that by changing the looking angle of the SAR acquisition the radar bandwidth is shifted by Δf. It just says that by changing the SAR looking angle, the backscattered signal contains different spectral components of the ground reflectivity spectrum. In other words, if we look at the signals received by two SARs separated by an angle $\Delta\theta$, equation 9.3 states that the same spectral components of the first signal are found in the second spectrum shifted by Δf. In figure 9.1 the wavenumber shift principle is sketched.

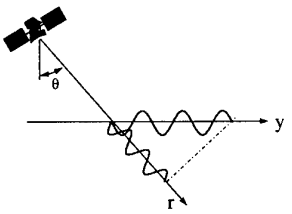


Figure 9.1: The wavenumber shift principle. By changing the looking angle of the SAR acquisition we get different wavenumbers of the ground reflectivity spectrum.

Different bands of the ground reflectivity spectrum (limited by the finite bandwidth of the SAR systems) observed from slightly different looking angles are shown. Notice that whenever the spectral shift Δf is smaller than the system bandwidth W, we have that only part of the ground reflectivity spectrum is common to both spectra. In the following, we shall refer to that as the "common band". In the case of distributed random scatterers, there can be a correlation of the two return signals only if there is such a common band; in such a case, by considering equation 9.3, it is clear that the spectral shift Δf between the two images can be measured and, if we know the geometric parameters B_n, θ and r_o, the local terrain slope α can be recovered. Then, by integrating the local terrain slope, the terrain topography can be derived. From equation 9.3, the expression can be easily derived of the critical baseline for which the two spectra become totally disjoint, the correlation of the two return signal goes to zero and the spectral shift cannot be computed. As an example, the ERS-1 critical baseline for flat terrain ($\alpha = 0$) can be computed is 1100 m.

9.1. Range resolution improvement

The range resolution achievable from a single SAR complex image is limited by the system bandwidth [3, 4]:

$$\rho_r = \frac{c}{2W}$$

However, we will show how the image bandwidth (and so the resolution) can be enhanced by combining multiple surveys from different view angles. As an example, in the case of two surveys, the spectral shift caused by

a change in the view angle of $\Delta\theta$ could give a single image with an enhanced bandwidth: $W_e = W + \Delta\omega/(2\pi)$.

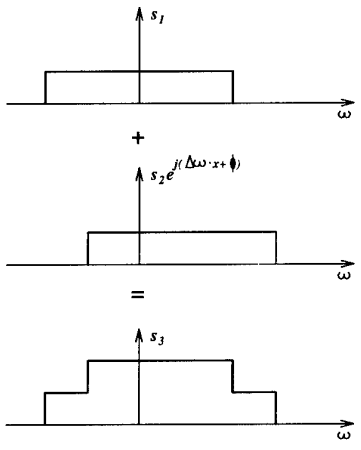


Figure 9.2: The wavenumber shift principle is used to improve slant range resolution.

This is shown in figure 9.2: it can be noted that the parts not common to the two range spectra, that were filtered out in the interferometric processing, now represent the innovation present in one image with respect to another. It's worth to note that the resulting band expansion, and hence the improvement of resolution in slant range is not constant in all the image points, but it is a function of the baseline and of the local slope angle of the ground α. In the points where α is positive, the spectrum shift and the obtainable improvement are maximum. The processing required to get the super-resolution image, i.e. the coherent superposition of the surveys, requires the compensation for each scatterer, of the phase rotation due to the travel path distance between the scatterer and the couple of interferometric sensors.

9.2. Volume scattering

Up to now, we have considered only surface scattering, neglecting the effect of volume scattering. However, when backscattering comes from targets with different elevation z within the resolution cell, volume effects cannot be ignored. Two SAR images from different angles (θ and $\theta + \Delta\theta$) represent different bands of the complex reflectivity spectrum. From figure 9.1, it is clear that, in order to observe the same k_y from the two acquisitions (i.e. the condition that allows coherent interferometry in case of surface scattering), the looking angle change $\Delta\theta$ should be compensated by the already mentioned spectral shift

$$\Delta\omega = -\omega\Delta\theta/\tan(\theta)$$

that is equivalent to the following k_z wavenumber shift

$$\Delta k_z \simeq -2\Delta\omega/(c\cos\theta) = 2\omega\Delta\theta/(c\sin\theta).$$

The coherence of interferometric images depends on the degree of spectral correlation between the measurements in k_z and $k_z + \Delta k_z$. If we consider, as a special case, the flat box ($\Delta z = 0$), the signal is impulsive in the z domain and we have total correlation of the different spectral components and unitary coherence. In general, with a finite volume and a non zero value of Δz, we shall have the following expression of the spectral correlation function:

$$\frac{\sin(\Delta k_z \Delta z/2)}{\Delta k_z \Delta z/2}$$

So, we can state that whenever $\mid \Delta z \mid \ll 2\pi/ \mid \Delta k_z \mid$, volumetric effects can be neglected. A useful expression can be found as a function of transmitted wavelength, normal baseline and satellite altitude:

$$\mid \Delta z \mid \ll \mid \frac{\lambda H \tan\theta}{2B_n} \mid = \mid \Delta z_o \mid$$

If we consider the case of $\mid \Delta z \mid \geq \mid \Delta z_o \mid$, we have spectral decorrelation and low coherence. In the ERS-1 case with a baseline of 250 meters decorrelation is expected for $\mid \Delta z \mid \geq$ 38meters. From those observations, it is then possible to derive a LMS estimate of the surface and the volumetric contribution. These concepts have been further tested by means of the anechoic chamber at JRC-EMSL [20].

10. Interferometry from airplanes

Apart from satellites, airplanes are the other source of interferometric data: in this case, however, the instability of the platform makes it practically very difficult to achieve multi pass interferometry; in the single pass technique, the two antennas are placed on board at the maximum distance compatible with the airframe to increase the baseline. The coherence of the interferometric picture is basically equal to 1, so this is not a source of information. On the other hand the quality of the fringes is topmost and therefore the quality of the DEM that can be generated very high. There are still problems with layover, but reduced due to the fact that the off nadir angle can be higher than 45^0, and at times close to 80^0. Then shadows become important: In that case the coherence drops to 0, and there is no signal. The real problem with airplane data is again the instability of the platform and the uncertainty of the Doppler centroid and of the actual value of the baseline, that depends on the airplane position. The spatial resolution can be as high as $0.5 \times 0.5\, m$ [21]. The vertical resolution can be also very good due to the high quality of the fringes and to the very good signal to noise ratio. A big difference with the satellite is the swath, that can never be much beyond a few kilometers. This makes airplane surveys more expensive than those from satellites. The other application of airplane interferometry is along track instead of across track; in that case, the two antennas are displaced along the

flight direction, so that the image taken is the same, but for the objects that moved in the time interval correspondent to the sensor displacement (say 20 $msec$). Then it is possible to see that movements along range at the speed of say $1\,m/sec$ correspond to fringes of about a wavelength in C band and therefore are very well visible. This leads to another very important application of interferometry [22] i.e. the measurement of sea currents.

11. The complete interferometric chain

In the case of multi pass interferometry a good control of the platform position and of the central frequency of the radar is needed to get good pictures; then the data should be processed with a phase preserving processor, i.e. without distorting or adding space varying phase shifts induced by inaccurate processing. Almost all processors can do as long as they are designed for that; we remember here the $\omega - k$ processor [2] but also range Doppler ones can be made to work [23]. Once the data are focused, then the two images have to be properly aligned in space; the precision of the alignment should be about .1 pixel or better; this is achieved by optimizing the quality of the fringes and using very high quality interpolators [24]; it is also possible to design the focusing processor to properly stretch the images avoiding too many interpolations [25]. The fringes thus generated have to be unwrapped: the problem can be solved either with a single very high quality image pair (the airplane case) or if we have several lower quality image pairs with different baselines (the satellite case); a precedent DEM is always useful. Finally the image has to be geocoded and if possible merged with other estimates of the DEM; but then this type of processing is not specific for interferometry anymore.

12. Actual and future missions

Many airborne systems have been designed to meet interferometric needs (AIRSAR - JPL; ESAR, DLR; EMISAR, TUDenmark; CV 580 - CCRS); satellite interferometry was found [1, 12] because the stability of the orbits and of the clocks were much beyond the specs. Recently, after many years of positive results with ERS -1, the first mission is flying that was designed with interferometry as one of the main goals, namely the tandem mission where ERS - 1 and ERS - 2 satellites of the European Space Agency revisit the same area after 1 day (revisiting delays of about 30 min. and of 8 days could also be envisaged); the quality of the images is very good. It is also important to mention the SIR - C mission, where the final days were devoted to multifrequency interferometry (L, C, and X band), with one day revisiting time, with results similar to that of the Tandem mission, with advantage of the multifrequency and the disadvantage of a limited duration of the mission and therefore a limited coverage of the globe. More satellite missions are envisioned: first ENVISAT, where a SCANSAR system will be available (the off nadir angle can be changed during the observation time), but then a mission with the shuttle where two C band and may be X band receivers will be available at the ends of a 60 feet boom, to achieve single pass interferometry from the satellite. This mission will merge the advantages of airplane and satellite interferometry, namely wide swath and unitary coherence. Further ahead in time, satellite fleets are being studied; the concept of the tandem mission will be exploited, i.e. the cooperation of two or more satellites to have images at a short time distance; the satellites may be staggered in frequency to have more resolution, in space to have the usual interferometry, in time to have along track interferometry, etc.

13. Conclusions

In this presentation we have tried to summarize the impact of interferometry on the design of SAR surveys, its possibilities and limits. First, interferometry is speckle free, since its effect disappears from the differential phase; the concept of radiometric resolution thus loses meaning. Further, we have seen that from the fringes, given short or zero revisiting times, a very good DEM is obtainable, with a vertical resolution that could be in the meter range or even less from airborne platforms. Millimetric motion of large areas of the terrain or of corner reflectors have been measured with good reliability and therefore the possibility of using the system to measure subsidence, landslides, coseismic motion has been demonstrated. However, several conditions have to be met, the most important being the maintenance of some coherence of the scatterers during the entire experience. This is possible in the case of exposed rocks, but in other cases artificial reflectors may be needed to link together scattering structures that may change due to vegetation, floods, storms, etc. Coherence is an important clue, that combined with the more usual backscatter amplitude, leads to high quality images segmentation. The wavenumber shift concept, finally, may lead to a 3 dimensional view of the terrain, provided that sufficiently low radio frequencies be used, that penetrate the target's internal structure. The same principle can be invoked to improve the resolution of objects that do not change with time, seeing them under different angles and therefore using the spotlight mode both in range and in azimuth. We have also presented the state of development of SAR interferometry. Codes have been developed that carry out automatic images registration (up to 1/100 of a pixel), fringe prefiltering and slope driven smoothing. Automatic phase unwrapping is not an easy goal with noisy or difficult images. However, the high quality of the tandem mission makes it likely that global DEM might be available soon in ERS - SAR geometry, so that only differential interferometry data with small phase deviations should be considered. However, both ascending

and descending passes are necessary to avoid foreshortening and layover effects. The combination of multiple images will allow improved resolution.

References

[1] Zebker, H. Goldstein R., 1986, Topographic Mapping from SAR Observation, Journal of Geophysical Research, Vol. 911, pp. 4993 - 4999.

[2] Prati C., Rocca F., Monti Guarnieri A., Damonti E., 1990, Seismic migration for SAR focusing: Interferometrical applications, IEEE Transactions on GARS, Vol.28, N.4, pp.627-640.

[3] Prati C., Rocca F., Range resolution enhancement with multiple SAR surveys combination, Proc. IGARSS'92, Houston, May 1992.

[4] Prati C., Rocca F., 1993, Improving slant range resolution of stationary objects with multiple SAR surveys, IEEE Transactions on AES, Vol. 29, No.1, 135-144

[5] Gatelli F., Monti Guarnieri A., Parizzi F., Pasquali P., Prati C., Rocca F., 1994, Use of the spectral shift in SAR interferometry: applications to ERS-1, IEEE Trans. on GARS, Vol. 32 , No 4, July 1994, pp.855-865.

[6] Goldstein R.M., Zebker H.A., Werner C.L., 1988, Satellite radar interferometry: Two-dimensional phase unwrapping, Radio Science, Vol.23, N.4, pp. 713-720.

[7] Giani M., Prati C., Rocca F., 1992, "SAR interferometry and its applications", ESA report N.8928/90/F/BZ.

[8] Spagnolini U., 1993, 2-D Phase Unwrapping and Phase Aliasing, Geophysics, Vol.58, N.9, pp.1324-1334.

[9] Tarayre H., 1996, Extraction de modele numeriques de terrain par interferometrie radar satellitaire: algorithmie et artefacts atmospheriques, These de doctoeur de l'I. N. Polytechnique de Toulouse.

[10] Hartl Ph., Reich M., Thiel K., Xia Y., SAR Interferometry applying ERS-1 - Some preliminary test results, Proc. First ERS-1 Symposium, Cannes, 4-6 November 1992, pp. 219-222.

[11] Prati C., Rocca F., Monti Guarnieri A., SAR interferometry experiments with ERS-1, Proc. First ERS-1 Symposium, Cannes, 4-6 November 1992, pp. 211-218.

[12] Monti Guarnieri A., Parizzi F., Pasquali P., Prati C., Rocca F., Developments in ERS-1 SAR interferometry, Proc. of 1st Workshop FRINGE, ESA-ESRIN, Frascati, 12 Ottobre 1992.

[13] Massonnet D. et al., The Displacement Field of the Landers Earthquake Mapped by Radar Interferometry, Nature, Vol.364, July 8, 1993, pp.138-142.

[14] Gabriel A.K., Goldstein R.M., Zebker H.A., Mapping small elevation changes over large areas: differential radar interferometry , J.G.R., Vol.94, N.B7, July 1989, pp.9183-9191.

[15] Rodriguez E., Martin J.M., 1992, Theory and design of interferometric synthetic aperture radars, IEE Proceedings-F, Vol.139, No.2, pp.147-159.

[16] Prati C., Rocca F., Monti Guarnieri A., 1994, Topographic Capabilities of SAR exemplified with ERS-1, Geo-Information-Systems, Vol. 7, No.1, February 1994, pp.17-22.

[17] Werner C., Hensley S., Goldstein R.M., Rosen P.A., Zebker H.A., Techniques and applications of SAR interferometry for ERS-1: Topographic mapping, change detection, and slope measurement, Proc. First ERS-1 Symposium, Cannes, 4-6 November 1992, pp. 205-210.

[18] Wegmueller U., and Werner C. L., 1995, Farmland monitoring with SAR interferometry, Proc. IGARSS 1995, Florence, Italy. pp. 544 - 546, 1995b.

[19] Ulander L. M. H., Dammert, P. B. G., Hagberg, J. O., 1995, Measuring tree height using ERS - 1 SAR Interferometry, Proc. IGARSS 1995, Florence, Italy.pp 2189-2191, 1995c.

[20] Fortuny J., Holmer E., Sieber A. J., Pasquali P., Prati C., Rocca F., 1994, Validating SAR interferometry applications by using EMSL, Proc. of IGARSS'94, Pasadena (CA), pp.736-738.

[21] Moreira J., 1996, personal communication.

[22] Gray L., Van der Kooj M., Mattar K., Farris Manning P., Progress in the development of the CCRS along - track interferometer. Proc. IGARSS 1994, Pasadena, CA, pp. 2285 - 2287, 1994d.

[23] Moreira J., et al., 1995, X-SAR Interferometry - First Results, in IEEE Trans. on GARS, vol 33, no. 4, pp 950–956, 1995.

[24] Gatelli F., Monti Guarnieri A., Prati C., Rocca F., 1996, Medium Resolution Efficient Phase Preserving Focusing for Interferometry, in Proc. IGARSS' 96 (Lincoln, Nebraska,) 27-31 May, 1996.

[25] Fornaro G., Franceschetti G., Marzouk E. S., 1994, A new approach for image registration in interferometric processing, in Proc. IGARSS' 94, Pasadena, CA, Aug. 94, Vol. 4, pp. 1983–1985.

Paper 2

F. Lanzl (GE)

With respect to reduction of speckle, I would like to know:

a) How terrain acts in the case of multi-directional observation?

b) If the spotlight mode leads to speckle reduction?

Author's reply:

a) In speckle reduction techniques, multi-looking or use of sub-apertures, the change in direction angle is fairly small (1-3 degrees). In general, the terrain reflectivity does not change over this range. True multi-directional observation may lead to improved terrain knowledge, provided the change in direction angle is sufficiently large, say more than 20 degrees. Since the terrain reflectivity is now changing, the measurements cannot be used for speckle reduction.

b) In spotlight mode speckle reduction can be obtained in the same way as in the strip-map mode before, by the incoherent summation of sub-apertures. Again the radiometric resolution is improved at the cost of geometric resolution. Direction angle change in spotlight mode is limited to approximately 10-15 degrees.

J. Shi (US)

What are the critical areas for future real-time SAR processor development?

Author's reply:

Critical areas are miniaturisation, reduction of power consumption and software programming flexibility. Processing speed is not so much an issue anymore for airborne systems.

Spaceborne real-time processors will need more time to develop due to the need for space qualification of the newly developed components.

SAR Systems and Related Signal Processing

P. Hoogeboom, R.J. Dekker and M.P.G. Otten
TNO Physics and Electronics Laboratory
P.O. Box 96864
2509 JG The Hague
The Netherlands

1. SUMMARY

Synthetic Aperture Radar (SAR) is today a valuable source of remote sensing information. SAR is a side-looking imaging radar and operates from airborne and spaceborne platforms. Coverage, resolution and image quality are strongly influenced by the platform. SAR processing can be performed on standard computers or dedicated hardware architectures, depending on required throughput and image quality. Besides imaging, SAR can perform other modes of operation and applications such as multi-frequency and polarimetric imaging, elevation mapping and moving target indication (MTI). Current developments in SAR systems are concentrated on higher resolution, and flexibility in performing advanced modes in the future.

2. INTRODUCTION

Synthetic Aperture Radar (SAR) is nowadays a well known source of remote sensing data. Space- and airborne SAR systems, with a range of characteristics, are available to provide information. Since SAR radiates and detects microwave radiation, it is able to make images under most weather circumstances, day or night. For these reasons SAR is a valuable remote sensing instrument in both militairy and civil information gathering.

SAR is a side-looking imaging radar instrument; it scans the surface of the Earth and focuses by coherently detecting the radar return and integrating over a distance flown, called the synthetic aperture. Another advantage of SAR compared to real-aperture systems is that the antenna can be smaller, which makes space-applications possible. Owing to the coherent detection process, SAR images also contain clearly visible speckle-noise, compared with images obtained from optical sensors. This noise is inherent to the system and can only be suppressed by averaging incoherently.

SAR has various and sometimes exclusive applications. Some are mapping, land classification, and change detection. All of these tasks can be done with both radar and optical instruments which can provide complementary information. Other, radar restricted applications are mapping of the sea-bed, interferometry and Moving Target Indication (MTI).

This article will give an overview of the state-of-the-art in present SAR systems, the related signal processing, their capabilities and future. The next section treats air- and space-borne SAR systems and the technology used today. Section 3 gives a general description of SAR signal processing and specific parts for air- and spaceborne systems, including real-time processing. Section 4 discusses several special SAR modes and applications. Section 5 closes with future developments.

3. AIRBORNE AND SPACEBORNE SAR

The platform is an important part of the SAR system. It strongly impacts its general characteristics. Two types of platforms can be distinguished, aircraft and spacecraft (satellite or space-shuttle). The major differences between the two types of systems are range, swath, resolution, platform motion and availability. For illustration, table 1 shows the characteristics of the European satellite SAR system ERS-1 [Vass 1992] and the Netherlands aircraft SAR system PHARUS [Hoogeboom 1992, 1994].

The table shows the extreme difference in range and the width of the scanned area (swath). Recording in along-track (azimuth) direction is unlimited, in principle. A long range SAR needs more power and a longer antenna because of free-space loss and smaller antenna beamwidth. Note that the need for smaller beamwidth is prescribed by the antenna-gain and not by resolution. The azimuth resolution of SAR is mainly determined by the length of the synthetic aperture, instead of the footprint of the antenna (real aperture). The synthetic aperture is formed by flying the antenna over a certain distance. Platform motion behaviour is very important: spacecraft are extremely stable platforms; they can repeatedly scan areas from almost identical tracks (repeat-pass imaging). Aircraft are much less stable. However, because of motion compensation techniques, focus and geometry of SAR images can be accurately reconstructed. It is also clear that satellite systems are less maneuverable than aeroplanes, and that they are less versatile. The suitability of a SAR system is very dependent on the task or application. For any task the best trade-off has to be found between the size of the area to cover and the resolution. For repeated monitoring of the same areas, for instance, satellites are extremely suitable. For more flexible gathering of SAR data, aircraft are often more suitable.

Many technologies are used in building synthetic aperture radar systems. Classical components that are still being used are traveling-wave-tubes (TWT) for the transmitters and waveguide antennas. The ERS-1 satellite, for instance, has been built with TWT's for reasons of robustness. New technologies such as monolithic microwave integrated cicuits (MMIC) and phased array antennas are emerging for the purpose of smaller and more flexible systems. Phased array antennas have the advantage of non-mechanical beam steering, which increases the flexibility of the system. The PHARUS system is equipped with MMIC's and an active phased array antenna. Obviously, there are many more systems, both satellite and aircraft, available. To name a few, the successor of the ERS-1, the ERS-2, and the Japanese JERS-1, the recent Canadian RADARSAT-1 SAR satellite,

Paper presented at the AGARD SPP Symposium on "Remote Sensing: A Valuable Source of Information", held in Toulouse, France from 22-25 April 1996, and published in CP-582.

		ERS-1	PHARUS
Peak Power		4.8kW	475W
Antenna	length	10m	1m
	beamwidth	0.288°	2.3°
Altitude		785km	6km
Range		875km	16km
Swath-width		102.5km	11.2km
Resolution	range	30m	4m
	azimuth	6m	1m

Table 1. Technical characteristics of the SAR systems ERS-1 and PHARUS in single polarization mode.



Figure 1. SAR image recorded with the PHARS system, the testbed of PHARUS, of the Hague, the residence of TNO Physics and Electronics Laboratory, the Netherlands.

the well known airborne SAR sytems AirSAR of JPL, and the Danish KRAS, see [Raney 92] for an extensive overview.

4. SAR PROCESSING

The signal processing related to SAR systems is usually called SAR processing and roughly consists of two resolution processing stages. The first is a common radar compression technique, in which the received modulated pulse (chirp) is compressed by convolution with a replica. This technique combines the advantages of high power and an effectively short pulse. This stage is also called the range compression stage, because it achieves high resolution in across-track direction. The next stage compresses in azimuth, or along-track direction. Here, the radar returns of the synthetic aperture are correlated with the doppler history of a point target. Between the two stages the data matrix is transposed to match the change in compression direction. Because of the side-looking geometry, the image is converted from the slant view into a plan-view image (i.e., as viewed from above). Full registration to a map coordinate system is then also possible. To reduce the amount of speckle-noise multi-looking can be applied during azimuth compression. With this technique the synthetic aperture is divided in a set of smaller apertures, called looks, that are separately compressed and averaged incoherently.

In SAR processing special attention has to be paid to antenna motion in case of airborne SAR, and to the motion and curvature of the Earth in case of spaceborne SAR. Due to cross-track motions of the aeroplane, the doppler history along the synthetic aperture is distorted. By measuring the motion of the aeroplane, this history can be reconstructed, as well as the image geometry. Motion information can mostly be obtained from a high performance navigation system in the aircraft, or otherwise a radar embedded navigation system. In airborne SAR imaging, the surface of the Earth is often assumed to be flat because of the small swaths relative to the circumference of the Earth. In spaceborne SAR this assumption is no longer valid. Also, an airborne radar operates in the atmosphere, and so has the same rotation as the Earth, contrary to satellites, which orbit around it. Due to Earth curvature and -rotation, every part of the imaged scene has a different speed towards the sensor, yielding a different doppler offset-frequency, in case of spaceborne SAR.

An important feature of a SAR processor is the accuracy of reconstructing the recorded scene spatially as well radiometrically. The required accuracy depends on the application. Moreover, in some applications it is very important that the SAR processor preserves the phase. On the other hand, there are applications where the throughput of the processor is the main concern. Different kinds of SAR processors have been built throughout the years, ranging from the early optical processors, to software architectures on standard computers, to dedicated hardware parallel architectures. Some examples of processors are the ERS-1 SAR processors of several ERS-1 Processing and Archiving Facilities (PAF). For instance, the United Kingdom PAF uses a Micro VAX II based system with three array processors. The German PAF uses a set of general purpose workstations and micro computers linked by a local area network [Guignard 90]. Another example of a processor on a standard platform is the Netherlands Generic SAR Processor (GSP). This processor, which is able to process data from airborne and spaceborne systems as PHARUS and ERS-1, runs on a SUN SparcStation with an optional external vector processor [Otten 94]. An example of a dedicated hardware processor is the Norwegian CESAR processor [Toverud 88]. This processor is based on a systolic array parallel architecture. A special kind of SAR processors are the real-time processors [Bierens 95] which process the SAR data on-board during recording. Technologies used for dedicated hardware and real-time processors are transputers, DSP pocessors and VLSI techniques.

5. SPECIAL SAR MODES AND APPLICATIONS

Beside the basic imaging modes a SAR system can be made to operate in other modes to perform special applications. The next section gives an overview of current special modes and applications of SAR.

5.1 Multi-frequency imaging

Like multi-spectral images in optical remote sensing, multi-frequency SAR images yield more information. Objects give different backscatter intensities for different radar wavelengths. To show the multi-frequency information, the different wavelength images can be viewed in composite colours. The differences can be amplified by applying techniques such as principal-components analysis. Sometimes, images from different sensors can be used like the ERS-1 operating in C band and JERS-1 operating in L band. Some sensors operate at different frequencies at the same time, like the AirSAR of JPL operating in C, L and P band. The advantage of one sensor operating at different frequencies is that spatial registration is almost automatic. A successful application of multi-frequency imaging is land-use classification.

5.2 Polarimetry

Polarimetry is a technique that exploits the polarization of microwaves. Generally two polarization directions are used, horizontal and vertical. As in multi-frequency imagery, most objects also have different backscatter properties for different polarizations. Normal SAR systems transmit and receive in equal polarization directions, like the ERS-1 system which is vertical-vertical (VV) polarized. Polarimetric SAR systems transmit alternating polarizations, and simultaneously receive both. This way all polarizations are recorded. An example of a polarimetric SAR is the Netherlands PHARUS system, which can operate in single and full polarization mode. Applications are land-use and crop classification, contrast enhancement etc.

5.3 Interferometry

In this technique [Rocca 96], the phase difference between two images recorded from different tracks with different points of view, is used to make a three-dimensional reconstruction of the sensed area. The different points of view have to be separated across-track and can be obtained by two sensors or by one sensor in repeat-pass mode. In the last mode, the difference in consecutive satellite passes provides the required across-track displacement. Single pass interferometric SAR systems use mostly one transmitter and two receive antennas to obtain the different points of view. Because the receive antennas must be displaced across-track, this technique is also called across-track interferometry. Since interferometry is based on the phase difference, the phase-preserving characteristics of the used SAR processor



Figure 2. The Phased Array Universal SAR (PHARUS) mounted under a Cessna Citation II aircraft.

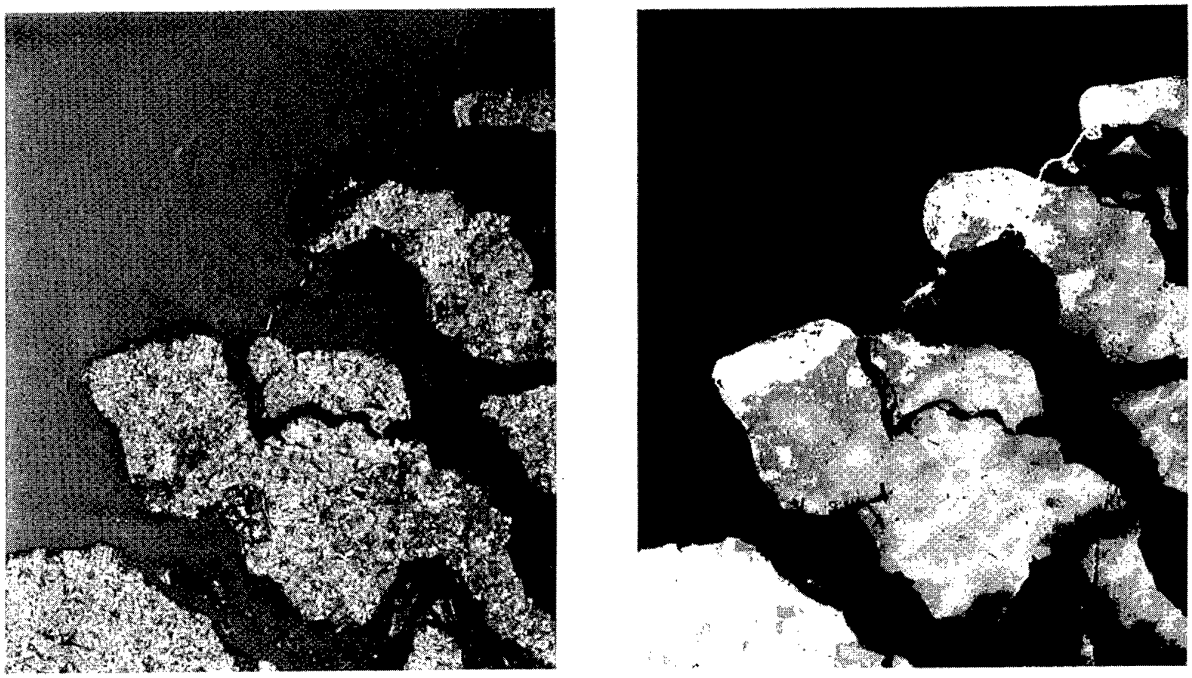


Figure 3. SAR intensity image (left) and phase-difference image (right) of Zeeland, the Netherlands, generated with the interferometric SAR processor of TNO Physics and Electronics Laboratory, the Netherlands, using repeat-pass ERS-1 imagery.

are very important, and pose additional quality requirements on SAR algorithms. Some applications of interferometry are elevation mapping and land subsidence monitoring [van der Kooij 95].

5.4 MTI

Moving Target Indication (MTI) is the proces of instantaneously discriminating moving targets from the stationary background. The target's speed in the direction of the radar can be accurately determined. In an airborne radar, radially moving targets can often be discriminated because of their different Doppler frequency. This does work well on very fast platforms, such as satellites, because the ground targets are relatively slow. In this case clutter suppression techniques involving multiple receive apertures are required. Multiple receive apertures can also be used to perform along-track interferometry, to accuractely determine target speed or direction. While MTI techniques exist in airborne systems, future satellite SAR systems are also expected to have MTI capabilities.

5.5 ScanSAR mode

ScanSAR is a technique which enables SAR to image wider swaths. A system that operates in this mode uses a steerable beam that scans the swath in range-direction. Because of the decreased observation time per unit area, the resolution in ScanSAR imagery is lower. The Canadian RADARSAT-1 satellite includes a ScanSAR mode [Raney 92].

5.6 Spotlight mode

The SAR spotlight mode makes it possible to increase azimuth resolution by keeping objects longer in its beam, thus extending the synthetic aperture [Munson 83]. This can be done by keeping the beam on the imaged scene during some time in-flight. Spotlight SAR processing has to deal with very long integration lengths, and large range migration. Therefore, spotlight algorithms are typically different from strip-mapping algoritms. The advantage of phased array antenna SAR systems such as PHARUS is that the beam can be pointed electronically, without any delay, from pulse to pulse.

6. FUTURE

The use of SAR systems in remote sensing is increasing. Moreover, there are many further developments in the understanding of the interaction between microwaves and what takes place on the ground, and there are many new system developments. SAR system development concentrates on several subjects. One is the increase of resolution, which is necessary for military applications, such as identification and verification. Another is the possibility of imaging areas on demand. Especially for spaceborne systems, this is an important item. The Canadian RADARSAT-1 satellite already proved this to be possible. Smaller airborne platforms like unmanned aircraft also expand the use of SAR. Intergration of several SAR modes and applications in one system also qualify the use of a SAR system in the future. Important developments are active phased array antennas, and interferometric and polarimetric capabilities. Microwave integrated circuits allow miniaturization of radar hardware. Advanced hardware will also lead to more advanced algorithms for processing. Such algorithms are being developed at many places throughout the world.

7. REFERENCES

Bierens, L., "Architectures for Real-Time On-Board Synthetic Aperture Radar Processing", Technical University Delft, The Neteherlands, 1995 (ISBN 90-6743-402-7).

Guignard, J.P., "ERS-1 SAR Processors", ESA er tn esa gs 0051, 1990.

Hoogeboom, P., P. Snoeij, P.J. Koomen and H. Pouwels, "The PHARUS project, results of the definition study including the SAR testbed PHARS", IEEE Trans. On Geosci. And Remote Sensing, 30(4):723-735, 1992.

Hoogeboom, P., H. Greidanus, P.J. Koomen, P. Snoeij and H. Pouwels, "PHARUS: An Autonomous Airborne SAR Capability", Symposium on Satellite and Airborne Synthetic Aperture Radar, Jakarta, Indonesia, 1994.

Kooij, M.W. van der, D. van Halsema, W. Groenewoud, B.A.C. Ambrosius, G.J. Mets, B. Overgaauw and P.N.A.M. Visser, "Satellite radar measurements for land subsidence detection", *in* Proceedings of the 5th International Symposium on Land Subsidence, The Hague, The Netherlands, 1995.

Munson, D.C., J.D. O'Brien and W.K. Jenkins, "A Tomographic Formulation of Spotlight-Mode Synthetic Aperture Radar", *in* Proceedings of the IEEE, 71(8):917-925, 1983.

Otten, M.P.G., J.S. Groot and H.C. Wouters, "Development of a Generic SAR Processor in the Netherlands", *in* Proceedings of IGARSS'94, Pasadena, pp. 903-905, 1994.

Raney, R.K., "Review of Spaceborne and Airborne SAR Systems", *in* 'Fundamentals and Special Problems of Synthetic Aperture Radar (SAR)', AGARD Lecture Series 182, Ref. 11, 1992.

Rocca, F., "Possibilities and Limits of SAR Interferometry", *in* Proceedings of AGARD Symposium 'Remote Sensing: A Valuable Source Of Information', Toulouse, France, 1996.

Toverud, M., "CESAR - A Programmable High Performance Systolic Array Processor", *in* Proceedings of IEEE Int. Conf. On Computer Design: VLSI in Computers and Processors, New York, 1988.

Vass, P. and B. Battrick, "ERS-1 System", SP-1146, ESA, 1992.

Paper 3

G.S. Brown (US)

You said there would be no speckle in the phase of a pixel element, but there would be in the amplitude. Could you please explain this?

Author's reply:

When the baseline is small, the combination of the scatterer's complex amplitudes is the same in the two views and therefore a random variate; their phase **difference**, however, is dependent on the geometry only and it is not random (not speckled).

Spotlight Synthetic Aperture Radar: System Overview and Application to the Maritime and Land Environments

Anthony Damini and George E. Haslam

Aerospace Radar and Navigation
Defence Research Establishment Ottawa/Department of National Defence
3701 Carling Avenue
Ottawa, Ontario
Canada K1A 0Z4

1. SUMMARY

Spotlight Synthetic Aperture Radar (SSAR) is a research and development project sponsored by the Canadian Department of National Defence. Its mandate is the development of the technology to retrofit the AN/APS-506 search radar, on Canada's CP-140 maritime patrol aircraft, with Synthetic Aperture Radar (SAR) imaging modes. Three SAR modes have been developed: Strip Map Mode in which the antenna is pointed side-looking to the aircraft's desired track and continuous imagery of the terrain flown by is generated; Range-Doppler Profiling (RDP) Mode in which the antenna is pointed at an ocean going-vessel and an animated image of the target is produced with time; and Spotlight Mode in which the antenna is again trained on a target of interest and a single high-resolution image of the target is produced. Spotlight has two submodes: Adaptive Spotlight for ship imaging and Non-adaptive Spotlight for land imaging. The Spotlight Synthetic Aperture Radar eXploratory Development Model (SSAR XDM) was completed in 1990 and work since then has been oriented towards improving the system's performance and resolution, enhancing the signal processing routines and developing a compact, real-time, airborne signal processor. An overview of the SSAR system is given. The original motive for the development of SSAR was to aid in maritime surveillance. The SSAR system would produce at long range, high-resolution images of ocean-going vessels from which the radar operator could discern vessel type. The land imaging capability and enhanced resolution of SSAR, however, has created interest in its potential application to other areas such as arms verification and situational awareness.

The imaging capability of SSAR was exercised in 1994/1995 in trials which took place in both the maritime and land environments. Specifically, vessels were imaged in flight trials performed over Lake Ontario in Canada and as part of naval trials off the east coast of Nova Scotia, and land trials were conducted in terrestrial regions of Canada. This paper presents trials results and discusses future enhancements planned for the SSAR system.

2. INTRODUCTION

Synthetic Aperture Radar (SAR) is an airborne, or spaceborne, radar mapping technique for generating fine resolution images of targets [1,2]. Fine resolution in the range direction is obtained by transmitting a coded (usually Linear FM) signal that has a wide frequency bandwidth, sampling the echoes in the received signal at equal slant-range spacings, and range compressing the echo samples. Fine resolution in the cross-range direction is obtained by making use of the principles of Doppler imaging where reflected pulses for each range bin are coherently integrated. Doppler imaging uses the relative motion between the radar and the target as its imaging mechanism. For Strip Map Mode, the relative motion is attributed to the forward motion of the radar platform as the swath illuminated by the antenna crosses the target of interest. For Spotlight Nonadaptive, the relative motion is again due to the forward motion of the radar platform as the antenna illuminates the patch of interest. For Spotlight Adaptive and RDP imaging, the relative motion can be attributed to two sources. The first is caused by the forward motion of the radar platform, the second is caused by the motion of the target (Inverse SAR, or ISAR, is a mode of SAR where the imaging mechanism is usually attributed to target motion). Platform motion is a measurable quantity, whereas information on target motion, while embedded in the reflected signal, is usually not directly accessable. The combined effects of these two motions is an angular rotation of the target with respect to the radar. This angular rotation sweeps

Paper presented at the AGARD SPP Symposium on "Remote Sensing: A Valuable Source of Information", held in Toulouse, France from 22-25 April 1996, and published in CP-582.

out a long synthetic aperture which allows the subsequent synthesis of a narrow antenna beam pattern. A narrow synthetic antenna beam pattern corresponds to a fine cross-range resolution.

The SSAR XDM has been retrofitted to the AN/APS-506 maritime search radar and installed on a Convair 580 which is owned and operated by the National Research Council's Institute for Aerospace Research in Ottawa, Canada. All of the conventional capabilities of the AN/APS-506 radar have been retained, and are used in concert with Spotlight SAR. The AN/APS-506 radar is primarily used for wide area maritime surveillance (i.e. surface vessel detection and periscope and snorkel detection), navigation (over land and sea) and weather avoidance. Spotlight SAR augments the maritime surveillance capabilities and further provides the capabilities to perform harbour and coastal surveillance, and mapping. In the maritime environment this allows the radar operator to localize, image and classify targets (i.e. combatant or non-combatant). Other potential applications include search and rescue, fishery patrols and detection of environmental hazards such as oil slicks.

In the maritime environment, upon detection of a surface target using the conventional AN/APS-506 search radar capabilities, the radar operator can cue the radar into a Spotlight imaging mode. If the radar operator wishes to operate in a covert fashion, he can use the Spotlight Adaptive mode which illuminates the target with the radar for the minimum time required to form a single image. This mode minimises the likelihood of alerting the target to the fact that it is the subject of a surveillance action. If operating in a covert fashion is not a concern, he can use the RDP mode, which maintains illumination on the target and forms an animated series of images. Targets for the land imaging modes are specified by the operator entering the desired co-ordinates.

The performance of the system when operating in the maritime environment, where the RDP or Spotlight Adaptive modes are used, is less affected by aircraft operating altitude than when imaging land while using the Strip Map and Spotlight Nonadaptive modes.
In the land imaging modes, the utility of the imagery is limited by shadowing. Higher operational altitudes allow Strip Map and Spotlight Nonadaptive imaging to be performed at greater ranges. In the maritime environment, sea clutter is generally not a problem because of a lack of coherency as well as the fact that sea clutter power drops with range.

Section 3 provides background information on SAR and ISAR. In Section 4, high-level descriptions of the SSAR XDM and the post-processing facilities are given. The SAR algorithms upon which the SSAR XDM's digital signal processing software is based are described in Section 5. During 1994/1995 the SSAR XDM was involved in several sets of flight trials. Imagery from these trials are presented in Section 6. Finally, Section 7 summarises the results achieved to date.

3. SAR/ISAR CONCEPTS

Figure 1 illustrates a typical Strip Map Mode imaging geometry in which the antenna is pointed sidelooking to the desired aircraft track. The vectors R_o and R_i lie in the slant-range plane, which is the plane onto which the image is projected. R_o is the initial range vector, at the beginning of the imaging session, from the antenna phase centre to the centre of the swath being imaged. R_i is the actual range vector, during data collection, from the antenna phase centre on the actual flight track to the centre of the swath being imaged. i is the index into time or space. If the actual flight track corresponded to the desired flight track then R_i would be equal to, and lie in the same slant range plane as R_o for all i. Since the actual flight track does deviate from the desired flight track, the radar data must be motion compensated so as to appear to have been collected with the aircraft following the desired flight track.

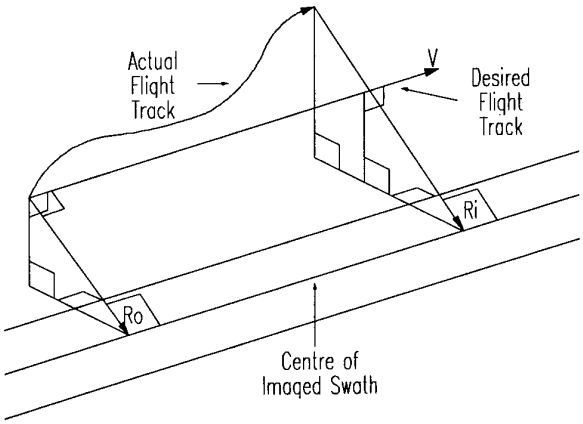


Figure 1: Strip Map Imaging Geometry

In Spotlight and RDP Modes, the SSAR XDM operates with the antenna fixed on a target of interest. Figure 2 illustrates a typical Spotlight imaging geometry. R_o, R_i and i are as defined above. η is the

squint angle about which the SAR data is initially motion compensated. If the actual flight track corresponded to the desired flight track then R_i would lie in the same slant range plane as R_o for all i. This plane contains both the line of the desired flight track and the point corresponding to the centre of the desired patch. Since the actual flight track does deviate from the desired flight track, the vectors R_o and R_i lie in different slant-range planes, and the radar data must be motion compensated so as to appear to have been collected with the aircraft following the desired flight track. The data collection points on the actual flight track are referenced to their corresponding motion compensation reference points on the desired flight track.

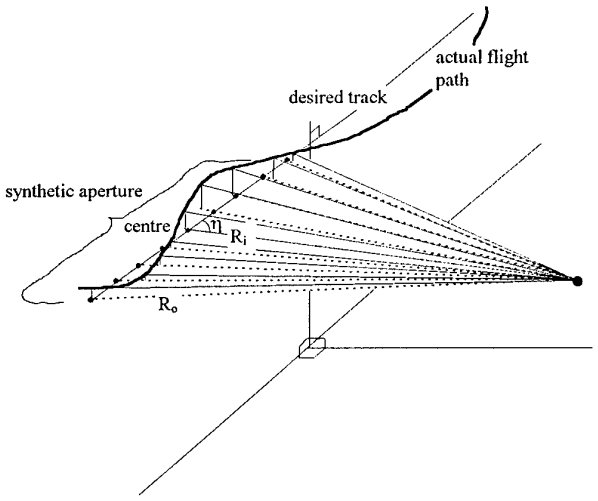


Figure 2: RDP/Spotlight Imaging Geometry

Conventional Strip Map SAR operates with the antenna pointed side-looking to the aircraft's desired track. As a point target passes through the antenna's real beam, its radial velocity and Doppler frequency decrease. Therefore, the returned signal representing such a target takes the form of a chirp signal. The signal processor compresses this chirp signal by producing a cross-range matched filter to integrate the target's response. Proper signal compression requires that the data be motion compensated for any deviations of the actual track of the radar platform from the desired track. The azimuth resolution of a conventional SAR is inversely proportional to the synthetic aperture size, or the length of the flight track over which the echo energy is coherently integrated to produce the response corresponding to a single point target. The finest possible cross-range resolution is proportional to the product of the wavelength and the range to the target, and is inversely proportional to the synthetic aperture length. Theoretically, the finest cross-range resolution obtainable with a SAR is independent of range and wavelength, and is equal to one-half the physical length of the real antenna.

Spotlight Nonadaptive imaging requires that the target remain in the antenna beam for the whole synthetic aperture. Therefore, the bandwidth of the collected signal can exceed (and usually does) the instantaneous bandwidth illuminated by the real antenna beam. The spectral history of the returned signal varies as a function of range and cross-range within the real beam. The signal processor compresses this signal by generating spatially variant filters which are matched to the position of individual scatterers within the real beam. The azimuth resolution of a Spotlight SAR in a land imaging mode is inversely proportional to the synthetic aperture size.

As with Spotlight Nonadaptive, Inverse Synthetic Aperture Radar (ISAR) tracks all of the target scatterers simultaneously for the duration of the synthetic aperture. The resolution in cross-range is obtained by coherently processing the echo energy returned from the target. The Doppler gradient in these echoes is once again caused by the relative angular motion between the radar and the target, but, in pure ISAR, the relative angular motion is attributed solely to the target's angular motion. The azimuth resolution of an ISAR image is inversely proportional to the angle through which the target rotates during the synthetic aperture, and is thus a function of the angular rate and aperture time. The essential difference between SAR and ISAR is thus summarised as follows: in SAR the radar induces rotations, in ISAR the target rotates. In the cases of turntable ISAR and planetary imaging, the Doppler gradient is known and the cross-range resolution of the image can be computed. When the Doppler gradient is unknown, the cross-range resolution cannot be computed.

Mixtures of SAR and ISAR are possible if both the radar and the target are moving. This is generally the case when ocean-going vessels are being imaged. Both the Spotlight Adaptive and the RDP modes embody hybrid SAR/ISAR imaging mechanisms. A ship in port is usually stationary (as ports are generally protected). Therefore, the imaging mechanism is primarily due to the motion of the radar platform. A ship at sea

experiences both translational and rotational motion. In this case a hybrid SAR/ISAR imaging mechanism contributes to the formation of the radar image. Spotlight Adaptive and RDP aperture times can range from a fraction of a second to several seconds. The larger of these values may represent a significant fraction of, or be greater than, the period of a ship's angular motion. The cross-range resolution and orientation of the SSAR image can undergo significant changes in this time which will cause severe image degradation [3]. Small aperture times are utilised with high sea states and high Pulse Repetition Frequencies (PRFs), and large aperture times are utilised with low sea states and low PRFs.

4. SPOTLIGHT SAR XDM AND GROUND PROCESSING FACILITY

4.1 Spotlight SAR XDM

The Spotlight SAR XDM system configuration is illustrated in Figure 3. A functional description of each of the subsystems follows.

The Control Computer Subsystem (CCS) is a Concurrent 3210 computer with an attached operator's console. The CCS acts as the command interface to the AN/APS-506 radar subsystem and the subsystems for SAR operation. The CCS initialises the subsystems, sequences the SAR operation, and responds to operator requests and error conditions.

The CSAR signal processor is VME bus based, consisting of a system controller card, a high-speed interface to the Motion Compensation Subsystem, six i860-based signal processing cards and a TMS 34020-based graphics display processor card. The signal processing algorithms are configured to execute on the six i860-based cards in a Multiple Instruction Multiple Data (MIMD) mode. Each processor executes roughly the same signal processing operations, but on different data, and communicates with each of the other processors for the purposes of sharing data and control information. The software has been designed to be scaleable so that the use of additional processors will improve the processing throughput in roughly a linear fashion. Conversely, the availability of a limited number of processing nodes can be accommodated in a real-time operational scenario by restricting the radar Pulse Repetition Frequency (PRF), and thus the required computational throughput of the CSAR.

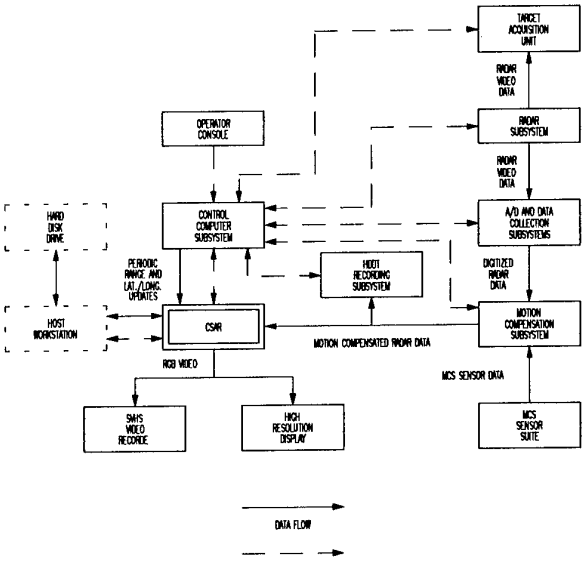


Figure 3: Spotlight SAR XDM System Configuration

The radar subsystem is a modified AN/APS-506 search radar. The radar pulse length is 0.5 microseconds and the peak transmitter power is 500 kWatts. The carrier frequency is 9.75 GHz and the system polarization is Horizontal/Horizontal (HH). The antenna has a physical width of 42", and azimuth and elevation beamwidths of 2.4° and 4.0°, respectively. For SAR operations, an Arbitrary Waveform Generator (AWG) is employed to generate highly precise digital Linear Frequency Modulated (LFM) waveforms, of various bandwidths, to modulate the Intermediate Frequency (IF) signal. Between data collection sessions, the digital waveform can be reprogrammed, and the sampling rate of the digitizer changed. This allows the range resolution to be traded off with the swath width since 1024 range cells are always digitized across the swath.

The Data Collection Subsystem coherently demodulates and digitizes to 8-bits the I (In-phase) and the Q (Quadrature-phase) channels of the analog SAR data. A range record is formed by sampling 1024 of these complex pairs during the inter-pulse period. Each record is temporarily stored and then forwarded to the Motion Compensation Subsystem. The A/D Conversion Subsystem supports five complex sampling rates. A synopsis of these sampling rates, with respect to range processing, is given in Table 1. For ship imaging, the sampling rates of interest are the super-high resolution and the high-resolution sampling rates. The lower sampling rates are also useful for maritime surveillance as they provide a synoptic view of the target(s) and its(their) surroundings. For land imaging, all of the sampling rates are useful, depending on the application. The processed range resolutions listed in

Table 1 assume a degradation in the main-lobe response of approximately 1.4 due to the Hamming window weighting imposed on the range filter. This weighting yields a maximum sidelobe of -43 dB. The upper limits on the bandwidths are determined by the digitizer sampling rates and the inclusion of guard bands. The guard bands take up a larger percentage of the Nyquist bandwidth at lower sampling rates, so the fraction of the bandwidth that is used is much smaller.

Table 1: Spotlight SAR XDM Swath Parameters

Mode	Sampling Rate (MHz)	ρ_{Pixel} (m)	Swath Width (m)	Bandwidth (MHz)	$\rho_{Range_processed}$ (m)
Super-High	>125	<1	-	>105	<1
High	>125	<1	-	>105	<1
Medium	125	1.2	1228.8	≤105	2.00
Low	62.5	2.4	2457.6	≤42.5	4.94
Super-Low	31.25	4.8	4915.2	≤11.25	18.67

Coherent integration over a long synthetic aperture requires that the radar data be motion compensated for undesirable platform motions which introduce pulse-to-pulse phase discontinuities in the azimuth direction. These motions usually consist of a non-uniform air speed and deviation of the actual flight track from the desired track, coupled with platform roll and pitch. The Motion Compensation Subsystem (MCS) corrects for these undesirable motion artifacts by [4]:

i) Controlling the PRF as a function of the platform velocity along the desired flight track, and the desired squint. This results in the radar pulses being transmitted at equi-spaced intervals along the desired track. The PRF is thus formulated as

$$PRF=S_s \cdot V \cdot \sin \eta, \qquad (1)$$

where S_s is the spatial sampling constant of 5.3333 Hz/(m/s). In Strip Map Mode η equals 90°or 270°;

ii) Controlling the triggering of the digitizers so as to maintain the range (time-delay) to the start of the digitization window centred over the area being imaged regardless of aircraft deviation from desired track; and

iii) Modulating the phase of the sampled radar data, on a pulse-to-pulse basis, so that the echo data appears to have been collected with the phase centre of the radar antenna following the desired track and pointing at the centre of the desired patch. For RDP and Spotlight Adaptive, the SAR processor also corrects for the deviation from the expected phase history added by the target motion.

For the purposes of measuring the composite platform motions, highly sensitive motion sensors (gyros to measure angles and accelerometers to measure displacements) are employed. The optimal position on the aircraft for mounting the motion sensor equipment is at the phase centre of the radar antenna. The sensor suite employed by the Motion Compensation Subsystem consists of an Inertial Measurement Unit (IMU), which is placed as close to the antenna phase centre as possible, and an Inertial Navigation System (INS), which is located near the centre of gravity of the platform. The IMU provides high-rate (50 Hz) data in the form of angle and velocity increments, and the INS provides low-rate (16 Hz) data in the form of aircraft position, velocity, roll, pitch and heading. Information derived from the INS data is used to provide long-term stability to information derived from the IMU measurements by comparing the derived results via a Kalman filter. A Global Positioning System (GPS) receiver has also been incorporated into the MCS to further enhance the long-term stability to the IMU derived data, as well as augment the navigational and targeting accuracy of the system. GPS allows position and velocity to be ascertained to accuracies of 100 metres and 0.1 m/s (rms), respectively.

The output of the Kalman filter, which is used to control the radar system and correct the collected data in the manner described above, effectively removes the short term, high-frequency, and some of the long-term, low-frequency motion artifacts from the radar data. There are, however, residual long-term, low-frequency artifacts which introduce errors into the deterministically calculated quadratic phase rates, and which must be compensated prior to the execution of the cross-range image formation steps. This compensation, along with the compensation for any

target motion, is carried out by Autofocus routines. The MCS forwards the motion corrected raw data to both the CSAR and the High Density Digital Tape (HDDT) Recording Subsystem.

The Target Acquisition Unit (TAU) is enabled once the radar operator has designated the target. The TAU carries out data collection by directing the radar to sweep the antenna so as to illuminate the target area and then four range cells are digitized (each spanning approximately one hundred metres in range) to ten bits for 256 pulses. The Automatic Gain Control (AGC) levels, signal amplitude, antenna azimuths and timestamps are then stored with each range sample by the CCS. The CCS then correlates this data with a stored replica of the antenna power spectrum. A very precise latitude and longitude are determined for the target. The AGC levels are used to determine an optimum receiver gain setting which prevents the digitizers from being saturated.

The raw, motion compensated radar data is recorded on High Density Digital Tape (HDDT), as it is collected, so that scenes of interest may be reprocessed to higher resolutions using a ground-based processing facility. The HDDT recording system is based on three T-826 linear tape recording devices. Each device can sustain a data transfer rate of 800 kilobytes/second. The recording medium consists of a one-half inch tape cartridge with a storage capacity of 2.6 Gigabytes. The data are collected in scenes consisting of 32772 records.

4.2 Target Selection

The location of the area to be Strip Mapped, or patch to be Spotlighted, can be selected in one of two ways. First of all, the area to be targeted can be designated by the operator entering the target co-ordinates in latitude and longitude. Secondly, an area can be designated by selecting a target from the coarse resolution Plan Position Indication (PPI) display, or the finer resolution B-scan display, while the radar is in one of its search modes. For both methods, the CCS computes the antenna depression angle and slant-range offset (since it already has its own altitude and position information) and uses the MCS sensor data to control: the radar antenna so that the centre of the desired patch (in azimuth) is coincident with the boresight of the antenna beam pattern; and the radar timing circuitry and analog to digital converter so that the centre of the desired patch (in range) is coincident with the centre of the radar's digitization window. For targets of a known fixed location, the first approach is much more reliable due to the integration of the highly accurate Global Positioning System (GPS) receiver into the MCS sensor suite. The second approach is almost exclusively used for ship targets at sea.

4.3 SAR Ground Processing Facility (SAR GPF)

Practical limitations on the size, cost and complexity of the airborne signal processor do not allow data which has been collected at long range to be processed to the finest possible cross-range resolution in the land imaging modes. The processing of this data is done in a non-real-time mode. The SAR Ground Processing Facility (SAR GPF), which is a derivative of the CSAR design and based on two quad-i860 cards, is used to reprocess the radar data to the finest azimuth resolution allowed by the bandwidth of the radar antenna. The SAR GPF is a portable, networkable system. The user interface is implemented through a Sun Workstation, and communication is over an Ethernet connection. The digital imagery generated by the SAR GPF can be stored on either digital tape or the Sun's hard disk drive. An X-Windows based graphics package, running on the host, is used to review the digital imagery. This package allows the SAR imagery to be repackaged in industry standard file formats, such as Postscript, tiff or pcx, for dissemination.

5. SAR ALGORITHM OVERVIEWS

SAR signal processing for land imaging can be theoretically formulated as a two-dimensional task. This is primarily due to the range migration of the targets being deterministic during the formation of the synthetic aperture. The Strip Map Mode algorithm is based on the Range-Doppler algorithm [5]. The algorithm decouples the processing into the range and cross-range directions by performing matched filtering in each direction. A range cell migration correction step is used to correct for the variation in slant range between the radar antenna and the target during the synthetic aperture. For moderate squint angles (deviations from sidelooking), the two one-dimensional filters of the Range-Doppler algorithm, combined with the range cell migration correction process, closely approximate the result which would be obtained by the use of an exact two-dimensional matched filter. Fast convolution in the frequency domain is used for the matched filtering because the long Doppler histories preclude the use of a time-domain implementation. An Autofocus step, based on the correlation of subapertures, is used to correct for undesirable platform

motions not compensated for by the MCS, as well as low-frequency errors inherent to the MCS.

The Spotlight Nonadaptive algorithm processes the radar data so as to produce a single high-resolution image of the target. The algorithm images in the cross-range direction using one of two methods. It can use a coherent subaperture processing technique which first forms NRAMP coarse resolution images, referred to as subapertures, each of length LFFT records and overlapping by OVERLAP records, and then coherently combines these subapertures to form a single fine resolution image. Alternatively, it can use a single subaperture to form a single fine resolution image of length LFFT records. The Spotlight Nonadaptive coherent subaperture processing technique is combined with a range cell migration correction algorithm to implement the cross-range compression.

Hybrid SAR/ISAR imaging, however, which involves the imaging of a target which has its own components of motion, is traditionally decoupled into two one-dimensional tasks, since the motion of the target in each direction, during the formation of the synthetic aperture, is unknown. The Range-Doppler Profiling (RDP) algorithm [6] is a SAR processing algorithm, based on spectral analysis techniques, which decouples the processing into the range and azimuth dimensions. The algorithm first performs range compression and tracking (one target is tracked) in the range direction, and then performs spectral analysis in the cross-range direction, in conjunction with an Autofocus procedure, to form the final image. The RDP algorithm achieves a processing throughput commensurate with the rate at which the data is collected. The RDP algorithm processes the incoming radar data so as to produce a continuous series of images of the target with time. Each RDP image is formed from LFFT range records, which collectively are referred to as a Major Frame. LFFT is computed from the operator entered synthetic aperture time, T_a, in seconds, as follows:

$$\text{LFFT} = \text{nearest power of } 2 \leq T_a \times \text{PRF} . \quad (2)$$

The same range records may contribute to both the end of one Major Frame and the beginning of the next Major Frame. In this case an operator entered overlap ratio of Overlap exists between the two Major Frames. The overlap in range records between Major Frames is then

$$\text{OVERLAP} = \text{Overlap} \times \text{LFFT} . \quad (3)$$

The Spotlight Adaptive algorithm processes the incoming radar data so as to produce a single high resolution image of the target. The cross-range processing uses one of two methods: a coherent subaperture technique similar to that of the Spotlight Nonadaptive algorithm; or a single subaperture to form a single fine resolution image of length LFFT records. The Spotlight Adaptive algorithm performs range compression and tracking in the range direction (multiple targets are tracked and corrected as opposed to using a deterministic range cell migration correction process such as is employed in Spotlight Nonadaptive), tracks and corrects the same targets in the cross-range direction (Doppler domain), and then coherently processes the data in the cross-range direction to form the final image. Neither Spotlight Adaptive nor Nonadaptive are real-time modes and thus neither require a sustained high throughput rate.

All the image frames produced are accompanied with the following annotation: Time, Scene Number, and Latitude and Longitude. The Time, in seconds, is measured from the start of the current mission. The Scene Number indexes to the scene containing the raw data from which the current image frame was produced. Latitude and Longitude are the co-ordinates of the beam centre and are expressed to a precision of seconds.

6. FLIGHT TRIALS RESULTS

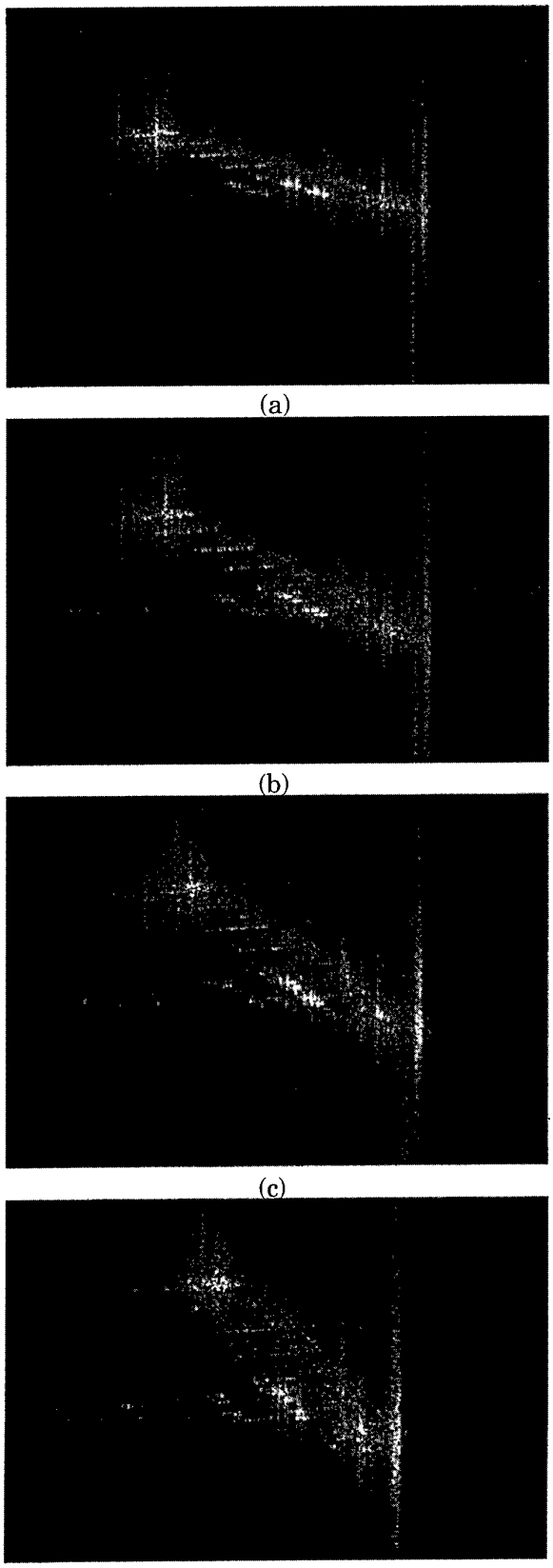
The Spotlight SAR XDM was flight tested in several sets of flight trials during 1994 and 1995.

The MARCOT trials were a set of tactical naval exercises held off the east coast of Nova Scotia. The SSAR XDM participated on a non-interfering basis, for the purposes of imaging vessels, from 22 June 1995 to 27 June 1995. Figure 4 represents a Spotlight Adaptive image of a vessel imaged during MARCOT. The near-range corresponds to the left-hand side of the image. The super-structure indicates that this view represents a vertical profile of the vessel.



Figure 4: Spotlight Adaptive Image

The Lake Ontario trials consisted of a series of flights between 1 June 1994 and 29 June 1994. The targets of opportunity were merchant vessels on Lake Ontario. Figure 5 represents a series of RDP images of a vessel imaged during these trials. The images span approximately 20 seconds of data. In Figure 5, the near-range corresponds to the right-hand side of the image. The returns in range which are located roughly at the vessel's midstructure are due to two cargo hatches being open and part of the transmitted signal being reflected within the hull of the ship before a portion of it is reflected back to the radar. The intensity of these repeated returns is seen to diminish with range. Figure 6 contains a two-dimensional plot of the first image in Figure 5. This plot appears as a two-dimensional plot since it has been rotated to present a profile view of the target. Evident in the plot is a row of cargo hatches.



(a) (b) (c) (d)

Figure 5: SSAR RDP Imagery

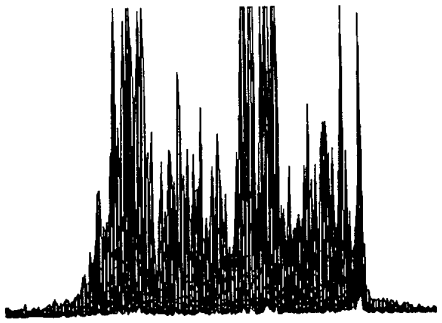


Figure 6: Two-Dimensional Plot of First Image in Figure 5

Figure 7 is a Spotlight Nonadaptive image of Ottawa International Airport. The near-range is along the left-hand side of the image. The aircraft hangars appear as L shapes which are characteristic of buildings imaged under the conditions of shallow depression angles. The roof of the terminal is strongly reflective. One aircraft is visible to the right of the main terminal area.

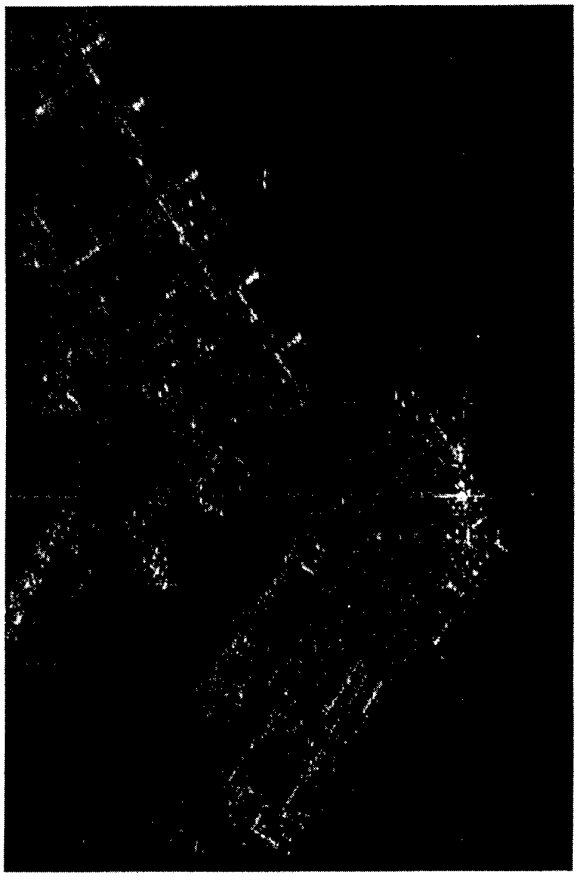


Figure 7: Spotlight Nonadaptive Image of Ottawa International Airport.

Figure 8a is a Strip Map image of the National Aviation Museum in Ottawa, Canada. The near-range is along the right-hand side of the image. A row of aircraft are on display along one side of the triangular building. Two rows of light aircraft are also visible along the edge of the runway. A zoomed view of the building and aircraft are given in Figure 8b.

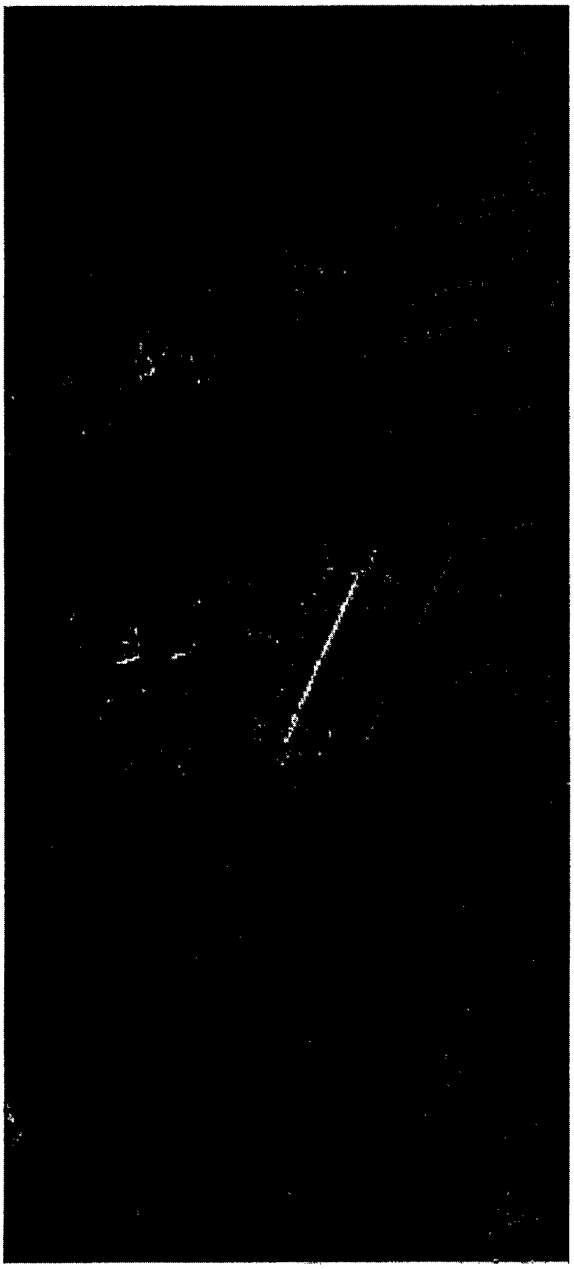


Figure 8a: Strip Map Image of the National Aviation Museum, Ottawa, Canada

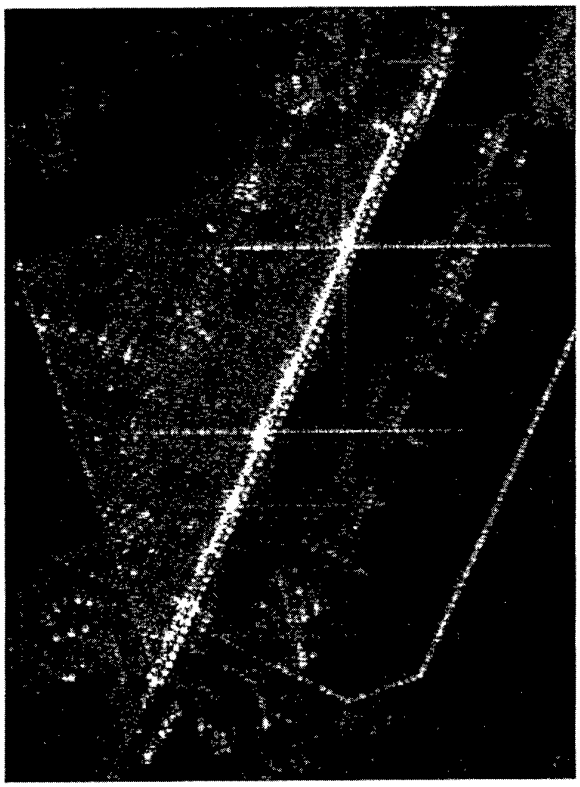


Figure 8b: Zoom of Building and Aircraft in Figure 8a

7. SUMMARY

The Spotlight SAR eXploratory Development Model (SSAR XDM) has been described. The concept of operation with respect to its operation in both the land and the maritime environments has been discussed. The principles behind the generation of SAR and SAR/ISAR images have also been reviewed. The SSAR XDM has two ship imaging modes (Range-Doppler Profiling (RDP) and Spotlight Adaptive), and two land imaging modes (Strip Map and Spotlight Nonadaptive). A high-level description has been given of the SSAR XDM's signal processing algorithms for these modes. Spotlight SAR's long-range imaging capability makes it useful for the surveillance of hostile territory or targets without endangerment of the aircraft. It also acts as a force multiplier by allowing surveillance of large areas while minimising the amount of transit time required by the aircraft. Onboard processing allows greater situational awareness to be realized in near real-time by the operational community.

Imagery has been successfully produced in the airborne environment for each SAR mode. Both land imagery and imagery of vessels were presented and evaluated. The Spotlight SAR project is currently in its next phase with the development of the Spotlight SAR Advanced Development Model (SSAR ADM). The SSAR ADM builds upon the proven technology of the SSAR XDM by increasing the swath width coverage and enhancing both the signal processing capability, so as to improve the performance of the algorithms, and the airborne processor's throughput (up to 2 Gigaflops), so as to allow the data to be processed to higher resolutions than the XDM airborne processor allowed. The SAR subsystems developed for the ADM will be ruggedized and of a VME bus form factor. The majority of the VME modules are off-the-shelf. Custom hardware development is required for the A/D converters, SAR waveform generator and SAR waveform timing circuitry. The SSAR ADM will be flight tested on a CP-140 aircraft. Future work also includes the inclusion of a data link for transmission of digital imagery to ground stations.

8. REFERENCES

1. D.R. Wehner, "High Resolution Radar", Artech House, Massachusetts, 1987.

2. J.C. Curlander and R.N. McDonough, "Synthetic Aperture Radar Systems and Signal Processing", John Wiley and Sons, New York, New York, 1991.

3. A. Damini and G.E. Haslam, "Target Angular Motion and Its Effect on Simulated SSAR Ship Images", Canadian Conference on Electrical and Computer Engineering, 25-27 September 1991, Quebec City, Canada.

4. D. DiFilippo, "Description of Data Processing in the Phase III Motion Compensation Subsystem for the Spotlight SAR Project, Volume II: Targeting Algorithms", Defence Research Establishment Ottawa, Report No. 1111, November 1991.

5. J.R. Bennett, I.G. Cumming and R.A. Deane, "The Digital Processing of SEASAT Synthetic Aperture Radar Data", IEEE International Radar Conference Proceedings, pp. 168-175, 1980.

6. A. Damini and G.E. Haslam, "SAR/ISAR Ship Imaging: Theoretical Analysis and Practical Results", EUSAR '96, 26-28 March 1996, Konigswinter, Germany.

MM-WAVE AIRBORNE CLOUD RADARS

John Galloway, Andy Pazmany, Robert McIntosh
Microwave Remote Sensing Lab
Knowles Engineering Building, University of Massachusetts at Amherst,
Amherst, MA 01003
Robert Kelly and Gabor Vali
University of Wyoming, Department of Atmospheric Science
Laramie, WY 82071

Summary

The demand for fine scale radar measurements of clouds from an airborne platform motivated the development of a 95 GHz imaging polarimeter used on a King Air research aircraft. The basic system components and operation of the polarimeter are discussed and the data processing techniques used to obtain conventional measurements of clouds and precipitation are presented. Examples of data obtained from a melting layer, mixed phase cloud, and marine stratus cloud during three different field experiments are provided with a description of flight and weather conditions prevailing during each observation.

TABLE I
WEIGHTS AND SIZES FOR ELEMENTS OF AIRBORNE 95 GHz POLARIMETER.

Item	Wt. (lbs.)	size (W x L x H)
EIA controller	28	19" x 12.5" x 5.25"
Radar RF	105	16" x 32" x 14.5"
antenna	20	14" x 15" (conical)
Power supply	39	19" x 14" x 5.25"
IF	15	19" x 14" x 5.25"
Hard Drive	35	19" x 23" x 5.25"
VXI cage	78	19" x 24" x 12.25"
oscilloscope	15	19" x 12" x 7"
System	335	

List of Symbols

Z_{eHH}– copolar reflectivity (H pol.) (mm^6m^{-3})
ρ_{HV}– copolar correlation coefficient
LDR – linear depolarization ratio
Z_{DR}– differential reflectivity
PRF – pulse repetition frequency
SSB – single side band
VXI – VME bus with extensions for instrumentation
HP – Hewlett-Packard Company
EIA – Extended Interaction Amplifier

1 Introduction

Ground-based weather radars, operating at S and C bands, have been in use for many decades and have principally been used to characterize precipitation [1]. These operating frequencies place fundamental limits on the spatial resolution possible with a given antenna system and on the sensitivity of the system to small hydrometeors [3]. X-Band radars have recently taken advantage of the benefits of airborne platforms for use in measuring weather systems. Examples include the National Center for Atmospheric Research (NCAR) ELDORA system [4], and the National Oceanographic and Atmospheric Administration (NOAA) P-3 system [1].

The advantages of a 3 mm wavelength radar system in making measurements of clouds include the physical size of the antenna required for a given angular resolution [3], the increase in sensitivity to small cloud particles as compared to lower frequency systems and the physical size of the RF components and overall system. Comparisons between the sensitivity of Doppler weather radar systems operating in different frequency bands may be found in [6], which emphasizes the use of 94 GHz as the operating frequency for a ground based system. The increase in sensitivity to small hydrometeors given small wavelength may be seen simply by observing the λ^{-4} dependence of backscatter efficiency for dielectric spheres in the Rayleigh region [12]. The impact of the small size of W-Band waveguide and associated components may be seen in the table of system component weights and sizes (see table I).

The University of Massachusetts (UMass) and University of Wyoming (UWyo) have collaborated during three field experiments making use of the combination of the UWyo King Air airplane and UMass 95 GHz polarimeter. The first field effort was undertaken in late fall of 1992 in the region around Laramie, Wyoming and established the feasibility of making fine scale measurements of clouds using the combination of polarimeter and small aircraft [9]. The second field effort took place during the Win-

Paper presented at the AGARD SPP Symposium on "Remote Sensing: A Valuable Source of Information", held in Toulouse, France from 22-25 April 1996, and published in CP-582.

ter Icing and Storms Program of 1994 (WISP94) during February and March in the Colorado Front Range region of the Rockies. The research objectives of WISP94 included winter precipitation production in upslope and wave clouds and aircraft icing conditions [10]. The third field effort covered the Small Cumulus Microphysics Study (SCMS) in central Florida during July and August of 1995 and the Coastal Stratus (CS) experiment during August and September in Oregon. The focus of the SCMS effort was precipitation initiation and convective dynamics in small cumuli and that of CS was characterization of marine stratus clouds and precipitation.

This paper presents a summary of the 95 GHz polarimeter's specifications and operation, its measurement capabilities, the in situ measurements available from the King Air probes and examples of data from the three field experiment periods undertaken thus far. Specification of the polarimeter's operation is broken down into the performance and operation of the RF section and the data acquisition. The fundamental measurement relation for polarimeters in terms of the target scattering matrix is reviewed. The relations between conventional polarimetric weather radar measurements and the scattering matrix are summarized and the influence of the hydrometeor properties on the measurements is discussed. Data from each of the three field experiment periods are then presented in the form of images or plots of the radar data and a description of the King Air measurements of weather and flight conditions prevailing during the radar observations.

2 Radar System Description

The UMass 95 GHz airborne cloud profiling polarimeter is a pulsed, fully coherent, dual receiver radar system with programmably switched polarization from pulse to pulse. An antenna mounted to the wall of the King Air is shielded by a faring containing a motor-driven reflector plate and two dielectric windows, providing the capability to direct the radar beam up or to the side. A block diagram of the RF portion of the radar system is shown in figure 6. The outputs of the two receivers (horizontal (H) and vertical (V) polarizations) are sampled by digitizers controlled by 4 DSPs running in parallel and the signal processing to obtain the initial data products is handled by the DSP's. A system control computer configures and runs the DSP's and arbitrary function generator (AFG) used to key the transmitter. The entire data acquisition system is contained in a VXI cage which provides the power supply and necessary cooling. The system uses standard line power (120 VAC), with a maximum power requirement of 2 kW.

TABLE II
95 GHz Polarimeter Specifications

Transmit Frequency:	94.92 GHz
Peak Power:	1.2 kW
Pulse Duration:	50 ns - 2 μs
Pulse Repetition Frequency:	5-40 kHz
Antenna Diameter:	30.5 cm
Antenna Beam Width:	0.7 degrees
Receiver Noise Figure:	11 dB (SSB)
Receiver Bandwidth:	5 MHz

2.1 RF Subsystem Characteristics

The frequency reference for the system (93.72 GHz) is generated by a phase-locked oscillator (PLO) referenced to a 120 MHz crystal oscillator. Dual upconversion, using a PLO at 1.08 GHz, the 120 MHz reference, and the 93.72 GHz reference provides a pulsed 94.92 GHz input to the Extended Interaction Klystron Amplifier (EIA). The EIA generates peak power output of 1.2 kW with a pulse length ranging from 50 ns to 2 μs. A network of switching circulators connected to an orthomode transducer (OMT) allows for transmit/receive and polarization switching at the front end just before the antenna. Each of the two dual downconversion receivers consists of a W-Band mixer using the 93.72 GHz reference, L-Band low noise amplifier, L-Band mixer using the 1.08 GHz reference, followed by I/Q and log detectors using the 120 MHz reference. The I/Q detectors use a hard-limited copy of the return signal in order to avoid loss of phase resolution in the I/Q plane due to low signal levels. The expected dynamic range of return power is about 70 dB. This magnitude information is retained in the output of the log detector. The operating specifications of the RF system are summarized in table II.

2.2 Data Acquisition Characteristics

Processing of the output of the RF section takes place in the VXI-based data acquisition system. An HP control computer tasks an arbitrary function generator (AFG) and DSP subsystem to obtain a given radar measurement. The AFG sends a set of pulses to the radar RF section which are used to key the transmitter. The length of the pulses is used to indicate the desired polarization state from pulse to pulse. The AFG also generates the clock used to trigger the A/D's used in the DSP subsystem to sample the output of the IF. The DSP subsystem reads the FIFO memories of the A/D's and com-

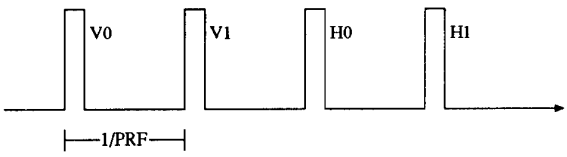


Fig. 1. Transmit pulse polarization sequence used in pulse-pair and sample average algorithms to find Doppler velocity and spectrum width along with copolar and crosspolar power.

putes running averages of covariance estimates of the backscattered return from each set of 4 transmit pulses (see transmit waveform diagram in figure 1). These estimates are periodically stored on the hard drive connected to the control computer. Algorithms currently implemented with this system include the pulse-pair estimator of Doppler velocity and spectral width, sample averages in power for copolar and crosspolar power, and full Doppler spectrum estimates using windowed periodograms (FFT used to produce power spectrum estimates). A more detailed discussion of these algorithms in the context of atmospheric radar signal processing may be found in [3].

3 Polarimetric Measurements of Clouds

3.1 Fundamental Relation

The fundamental equation describing the relationship between a target and the response measured by a polarimeter is [11]:

$$\begin{bmatrix} E_h^s \\ E_v^s \end{bmatrix} = \frac{e^{-jkr}}{r} \begin{bmatrix} S_{hh} & S_{hv} \\ S_{vh} & S_{vv} \end{bmatrix} \begin{bmatrix} E_h^i \\ E_v^i \end{bmatrix}, \quad (1)$$

where E^s is the scattered electric field at the receiving antenna and E^i is the field incident on the target. Each of the scattering matrix parameters, S_{ij}, is related to a corresponding volume backscatter cross section by:

$$\sigma_{ij} = 4\pi |S_{ij}|^2, \quad (2)$$

where ij takes on the four polarization combinations (received and transmitted): hh, hv, vh, and vv.

3.2 Derived Parameters

Covariances of the elements of the scattering matrix defined in (3.1) are related to the polarimetric measurements most often used to characterize atmospheric targets by [8]:

$$Z_{eHH} = 4\frac{\lambda^4}{|K|^2\pi^4} \langle |S_{hh}|^2 \rangle \quad (3)$$

$$Z_{DR} = 10\log_{10} \frac{\langle |S_{hh}|^2 \rangle}{\langle |S_{vv}|^2 \rangle} \quad (4)$$

$$LDR = 10\log_{10} \frac{\langle |S_{hv}|^2 \rangle}{\langle |S_{vv}|^2 \rangle} \quad (5)$$

$$\rho_{HV} = \frac{\langle S_{hh} S_{vv}^* \rangle}{\sqrt{\langle |S_{hh}|^2 \rangle \langle |S_{vv}|^2 \rangle}} \quad (6)$$

The indicated expectations, $< \bullet >$, are taken over the ensemble of particles in a radar resolution volume, and $*$ denotes a complex conjugate.

The units of reflectivity, Z_{eHH}, are customarily $mm^6 m^{-3}$. Values of Z_{eHH} are determined by a combination of hydrometeor density and size. One of the principal applications of Z_{eHH} measurements has been the attempt to retrieve cloud liquid water content (LWC) or rain rate, R, from observations of clouds or precipitation. Z_{DR} and LDR are unitless and serve as indicators of hydrometeor orientation and geometry. Particles with a greater extent in the horizontal as viewed in the polarization basis of the polarimeter will have Z_{DR} values greater than 0 dB (oblate particles viewed at side incidence for instance), while prolate particles viewed at side incidence and falling with their major axis vertically aligned will produce Z_{DR} values lower than 0 dB. The best known application of Z_{DR} in conjunction with Z_{eHH} measurements has been in the discrimination of liquid and ice hydrometeors [3]. LDR indicates the degree of asymmetry or randomness of the viewed hydrometeors. The clearest indication of LDR appears in observations of melting layers present in precipitating clouds, but LDR also finds application in research into determining the crystal habit of hydrometeors observed in ice clouds [7]. The copolar correlation coefficient provides information about the degree of orientation present in a volume of hydrometeors. Research into the meaning and use of ρ_{HV} measurements may be found in [14] and [5].

4 Available Aircraft Data

The King Air data acquisition system provides a variety of measurements of the cloud and weather conditions present at the wingtips as the aircraft travels through a given cloud region (see [13]). Cloud hydrometeor habit and phase may be determined from the 2D laser spectrometers manufactured by Particle Measurement Systems (PMS). Other meteorological parameters of interest in interpreting the radar data include: temperature, humidity, liquid water content, and pressure. Of key interest in the understanding of polarimetric and Doppler data available from the radar are the measurements of aircraft position, orientation and velocity available from the on-board INS system.

5 Examples of Observations

5.1 Melting Layer 1992

A case of well-defined melting band was observed on October 31, 1992 about 200 km to the NE of Laramie, WY. Light rain was falling at the surface, consisting of drops to a maximum of 1 mm diameter. Above the melting level, the cloud consisted entirely of ice crystals, except for transient patches of low cloud liquid water content around 3.5 km altitude. Below 3 km, the air was not fully saturated, leading to evaporation of ice crystals falling from above; this decreased the precipitation rate but did not lead to complete evaporation. The radar data shown in figure 7 were obtained during a gradual descent from 1.7 km to 1.45 km. The horizontal stratification of the reflectivity above 3.5 km shown in figure 7 is the result of wind shear in the regions of generating cells at cloud top. The weak echo region below 3.5 km is due to the partial evaporation mentioned earlier. The top of the lower echo layer is the 0 $°C$ level. The rapid increase in fall velocity at the melting layer is characteristic of such observations [14]. The region of fall velocity increase corresponds to a region of heightened LDR (-15 to -10 dB) at about 2 km altitude.

5.2 Melting Snow 1994

One observation of note from March 8, 1994, was of melting snow at horizontal incidence. Just about all of the polarimetric measurements give indications of some kind of variability in the cloud with range (figure 8). There is a region centered about 40 s along track through which the SNR decreases rapidly with range which may represent the presence of significant attenuation. This corresponds to the high reflectivity band close to the aircraft around the same time. The copolar correlation magnitude drops substantially in the region of high reflectivity as does the Z_{DR}. The LDR response shows up just inside the ring of high reflectivities and extends out to the visible range just before 40 s along track. During this observation, the temperature was approximately 0$°C$ and the ice and liquid water contents were nearly equal along the flight line, fluctuating between 0 and 0.20 g/m^3 and 0 and 0.40 g/m^3 respectively. The altitude was about 3310 m along most of the track up until 70 s. The aircraft maintained a steady heading with minimal changes in roll ($\pm 2°$) and pitch ($\pm 1°$) up until 70 s. An example of the 2DC PMS particle images is seen in figure 5. The total width of the image is 800 μm. Large aggregates are clearly visible in this time segment and extend well beyond the edges of the imaged region.

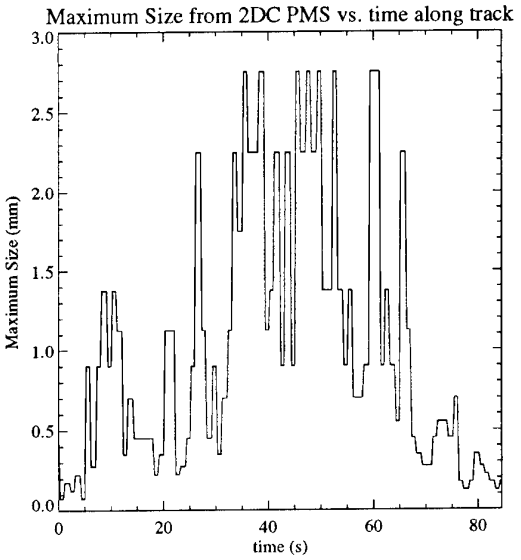


Fig. 2. Maximum particle size indicated by 2DC PMS probes for observation of melting snow.

The Z_{DR} throughout much of the cloud indicates oblate particles are being imaged ($Z_{DR} > 0$ dB), but in the region of high reflectivity, actually goes below 0 dB, which might indicate prolate particles. Snow at horizontal incidence would be expected to be oblate and to therefore produce a Z_{DR} greater than 0 dB. For scatterers in the Rayleigh region, a Z_{DR} of less than 0 dB indicates that the mean axial ratio of the scatterers observed is less than one. However, resonance region backscatter can bring about oscillations in the backscattered power that foil simple interpretation of Z_{DR} measurements [2]. The maximum particle sizes as measured by the 2DC PMS are presented in figure 2. The times along track for which the Z_{DR} decreases substantially from 0 dB (by more than 0.5 dB) correspond to maximum particle sizes above 2 mm, which could mean that the backscatter in this region was influenced by resonance region scattering. The copolar correlation magnitude decreases throughout this region as well.

5.3 FFT Data on Stratus 1995

Full Doppler spectrum estimate data was taken on coastal marine stratus clouds during the CS experiment in Oregon during August and September of 1995. Images of Z_{eHH}, fall velocity and Doppler spectral width estimated using the measured spectra are provided in figure 3. A plot of a typical Doppler spectrum estimate obtained during this period is shown in figure 4. The number of bins used in this case was 64 and the PRF was 15 kHz, yielding a fold-

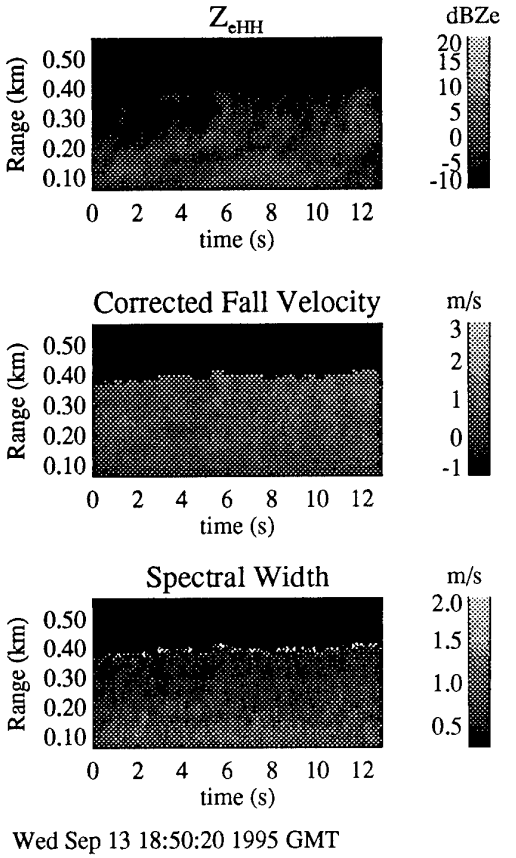


Fig. 3. Z_{eHH}, fall velocity and spectral width for observation of drizzle from marine stratus observed during CS experiment in Oregon.

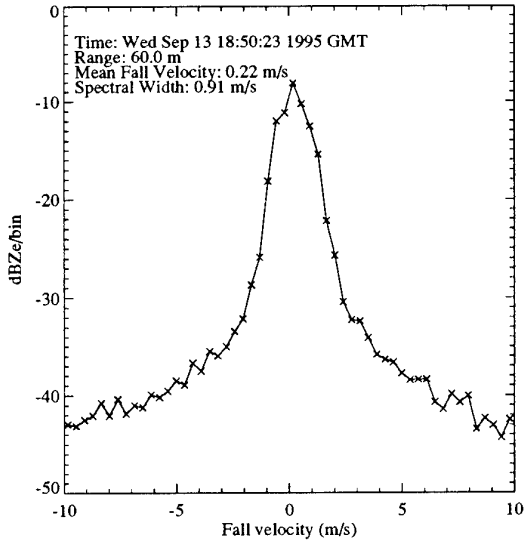


Fig. 4. Example of Doppler Spectrum Estimate for observation of drizzle from marine stratus observed during CS experiment in Oregon.

ing velocity of 11.9 m/s and bin width of 0.37 m/s. The data acquisition was averaging 32 spectrum estimates along track at a rate of 15.6 stored, averaged spectrum estimates a second. Given a mean airspeed of 85 m/s and a digitizer sampling rate of 10 MHz, this yielded image pixels about 5.5 m wide and 15 m long. The altitude was about 143 m for this time period. The pitch variation was less than $\pm 0.5°$ and the roll was $-1° \pm 0.3°$. The liquid water content estimated from the 2DC measurements was between 0.01 and 0.04 gm^{-3} and the temperature was about 12 $°C$.

References

[1] D. Atlas. *Radar in Meteorology.* American Meteorological Society, 1990.

[2] K. Aydin and Y. Lure. Millimeter Wave Scattering and Propagation in Rain: A Computational Study at 94 GHz and 140 GHz for Oblate Spheroidal and Spherical Raindrops. *IEEE Transactions on Geoscience and Remote Sensing*, 29:593–601, 1991.

[3] R.J. Doviak and D.S. Zrnić. *Doppler Radar and Weather Observations.* Academic Press, Inc., 1993.

[4] Peter H. Hildebrand, Craig A. Walther, Charles L. Frush, Jacques Testud, and Francois Baudin. The ELDORA/ASTRAIA Airborne Doppler Weather Radar: Goals, Design, and First Field Tests. *Proceedings of the IEEE*, 82:1873–1890, December 1994.

[5] A. R. Jameson. The Interpretation and Meteorological Application of Radar Backscatter Amplitude Ratios at Linear Polarizations. *Journal of Atmospheric and Oceanic Technology*, 6:908–919, December 1989.

[6] R. M. Lhermitte. Cloud and precipitation remote sensing at 94 GHz. *IEEE Trans. Geosci. Remote Sens.*, 26:207–216, 1988.

[7] Sergey Y. Matrosov, Roger F. Reinking, Robert A. Kropfli, and Bruce W. Bartram. Estimation of Ice Hydrometeor Types and Shapes from Radar Polarization Measurements. *Journal of Atmospheric and Oceanic Technology*, 13:85–96, February 1996.

[8] J.B. Mead, A.L. Pazmany, S.M. Sekelsky, and R.E. McIntosh. Millimeter-wavelength radars for remotely sensing clouds and precipitation. *Proceedings of the IEEE*, 82(12):1891–1906, December 1994.

[9] A. L. Pazmany, R. E. McIntosh, R. Kelly, and G. Vali. An Airborne 95 GHz Dual Polarization Radar for Cloud Studies. *IEEE Transactions on Geoscience and Remote Sensing*, 1, 1994.

[10] R. Rasmussen et al. Winter Icing and Storms Project. *Bulletin of the American Meteorological Society*, 73:951–974, July 1992.

[11] F. Ulaby and ed. Elachi, C. *Radar Polarimetry for Geoscience Applications.* Artech House, 1990.

[12] F. Ulaby, R. Moore, and A. Fung. *Microwave Remote Sensing; Active and Passive vol. 1.* Artech House, 1982.

[13] G. Vali, R. D. Kelly, A. Pazmany, and R. E. McIntosh. Airborne Radar and In-situ Observations of a Shallow Stratus with Drizzle. *Atmospheric Research*, 38:361–380, 1995.

[14] D.S. Zrnić, N. Balakrishnan, A.V. Ryzhkov, and S.L. Durden. Use of Copolar Correlation Coefficient for Probing Precipitation at Nearly Vertical Incidence. *IEEE Transactions on Geoscience and Remote Sensing*, 32(4):740–748, 1994.



212606.10 212606.74

Fig. 5. 2DC PMS probe images for observation of melting snow.

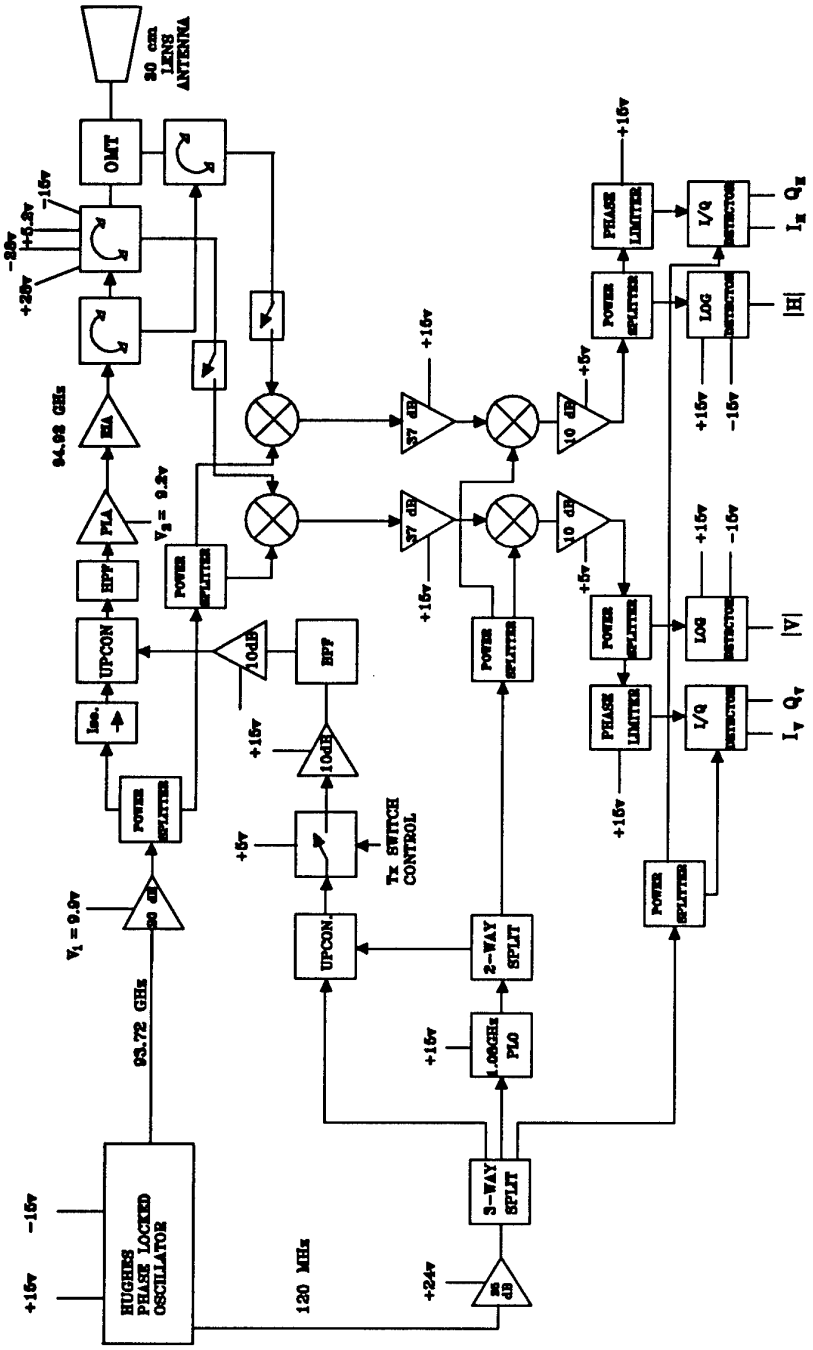


Fig. 6. Block diagram of RF and IF sections of 95 GHz airborne polarimeter.

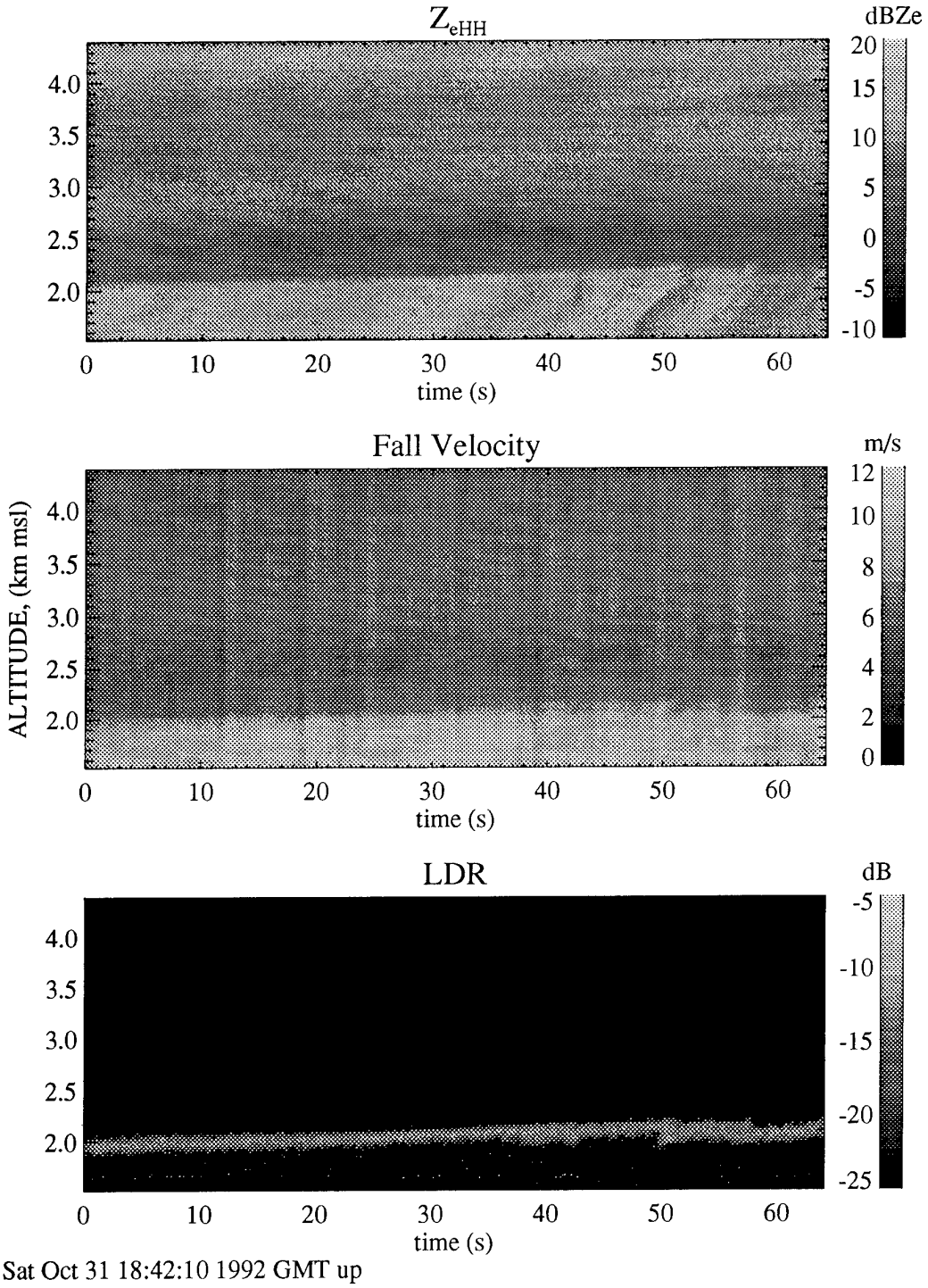


Fig. 7. Images of data taken on melting band on October 31, 1992 over Laramie, WY.

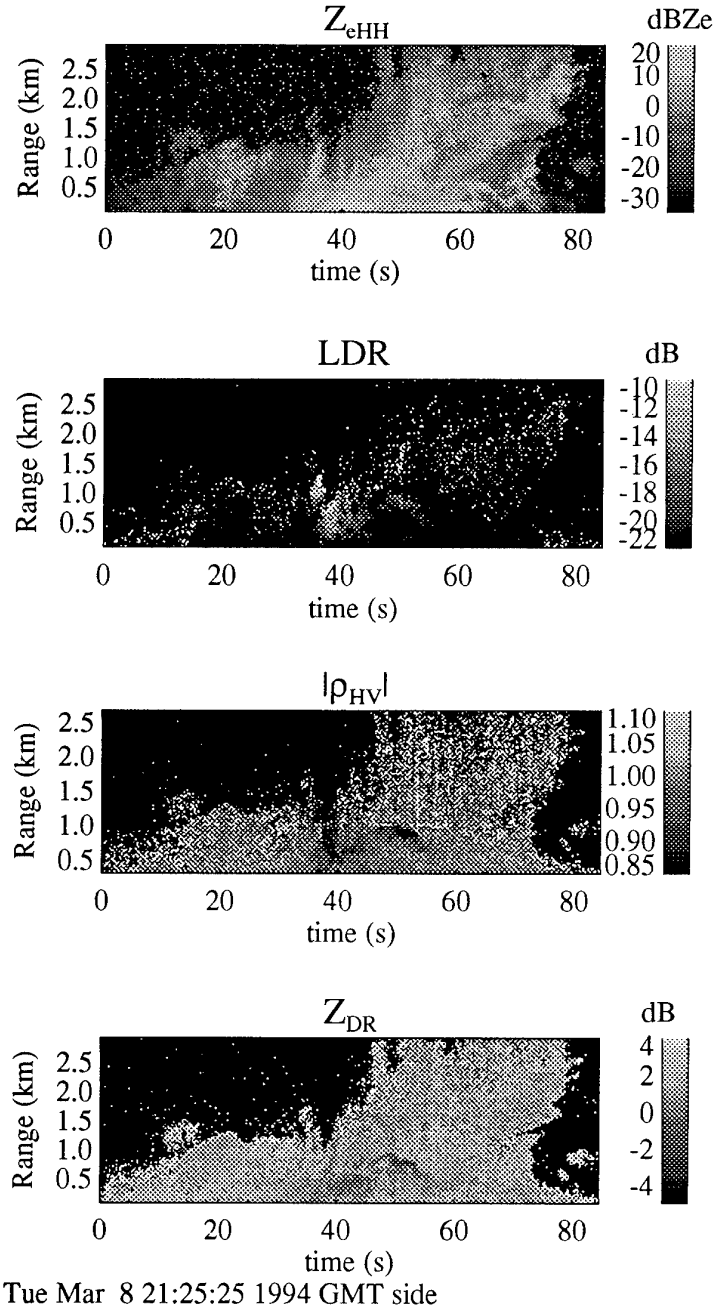


Fig. 8. Images of data taken on melting snow observed on March 8, 1994 over Oklahoma City, OK.

Imaging tools for spaceborne SAR programme

J.M. Hermer, E. Normant, P. Martynerie, J.P. Hardange
THOMSON-CSF, RADARS & CONTRE MESURES, 1 Bld J. Moulin, 78852 Elancourt Cedex - France
J.M. Boutry
ONERA, Fort de Palaiseau - 91120 Palaiseau - France
U. Erich
Daimler-Benz Aerospace, Dornier Satellitensysteme GmbH, 88039 Friedrichshafen

1. ABSTRACT

The image chain engineering for a spaceborne SAR requires the support of simulation tools and airborne demonstrators.

In the frame of the future German/French spaceborne SAR programme, several of these means are already currently used. Others are under development or under project.

The main simulation tool in France is the SIROS simulator, specified and operated by CELAR in Rennes. In Germany related tools are presently being developed by Dornier and DLR, i.e. SAR parameter synthesis and SAR performance analysis.

Four airborne demonstrators are actually in operation:

In Germany, DOSAR from DORNIER and E-SAR from DLR,

In France, RAMSES from ONERA and RENE from THOMSON-CSF and CETP.

The airborne radar demonstrators are useful to collect data in order to determine the characteristics of the clutter and of the targets and identify all the physical phenomenons related to the interactions between the electromagnetic wave and the targets in their environment.

The future spaceborne radar will have image quality characteristics and detection capabilities which will require the use of advanced techniques such as Spotlight imaging, space time adaptive processing, jammer cancellation. The validation of those techniques will be performed through a complete programme of qualification and test requiring simulations and in flight demonstration.

Besides ground based active antenna demonstrator, an advanced airborne demonstrator is needed, in order to enhance the existing means.

This advanced airborne demonstrator is to be made with existing sub-assemblies, while being equiped with a wide active antenna and installed onboard a high altitude high speed aircraft such as a MYSTERE 20 or a TUPOLEV.

2. INTRODUCTION

2.1. Challenging aspects of a Spaceborne SAR programme

The primary mission objectives of a spaceborne surveillance system comprise continuous and comprehensive intelligence gathering in peacetime and in crisis or conflict scenarios, for instance with respect to verification of international treaties and crisis monitoring.

Due to the necessity of weather independent observations, a satellite with a 'Synthetic Aperture Radar (SAR)' as a major system element will be applied for the detection, recognition and identification of man-made and natural targets. The potential requirements for this radar system, e.g. different measuring modes providing high or even extremely high resolution and highly flexible operations and agility within a wide access range, result in a challenge to both SAR *system design* and SAR *technology*.

2.2. Image Chain engineering

In order to simplify the SAR *system design* and have unified engineering activities, it seems reasonable to share the engineering process between functional chains. This functional approach is complementary to the physical decomposition of the system, which comprises the spacecraft with its radar payload, the communications and the ground segment (see figure 1).

For the spaceborne SAR system, the following functional chains can be proposed:

The mission functional chain, dealing with system time performance, platform service module, number and location of ground stations, image delivery, ...

The exploitation functional chain, dealing with image databases and image exploitation (including multi-sensor fusion),

And the image functional chain, dealing with all aspects related to image quality.

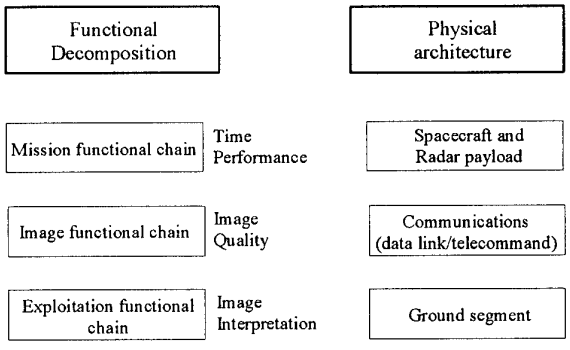


Figure 1: Physical and functional system decompositions

Seeing the number of possible configurations, the most important challenge for the image chain engineering is

Paper presented at the AGARD SPP Symposium on "Remote Sensing: A Valuable Source of Information", held in Toulouse, France from 22-25 April 1996, and published in CP-582.

to yield the exact configuration matched to the requirements of the end-user of the system. This challenge expresses in the following activities:

Image **Q**uality **R**equirement Specification (IQRS):

Analysis and quantification of the operational requirements; specification of the image quality parameters required to meet the operational requirements,

SAR **C**oncept **D**esign (SCD):

Design and specification of the main functions of the image chain, dealing with waveform and imaging modes, antenna architecture and patterns, raw data compression algorithms, ...,

Detailed SAR Specifications (DSS):

Specification of the lower level image chain components, including propagation effects,

Image **C**hain **B**udget **M**anagement (ICBM):

Revision of the image chain performances in function of the design and development results; management of the specifications allocated to the image chain components,

which also involve iterations at different engineering levels:

Because some of the specifications are shared by several functional chains and require system trade-offs (e.g., the altitude and 'range of incidence angles' requirements affect not only the image and mission functional chains but also the spacecraft and radar engineerings),

Because the solutions and specifications proposed for the image chain impact on the instrument feasibility, cost and development plan.

2.3. Imaging tools

In order to optimize the image chain, concurrent tools are required to support the engineering:

The simulators, which can simulate specific test patterns or operational image scenarii, participate in the image quality budget determination and in the image chain configuration validation,

The airborne demonstrators, which yield the real data required to model target and terrain characteristics, enable to verify that all the useful radar physics is allowed for and validate radar concepts and technology.

This paper, beyond the simple description of the existing simulators and demonstrators in France and Germany, describes the role that such means play in the image chain engineering activities. It also puts forward the need for an advanced airborne demonstrator, which will enhance the existing means and in operation validate the solutions proposed for the spaceborne SAR system (imaging modes, image quality performance, technology).

3. DEMONSTRATORS AND SIMULATORS IN THE IMAGE CHAIN ENGINEERING

3.1. Image Quality Requirement Specification (IQRS)

The IQRS activity makes use of performance charts, which summarize the link between the image quality specifications and the operational requirements. As an example, some of these charts may draw the link between target detection probability and contrast/resolution.

Different means exist to generate and validate those charts:

Theoretical calculations based on either standard or refined target and terrain models (e.g., Gaussian terrain models, Swerling target models),

Analysis of real images, to validate and improve the target and terrain models,

Photo-interpretation of images (the validation of the image chain performance by the end-user of the system is of primary importance) and analysis of photo-interpretation results,

Development of analysis software, for image performance evaluation (some of the operational exploitation software may, beyond the objectives of image chain engineering, be derived from this one) or image parameter optimization (e.g., from a multipolarisation image, synthesis of the optimal polarization matched to target detection).

Complementary use of demonstrators and simulators permits to optimize (in term of cost effectiveness) this chart generation activity. Simulation makes it possible to generate at lower cost a wide set of images covering different operational situations, while systematically scanning the image parameter values (resolution, incidence, contrast, ...). In return, airborne demonstrator is used to collect experimental data and real images.

Additional means, such as measurement stations (like SARAPE or STRADI operated by CELAR), also permit to gather experimental data (with limited flexibility and image quality performance, but lower cost than flight campaigns).

3.1.1. Simulation

It is of prior importance for the simulator to take into account, at the appropriate level[1], the interactions between the electromagnetic waves and the targets. For this, it is assisted by:

Electromagnetic simulators (capable of generating electromagnetic signatures from target CAD files),

Real target measurements (performed by demonstrators or measurement stations).

It can thus easily generate numerous realistic images for different incidence angles, polarisations or carrier frequencies, depending on the progress of the SCD activity (IQRS and SCD activities are complementary).

At this level, the simulator may not be representative from the operational radar (in terms of imaging modes for instance).

3.1.2. Demonstrators

The experimental data collected by the demonstrators (over different terrains and military targets) can be processed at different resolutions and with different number of looks. Different NEσ0 values can also be simulated by adding noise to the images.

It is required from the demonstrator chain to be calibrated in order to deliver images whose characteristics are duly controlled (resolution and impulse response parameters, NEσ0 value, geometry of observation, polarisation, ...). Absence of image colouring by the demonstrator must be verified.

In parallel to the image acquisition, ground truth must also be carefully recorded (in terms of target description and positionning, or terrain nature and wetness, ...).

3.2. SAR Concept Design (SCD)

This activity yields the main radar parameters (without allowing for the sub-systems limitations): carrier frequency, polarization requirements, geometry, imaging modes, PRF, peak power, pulse length, main antenna parameters, transmission rate, compression algorithm, ... It represents the first step in instrument feasibility assessment.

---

1 i.e., with an accuracy which depends on the image resolution and operational requirements.

3.2.1. Simulation

Parametric analysis software, which may be part of the simulator, are used to determine the main radar parameters.

Simulation, for its part, permits to validate the high level functions of the image chain by generating realistic SAR data (i.e., corresponding to the actual modes of the spaceborne SAR) and also processing them with the actual algorithms.

As a consequence, the realism of the simulator makes from it a flexible platform for the development and evaluation of advanced modes and algorithms. More than a simple tool, simulator becomes a real support to the engineering creativity.

Simulations are either performed on test patterns, which permit to evaluate the image quality parameters, or on operational scenarii, which deliver realistic images as suited by IQRS.

3.2.2. Demonstrators

Existing demonstrators and operational air- or space-borne SARs, which are part of the industrial experience, contribute in this respect to the SCD activity.

However, a specific and representative airborne demonstrator represents an experimental way to validate the actual solutions proposed for the SAR system, in terms of imaging modes implementation, image quality performances and also in operation technology demonstration.

The hardware of this representative demonstrator will not be developped specifically, but merely be picked up from (or based on) the sub-assemblies developped for the programme (e.g., the breadboards or the engineering models).

In consequence, for planning reason, it will mainly contribute to the ICBM activity, in terms of solution validation and risk reduction (cf. section 3.3.2.).

3.3. Detailed SAR Specifications (DSP) and Image chain budget management (ICBM)

3.3.1. Simulation

For DSP or ICBM activities, the image chain simulator permits to specify (or evaluate) the characteristics required for (or the degradations caused by) the equipements: e.g., transmitter and receiver non linearities, local oscillators phase noise, ambiguities, ... Trajectory perturbations (vibrations, errors in Doppler parameter estimation) and atmosphere effects are also taken into account in these simulations.

As introduced in 3.2.1., test patterns are used to evaluate the impacts of the specified (or measured) radar equipments on the images.

3.3.2. Demonstrators

With the representative demonstrator introduced in section 3.2.2., the following topics covering solution validation and risk reduction can from now be identified:

- Active antenna validation by operating the ScanSAR and Spotlight modes,
- In operation validation of the whole image chain, including calibration process and real time receiver gain adjustment,
- Validation of ECCM performance and evaluation of the impact on the image chain of non intentional intruders,
- Validation of moving target modes.

This representative airborne demonstrator is all the more desirable as there is presently no european airborne radar with an active antenna.

In consequence, due to the risk reduction that it permits and taking into account the limited expense for its development (especially when compared to the programme cost), the need for this representative demonstrator is beyond all question.

Preliminary design and development plan materials are presented in section 6.

4. CURRENT ACTIVITIES IN SIMULATION

4.1. SIROS

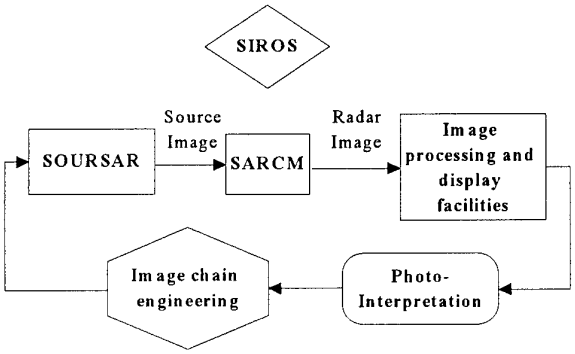


Figure 2: SIROS, a global environment for SAR image chain simulation

SIROS[2] is a global simulation environment developped under contract for french MOD to accurately simulate a spaceborne SAR. Its subsets have been specified and validated by CELAR[3], which, as an independant entity, warrants the performance level and credibility of this tool.

CELAR, which operates the simulator on its CRAY (YMP-EL, J-916) also warrants the representativity of the terrain and target signature databases required for the simulations.

SIROS (See figure 2) is made of several subsets:

- Generation of test patterns (SARCM module) and of operational image configurations (SOURSAR module), including terrain images and target signatures, with the positionning of the targets as indicated in the configuration plan,
- Simulation of the SAR images (SARCM module), according to the spacecraft geometry and image chain characteristics,
- Image processing and display as well as image evaluation facilities, in interface with the photo-interpreters (TREXIM module).

The SARCM subset [1], developped by Thomson-CSF, is an ideal tool to draw links between operational requirements and image parameters (IQRS activity). It permits to establish and validate image quality budgets at either SAR concept level (SCD activity) or at subsystem level (DSP and ICBM activities). To do this, SARCM comprises:

- A parametric analysis module (PRF calculation, Doppler parameter computation, signal to noise ratio and signal to ambiguity ratio calculations[4]),
- A test pattern generator (bright dot over dark background, uniform background, ...),
- A raw data generator,

---

2 SIROS stands for Simulateur de Radar d'Observation Spatiale (Spaceborne SAR Simulator)

3 CELAR stands for Centre d'ELectronique de l'ARmement

4 From a real or simulated antenna diagram.

A SAR processing unit (capable of processing real SAR data, like ERS-1/2),

A post-processing unit (including range and azimuth multilooks, in time and frequency domains),

An image quality performance measurement module.

The raw data generator includes the following:

Stripmap (including squinted version) and spotlight (fixed and extended) modes generation,

IFOV[5] and speckle generation,

Point targets and diffuse scatterers ambiguity simulation,

Raw data compression algorithms, including BFPQ, BAQ and VQ, and Bit Error Rate simulation (plus the means to measure the degradation of the impulse response quality or of the SNR due to the quantization noise),

Trajectory perturbations (vibrations, errors in Doppler parameter estimation),

Atmospheric perturbations (random delay process),

Transmitter/receiver and local oscillators characteristics.

The principal SAR processing algorithms are implemented in SARCM: Range Doppler, Secondary Range Compression and Exact Transfer Function.

SARCM is perfectly calibrated in amplitude and phase. Developped in standard C-ANSI language with high software quality standard (compatible with GAMT17-V2 DoD-STD 2167), it is portable on different types of workstations.

Optimized for a 4-processor CRAY YMP-EL, J-916, it takes 7 minutes to generate a 1 meter resolution X-band image from a source image of 2048 x 2048 pixels, oversampled by 2.

4.2. Relevant German Activities

Simulation of SAR images for a spaceborne surveillance system has been a field of high interest in Germany for a long time and related activities have been increased in connection with studies concerning the overall system, i.e. a national study called RESYS during the years '89 to '90 and the Main System Frasibility Study of the Western European Union in the period '92-'93.

Besides measurements of radar cross sections of individual targets in a test site, flight campaigns are at the centre of attention using the following systems: E-SAR from the German Aerospace Research Establishment (DLR) and DOSAR from Daimler-Benz-Aerospace/Dornier.

The primary goal of such campaigns is to measure SAR signatures of terrain with different vegetation cover and of relevant man-made targets and natural objects. The images are being used to verify theorerical findings w.r.t. measuring features in different frequencies and polarizations. But they are also used to proceed with the interpretation of SAR images, i.e. development of image data interpretation and analysis tools and training of interpreters.

Software simulation tools are also being developed in Germany. They represent a step-by-step approach to achieve finally a SAR End-to-End Simulator (see figure 3) which will be adapted to the special operational modes such as Spotlight SAR, ScanSAR and Look Steering SAR. Based on experiences from former simulations related to different spaceborne microwave-instruments in the civilian field, significant progress has been made concerning the individual elements of the End-to-End Simulator. Final results will be available in the near future.

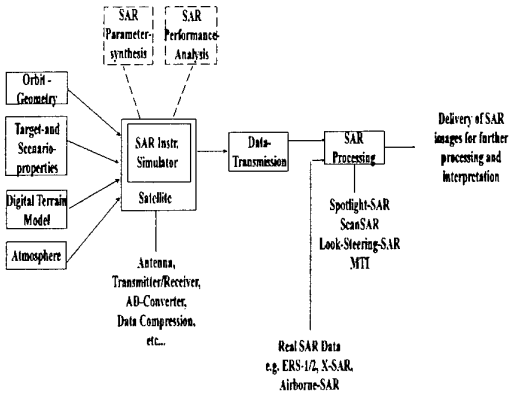


Figure 3: SAR End-to-End Blockdiagram

5. EXISTING AIRBORNE DEMONSTRATORS

5.1. DOSAR

DOSAR is an airborne SAR sensor. Its basic version is integrated in a DO-228 aircraft and can operate in C- and Ka-band. The second version is integrated in a Transall aircraft and can operate additionnally in X-band and is presently being modified to provide also S-band capability.

The DOSAR sensor consists of the following functional units:

Steerable Antenna Unit in Elevation,

RF Transmit and Receive Electronics,

SAR Data Unit,

Instrument Controller,

SAR Quick Look Processor,

Video Recording,

Tape Recorder for SAR Data Storage,

IMU and GPS for Platform Motion Compensation Registration.

The main system parameters of DOSAR are summarized in Table 1.

The DOSAR sensor is at this time capable of operating at 5.3 GHz, 9.6 GHz and 35 GHz. Using the basic version for a DO-228 aircraft the following radar modes are possible:

Polarimetric modes (HH, HV, VH and VV simultaneously),

Along track interferometry (MTI),

Accross track interferometry (stereo).

In the case of the DOSAR version for the Transall aircraft an additionnal X-band reflector antenna is available at the antenna platform, which is mounted in the parachuting door of the aircraft. Along track interferometric modes are not possible with this version because the required antennas for this purpose are not mounted. In a few months an S-band SAR will be implemented in which case the C-band subsystems will be replaced. Simultaneous imaging with X-band and S-band will then be possible.

The present DOSAR platform motion registration system consists of a sensor package (gyroscopes, accelerometers, GPS receiver, differential GPS), of a sensor computer for the acquisition, presuming and storage of inertial information as well as of the IMU

---

5 IFOV stands for Instantaneous Field Of View. IFOV simulation warranties the signal representativity in terms of power and statistics and makes it possible to simulate A/D converters or analyse compression algorithms.

software package. Thus even the highest spectral components of the aircraft motion can be monitored.

Velocity increments are transformed to the earth-fixed coordinate system and integrated to get velocity and position in all three coordinates. GPS data are used to reach long term stability of along track velocity and localization accuracy.

Parameter	C-band	X-band	Ka-band	S-band (in preparation)-
Transmit frequency	5.3.GHz	9.6 GHz	35 GHz	3.1 GHz
Wavelength	5.6 cm	3.1 cm	0.85 cm	9.6 cm
Peak power	1 kW	2kW	1 kW	2 kW
Antennas	Reflector (VV, VH, HV and HH-polarization) 3 slotted waveguide (HH-polarization) for interferometric modes	Reflector (VV, VH, HV and HH-polarization)	Slotted waveguide (HH-polarization)	Reflector (VV, VH, HV and HH-polarization)
Antenna gain	20 dB	23 dB	20 dB	18.5 dB
Interferometric mode	Azimuth basis = 0.6 m Elevation basis = 1 m			
Bandwidth	50, 100 or 200 MHz			
Resolution	1 - 3 m			
sampling rate	60, 120 or 240 MSamples/s			
Pulse Repetition Frequency	0.75/1.5/3.0 or 6 kHz			
Quantization	8 bits (I and Q)			
IF-attenuation	0 - 44 dB (adjustable for each operating mode)			
STC-attenuation	0 - 42 dB (adjustable by steps of 33 ns)			
Range line buffer	1024, 2048 or 4096 samples			
Data rate	105 MBPS			
Swath width	500 m - 10 km (depends on altitude and incidence angle)			
Altitude	100 m - 3600 m			
Incidence angle	30 deg - 89 deg			
Pulse generation	Programmable (digital signal generator)			
Operating modes	4 different modes for simultaneous measurements			
Positioning	GPS			
Motion compensation	Integrated IMU controlled by GPS and barometric altitude			
Image processing	Quicklook-processor FlexSAR-processor (on ground) EXSAR (on ground, UNIX-workstation)			
Storage	AMPEX DCRSi cassette for raw data Video-8 recorder for video data DAT cassette for quicklook images CCT for ground-processed SAR images Diskettes for IMU- and GPS-data			

Table 1: Main DOSAR characteristics

5.2. E-SAR

The E-SAR is a multifrequency high resolution SAR system mounted onboard a DO-228 aircraft, which is owned and operated by 'Deutsche Forschungsanstalt für Luft- und Raumfahrt (DLR)'. At present the radar is operational in P-, L-, C- and X-bands with selectable vertical or horizontal antenna polarizations. SAR interferometry and SAR polarimetry are new functional modes of the radar, which are on their way to be integrated during 1996. The system consists of the on-board segment with the radar sensor, motion compensation (Inertial Measurement Unit), real time processor and of the ground segment comprising radar raw data transcription (HDDT, Exabyte, CCT) and SAR processing (auxiliary data processing, motion compensation, focused multi-look SAR processing including radiometric calibration).

The key parameters of E-SAR are summarized in Table 2.

Recent developments resulted in SAR interferometry in X band (across and along track single pass). Presently, besides improvements of technical performances (e.g., higher resolution in the X-band by implementing a larger bandwidth up to 200 MHz) effort is being put by DLR into the development of a new S-band radar segment as well as into the capability of recording 4 channels quasi-simultaneously at full data rate for multi-polarization or multispectral measurements.

RF-band	X-band	C-band	L-band	P-band
RF-centre frequency	9.6 GHz	5.3 GHz	1.3 GHz	450 MHz
Transmit peak power	2500 W	50 W	360 W	180 W
Receiver noise figure	3.5 dB	5.0 dB	8.0 dB	4.5 dB
Antenna gain	17.5 dB	17 dB	17 dB	12 dB
Azimuth beamwidth	17 °	19 °	18 °	30 °
Elevation beamwidth	30 °	33 °	35 °	≈ 60 °
Antenna polarisation	H and V	H and V	H and V	H and V
System bandwidth	120 MHz	120 MHz	100 MHz	25 (or 60) MHz

Table 2: Main E-SAR Characteristics

5.3. RAMSES

The RAMSES[6] station makes use of a coherent airborne radar, which permits to improve the azimuth resolution of the observed ground targets to a better value than the resolution corresponding to the azimuth aperture of the antenna. In the range dimension, high resolution is achieved through programmable waveform generation and digital pulse compression.

The RAMSES station is designed to permit the simultaneous operation of several frequency bands, comprised between L-band (1.6 GHz) and W-band (94 GHz) band. This multi-spectral capability is completed by optical instruments, in the visible and infrared wavelengths.

Most of the radar heads have multipolarisation capability, making RAMSES a comprehensive support to innovative studies (e.g., in the field of target identification).

The RAMSES station is mounted in a TRANSALL C 160 aircraft, where it is also possible to embark a modular and programmable instrumentation, thus making from RAMSES a flying laboratory.

In order to limit the system complexity, the raw data are not processed onboard. The airborne station includes navigation measurement units (GPS, ...).

Several imaging modes are possible:

Side looking is performed through the rear right door,

Axial (with or without beam scanning) or squinted observations are realised by installing the heads on the rear ramp of the Transall.

The principal characteristics of the radar are given in table 3.

5.4. RENE

Since 1990, Thomson-CSF and CETP[7] have jointly been developping and exploiting the high resolution polarimetric radar RENE. This dual frequency system (S and X bands, non simultaneous) is mounted onboard an ECUREUIL helicopter.

The RENE radar makes use of Direct Digital Waveform Synthesis for the generation of its CW LFM waveform. H and V polarisations are switched from pulse to pulse at transmit, while the two receiver channels simultaneously receive the two polarisations. This continuous radar comprises two separate antennas for transmit and receive. It makes use of deramping in IF to reduce the signal bandwidth up to a few MHz. A video camera mounted beside the antennas records the ground truth during the radar acquisitions.

The range of incidence angles is comprised between 20 and 60 degrees for both frequencies. The raw data, which are onboard recorded, are ground processed. The main processor functionalities are:

Pulse compression (simple FFT),

Motion compensation and autofocus,

Azimuth correlation,

6 RAMSES stands for Radar Aéroporté MultiSpectral d'Etude de Signatures (Multispectral airborne radar for cross-section analysis)

7 CETP (Centre d'étude des Environnements Terrestres et planétaires) is a scientific research laboratory of CNRS.

Absolute calibration in both amplitude and phase, achieved through the exploitation of stationary flights on diedral and triedral corner reflectors.

The main characteristics of RENE are put in table 4.

The modest performances of RENE in terms of swath width and range of observation are sufficient for the acquisition of small targets like tanks and aircrafts. In return, RENE is flexible, low cost and reliable (with proven calibration process and image delivery procedure).

During the operational campaigns, the helicopter carrier makes it possible to keep a direct and permanent link between all the participants, thus allowing to adapt to any imponderable and keep control over the ground truth.

Frequency band	L	S	C	X	Ku	Ka	W
Bandwidth	200 MHz	300 MHz	300 MHz	300 MHz	300 MHz	500 MHz	500 MHz
Transmitter	Solid state amplifier		TWT				E.I.A.[8]
CW power	100 W	100 W	500 W	200 W	200 W	100 W	50 W
Elevation beamwidth	23 deg.	30 deg.	33 deg.	15 deg.	13 deg.		3 to 20 deg.
Azimuth beamwidth	16 deg.	10 deg.	8 deg.	15 deg.	13 deg.		3 to 20 deg.
Transmit polarisation	V or H[9]			V or H, L or R[6]		L or R[6]	
Receive polarisation	V, H	V, H	V, H	V, H L, R	V, H L, R	L, R	L, R
Antenna technology	Array of dipoles			Horn			
Range	1 to 10 km						
Swath width	1 to 7 km						
Range of incidence angles	30 to 85 deg.						

Table 3: Main RAMSES characteristics

Frequency band	X	S
Elevation beamwidth	3 deg.	11 deg.
Azimuth beamwidth	3 deg.	11 deg.
Bandwidth	200 MHz	200 MHz
Transmit polarisation	H or V (switchable)	H or V (switchable)
Receive polarisation	H and V	H and V
Pulse length	2.62 ms	2.62 ms
PRF	167 Hz	167 Hz
Transmitter	Solid State Amplifier	Solid State Amplifier
Peak power	0.063 W	0.063 W
Range	200 m to 2 km	200 m to 2 km
Range of incidence angles	20 to 60 deg.	20 to 60 deg.
Swath width	200 m	200 m

Table 4: Main RENE characteristics

8 Extended Interaction Amplifier

9 Switchable from pulse to pulse

6. ADVANCED DEMONSTRATOR DESIGN AND DEVELOPMENT PLAN

6.1. Requirements

The advanced demonstrator answers several requirements:

Reduce the cost of the flight campaigns and improve operational flexibility and efficiency,

Improve flight conditions in terms of flight stability and observation conditions (higher altitude and same illumination time as for the spaceborne radar),

Offer growth potential, in order to permit, during the progress of the programme, the validation of the actual spaceborne SAR hardware and modes.

6.1.1. Improvement of operational efficiency

Data gathering with the existing demonstrators has already begun in both France and Germany. This has permitted to start the image quality studies and get preliminary image specifications.

In the near future, intensive image acquisition will become necessary to validate the image specifications over a wide quantity of targets and operational conditions (observation angles, terrains, ...):

In a first time, these acquisitions will support the determination of the main image chain parameters,

In a second time, they will contribute to the image chain optimization (reduction of design margins together with accurate determination of system capabilities).

In this context, the cost of the campaigns and the efficiency of the demonstrator (capability to rapidly switch from one site to an other distant one or to rapidly process and deliver calibrated images) will be crucial.

Installation of the radar onboard a standard (and, hence, always available) aircraft is also required to favour efficiency and flexibility.

6.1.2. Improvement of flight conditions

The main objectives of these flight condition improvements are:

Aircraft altitude increase, in order to reduce the rate of the incidence variation accross the swath (constant incidence angle accross swath for a spaceborne SAR image),

Illumination time adjustment, in order to have it the same as for the spaceborne SAR (this can be important for the validation of an ocean mode, where the sea and ships always move).

Better flight stability only permits to use simplified processing algorithms or to reduce the risk of having bad images.

6.1.3. Capability of growth potential

Arguments for a representative demonstrator have been presented in sections 3.2.2. and 3.3.2. Additionnally, the idea of 'growth potential' puts forward the need for a flexible demonstrator, that will be gradually completed by representative hardware issued from the different developpements, along the programme.

6.2. Demonstrator design and development plan

In order to improve operational efficiency and flight conditions, it is proposed to make use of a MYSTERE 20[10] aircraft:

Its altitude (cat. 11 000 m) is 3 times heigher than the Transall altitude,

Its velocity is about 210 m/s (vs. 75 m/s for the Transall).

The required MYSTERE 20 altitude involves pressurized flight conditions. In consequence, it is proposed to install the antenna in a ventral pod. It will then become possible to change the antenna dimensions (up to say 2.5 m x 0.25 m), which were on a Transall aircraft constrained by the rear door dimensions.

The following development plan is proposed for the advanced demonstrator:

1. Development of the ventral pod and installation of existing radar hardware in the MYSTERE 20; adaptation of the existing software to process the raw data acquired by the new MYSTERE 20 demonstrator.

2. As far as the programme progresses, integration of actual spaceborne SAR equipements in the demonstrator for in operation validation.

As mentioned above, specific attention has to be paid to the active antenna. At each phase of this sub-system development (from exploratory development to engineering or flight model), extra components (such as T/R modules, radiating sources, ...) will be picked up from the production line in order to develop at lower cost the wide active antenna of the airborne demonstrator. This strategy will thus permit to secure the antenna progressively.

7. CONCLUSION

The incoming of the spaceborne SAR programme in Germany and France will involve intensive use imaging tools, like airborne demonstrators and simulators.

Required to specify the image chain of the future system and to validate the proposed technical solutions, these different tools are operated concurrently, so as to minimize their exploitation cost:

Simulators are used to generate rapidly a wide set of images corresponding to various operational situations and image parameters,

The airborne means permit to acquire realistic data, which also serve in validating the simulation models.

Even if the simulators permit to accurately model the spaceborne radar (including its geometry and operational modes), the airborne validation of the instrument and of its modes is of great interest, since it does not represent an additional cost with regard to that of the programme.

So, beyond, use of existing demonstrators, it is proposed to build a new and representative demonstrator, based on the radar sub-systemes developped along the spaceborne SAR programme. Progressively upgraded, this demonstrator will, in particular, permit to validate the use of an active antenna for ScanSAR or Spotlight modes.

---

10 The demonstrator could also be installed onboard a TUPOLEV (e.g. TU 154 M). With this solution, the SAR antenna could be placed under the aircraft and protected by a radome.

More, additional requirements, such as improved operational flexibility, improved flight conditions and growth potential requirement, tell in the favour of a new carrier. In this context, use of a high altitude, high velocity carrier (e.g., MYSTERE 20 aircraft), with the radar installed in a pod, constitutes an ideal solution to meet these requirements.

8. REFERENCES

[1] Normant, E., "SARCM, A SAR Simulator Dedicated to the Technical Specification and Operational Evaluation of an Image Functional Chain", to be published in EUSAR'96 proceeding.

SARAPE: UN OUTIL DE MESURES RADAR DU FOUILLIS

SARAPE: A TOOL FOR RADAR CLUTTER MEASUREMENT

Jean-Claude MOTET, B. MARECHAL
CELAR 35170 BRUZ FRANCE

Christian COCHIN
CRIL INGÉNIERIE 35517 CESSON-SÉVIGNÉ FRANCE

Régis GUILLEREY
Alcatel TITN Answare 35510 CESSON SÉVIGNÉ FRANCE

André BERGES
CERT/ONERA 31000 TOULOUSE FRANCE

Résumé:

Le CELAR (le Centre d'Electronique de l'ARmement) effectue des travaux d'essais et d'expertises concernant des systèmes radar d'observation et de détection au profit du Ministère de la Défense français. Pour remplir cette mission, il est amené à mettre en oeuvre des outils de simulation qui comprennent entr'autre la modélisation radar des cibles et de leur environnement. Ces modèles utilisent des données provenant de moyens de mesure radar haute résolution.

En ce qui concerne l'environnement, il met actuellement en oeuvre l'instrumentation SARAPE pour la mesure haute résolution du fouillis de sol. SARAPE a été conçu en laboratoire par l'ONERA/CERT, puis mis au point et intégré sur une plate-forme mobile par les équipes du CELAR. Cet outil complètement mobile sur le terrain permet de générer des images radar polarimétriques haute résolution de scènes à site d'observation variable dans la bande 2-18 GHz. Des résultats d'une campagne de mesure menée fin 1995 seront présentés.

Le système SARAPE comprend des outils logiciels permettant de produire des images radar polarimétriques étalonnées. Ces images sont utilisées pour caractériser des types de fouillis. En particulier, il est possible de calculer pour une partie de l'image correspondant à une surface homogène la rétrodiffusion moyenne (σ_0), les paramètres de la matrice de covariance Σ et des motifs de texture radar. L'outil logiciel permet de présenter à l'opérateur sur un écran informatique, outre l'image radar, des vues optiques de la scène mesurée.

Enfin le système SARAPE comprend une fonction d'archivage des résultats dans une base de données, actuellement en cours de constitution. Des outils d'interrogation permettent d'extraire en fonction de différents paramètres des données de rétrodiffusion permettant de faire des études thématiques ou d'alimenter des simulations radar.

Abstract:

CELAR (the Centre d'Electronique de l'ARmement) carries out trials work and expert analysis on surveillance and detection radar equipment for the french Ministry of Defence. For this role it has had to develop simulation tools which include among other things a facility for the radar modelling of targets and their environment. These models use data derived from high resolution radar measurement equipment.

As far the environment is concerned, it is currently using the SARAPE instrumentation system for the high resolution measurement of radar ground clutter. SARAPE was designed in the laboratory by CERT/ONERA and then finalized and installed on a mobile platform by a team from CELAR. This equipment, which is fully land mobile, enables the generation of high resolution polarimetric radar images of scenes at variable observation elevations in the 2-18 GHz range. Results from a measurement campaign carried out at the end of 1995 will be presented.

The SARAPE system includes software elements that enable the generation of calibrated polarimetric radar images. These images are used to characterise types of clutter. In particular it is possible to calculate the mean back scatter (s_0),for an area of the image corresponding to a homogenous surface, the covariance matrix parameters S and the radar texture pattern. The software enables presentation to the operator, on a computer screen, of an optical view of the scene being measured in addition to the radar image.

Finally the SARAPE system includes a function for the storage of results in a data base that is currently being created. Interrogation tools allow extraction, as a function of different parameters, of various back scatter data enabling thematic studies to be carried out, or for input into radar simulation.

Paper presented at the AGARD SPP Symposium on "Remote Sensing: A Valuable Source of Information", held in Toulouse, France from 22-25 April 1996, and published in CP-582.

1 - INTRODUCTION

Le CELAR effectue des travaux d'essais et d'expertises de matériels militaires divers. Dans ce cadre, il a en charge la mise en oeuvre de moyens de mesure et de calcul radar. Depuis plus de vingt ans, il met à la disposition du ministère de la Défense des moyens bien adaptés pour l'étude de la furtivité radar des cibles militaires. Les signatures radar de clutter étaient jusqu'à ces dernières années plutôt traitées dans des secteurs civils dans le but de mieux connaître tous les éléments de la surface du globe terrestre. Dans un passé récent des événements mondiaux ont mis en avant les domaines de l'observation de la terre. Mener une opération militaire à l'extérieur nécessite une connaissance du terrain. Il faut, pour cela, mettre en oeuvre des outils opérationnels capables de fournir les renseignements nécessaires. L'objectif est de détecter et de reconnaître des structures d'intérêt stratégique intervenant sur le théâtre des opérations militaires. Ces structures ne sont pas seulement des cibles militaires mais aussi des infrastructures telles que des bâtiments. Par ailleurs des armements nouveaux comme les missiles de croisière nécessitent pour être vraiment opérationnels, une connaissance précise du terrain qui ne peut être fournie que par des capteurs aéroportés ou satellisés très performants. Les progrès importants effectués dans le domaine de la furtivité radar sont tels que le niveau des cibles peut devenir de l'ordre de grandeur du fouillis environnant. Il est alors nécessaire de mettre en oeuvre des outils complexes de traitement du signal qui nécessitent une bonne connaissance de la signature du fouillis.

Les essais effectués au CELAR utilisent souvent la simulation. Notamment plusieurs projets consistent à simuler des équipements radar divers: radar SAR d'observation de la terre ou radar de détection et de poursuite. Pour simuler un capteur radar il est nécessaire de modéliser la cible et son environnement. La réalisation de modèle de fouillis passe par une bonne connaissance du phénomène physique et par conséquent par la mesure expérimentale. C'est ainsi que le CELAR met en oeuvre la station de mesure mobile SARAPE. Les objectifs visés ont conduit à la réalisation d'une instrumentation multifréquence, multipolarisation et haute résolution. L'instrumentation est actuellement opérationnelle et les résultats d'une campagne de mesure sont présentés. Les opérations de mesure comprennent plusieurs phases qui sont décrites dans le schéma de la figure N°1.

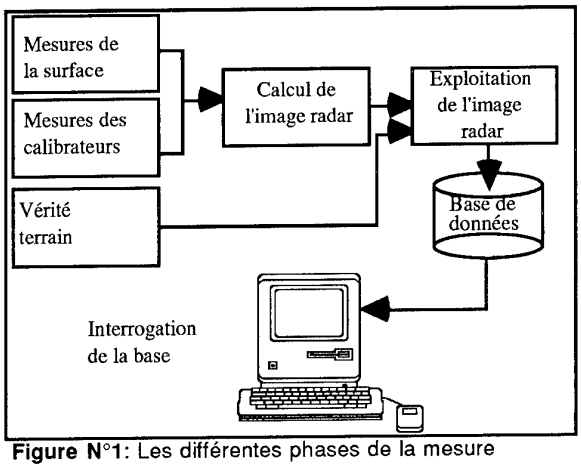


Figure N°1: Les différentes phases de la mesure

2 - PRÉSENTATION DE L'INSTRUMENTATION "SARAPE"

L'infrastructure de la base de mesure

La base de mesure SARAPE (Station Autonome Radar pour Acquisition de Paysages Electromagnétiques) est constituée d'un ensemble roulant motorisé, au gabarit routier, équipé d'une plate-forme élévatrice qui emporte la station de mesure jusqu'à 16 mètres au dessus du sol. La station fonctionne en visée latérale, par déplacement linéaire des antennes le long d'un rail de 3 mètres de longueur. Le choix de la hauteur de travail ainsi que de l'inclinaison des antennes permet de réaliser des mesures de réflexion de sol, avec des angles de site variant de 0° à 80°. La photographie ci-après illustre cette description.



Figure N°2: Photo de la station de mesure SARAPE

L'instrumentation de mesure (figure N°3)

Initialement conçue en laboratoire par le CERT/ONERA, elle a été ensuite mise au point et intégrée sur le porteur par le CELAR. Le principe de mesure consiste à émettre un signal hyperfréquence, dans la bande 2 à 18 GHz, qui peut être, au choix soit une onde continue, soit un signal modulé en impulsion. Ce signal est alternativement appliqué à une première antenne d'émission en polarisation linéaire verticale, puis à une seconde réglée en polarisation horizontale. La réception s'effectue alors de façon simultanée

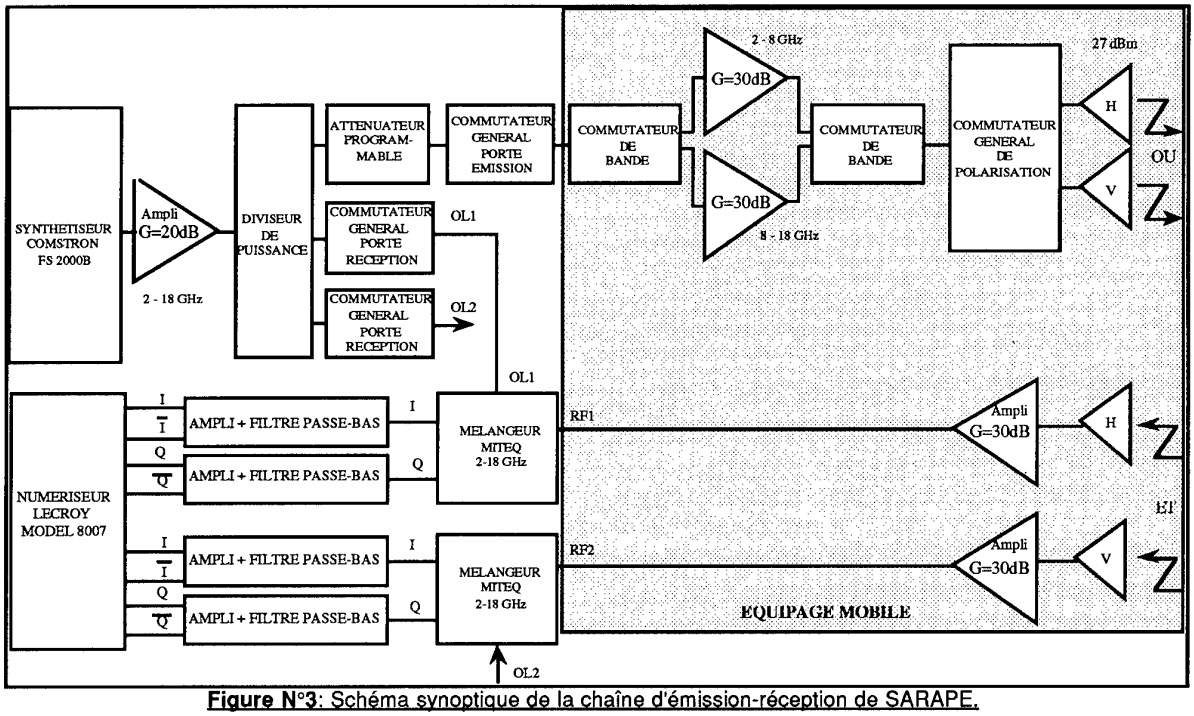


Figure N°3: Schéma synoptique de la chaîne d'émission-réception de SARAPE.

sur deux voies, l'une dédiée à la polarisation verticale et l'autre à la polarisation horizontale. La commutation de polarisation s'effectuant en moins de 200 ns, on considère que l'on accède ainsi à la matrice de rétrodiffusion quasi-instantanée des sols mesurés. Le signal reçu par les antennes est directement appliqué , après amplification, à des démodulateurs I/Q qui effectuent le mélange avec une partie du signal émis. Il en résulte, en sortie, des signaux continus, qui sont ensuite numérisés puis transférés vers le calculateur de mesure.

Paramètres caractéristiques

Hyperfréquence:

- Synthèse de fréquence directe 2-18 GHz.
- Temps de synthèse: 200 ns.
- Puissance émise: 1 Watt.
- Forme d'onde: CW ou impulsion rectangulaire.
- PRF: variable de 235 KHz à 1875 KHz.
- Facteur de forme: constant à 1/16.
- Ambiguïté distance: de 80 à 640 mètres.
- Polarimétrique linéaire horizontale et verticale.

Mécanique:

- Hauteur de travail: 3 à 16 mètres.
- Sites d'observation: de 0° (rasant) à 80° (spéculaire).
- Vitesse de balayage des antennes: 2 m/s maximum.
- Longueur du rail: 3 mètres.
- Échantillonnage sur le rail: 0.1 mm minimum.
- Dimension standard des scènes mesurées: 10m*10m.
- Mise en oeuvre: 2 personnes.
- Déplacement de l'ensemble: réseau routier normal.

divers:

- Prise en compte des paramètres météo : température, humidité du sol, pression atmosphérique ... pour chaque mesure

Principes de mesure et de construction d'images

Le principe essentiel consiste à émettre, pour une position d'antenne donnée, une rampe de fréquence à incréments discrets, dans une sous-bande choisie à l'avance. La transformée de Fourrier du signal reçu permet d'obtenir une réponse impulsionnelle et donc d'accéder à une localisation en distance des échos de sol. La résolution moyenne correspondante (selon l'axe radar) est donné par la formule suivante :

$$\delta r = \frac{c}{2*\Delta f}$$

où Δf représente la sous-bande explorée et c la vitesse de la lumière.

Ensuite, en déplaçant les antennes le long du rail, on recommence l'opération précédente pour une nouvelle position d'antenne. En balayant tout le rail, on accède alors à la résolution transverse qui est donnée en moyenne par :

$$\delta y = \frac{D*\lambda}{2*\Delta y}$$

où Δy représente la longueur balayée sur le rail, D la distance antennes/cible et λ la longueur d'onde moyenne de la sous bande explorée.

La reconstruction d'images holographiques radar est assurée par le logiciel LEADER [6], qui calcule les images reconstituées dans le plan horizontal. La résolution δr devient δx (par projection sur le plan horizontal) avec:

$$\delta x = \frac{\delta r}{\cos \phi}$$

où ϕ correspond à l'angle du site d'observation au centre de l'image.

Les fichiers de mesure sont enregistrés au format SEND [7], dont l'entête contient toutes les informations nécessaires pour la relecture, le traitement et l'enregistrement des résultats en banque de données.

Le calcul d'image est réalisé pixel par pixel. La valeur de chaque pixel est obtenue en identifiant, sur chaque réponse impulsionnelle (une par position d'antenne) la case distance qui correspond à la distance réelle du pixel considéré et en la sommant avec les autres cases correspondantes de toutes les réponses impulsionnelles qui ont constitué la mesure. On obtient alors le coefficient de rétrodiffusion du pixel choisi.

Des corrections sur les valeurs obtenues sont ensuite effectuées. Une correction appelée d^4 compense les différences d'atténuation de propagation en espace libre, car les points de l'image ne sont pas tous à la même distance des aériens. Une autre correction compense les atténuations dues aux lobes des antennes, lorsque les pixels ne sont pas situés directement sur l'axe radioélectrique des aériens, en approximant ces lobes à des lobes gaussiens.

L'instrumentation hyperfréquence est large bande, mais le principe de reconstruction d'image impose que la scène mesurée soit toujours, pour chaque position d'antenne, éclairée par le lobe d'antenne à -3 dB. Cela implique alors des contraintes sur le choix et le dimensionnement des aériens. Pour les scènes couramment mesurées (parcelles de 10 m par 10 m), des antennes de type cornet rectangulaire de 10° à 40° d'ouverture suffisent, si la sous-bande explorée ne dépasse pas 1 GHz. Cela donne des résolutions typiques de 15 à 20 cm en bande X.

Mode opératoire de mesure et principe de calibration

Afin de pouvoir exploiter les résultats de rétrodiffusion des sols, il est indispensable de calibrer soigneusement les mesures effectuées. La calibration polarimétrique n'est pas facile à mettre en oeuvre, de façon opérationnelle, sur le terrain.

La démarche utilisée est la suivante : la mesure de calibration s'effectue juste avant (ou juste après) la mesure de sol, sur un trièdre puis sur un dièdre, exactement comme pour des mesures aéroportées. La méthode de calibration, dérivée de la méthode IACT [3] nécessite un étalon dont la SER théorique co-polar est parfaitement modélisée (le trièdre) et un réflecteur ayant un coefficient de réflexion cross-polarisation élevé, mais pas forcément modélisé (le dièdre incliné à 22.5° ou 45° autour de son axe radioélectrique). Une fois ces deux fichiers de calibration réalisés, on effectue aussitôt la mesure des sols en supposant que les dérives de l'instrumentation sont négligeables.

La caractérisation de la scène

Un aspect incontournable pour l'utilisation et l'interprétation des mesures est la caractérisation de la scène. Un effort très important a été réalisé de manière à disposer pour chaque scène mesurée d'une vérité terrain très complète. Ainsi la météo est relevée (vitesse et direction du vent, pluviométrie, hygrométrie...). Les zones homogènes de la scène sont caractérisées par la nature du sol (type, humidité), la culture existante (variété, espèce, phénologie, hauteur, biomasse, direction et espacement des rangées, pentes du terrain).

L'exploitation de l'image radar (Planche 1)

L'exploitation consiste à extraire des caractéristiques de rétrodiffusion sur les zones homogènes de l'image .

L'exploitation comporte une partie graphique et interactive qui permet à l'opérateur de visualiser les zones homogènes définies lors de la mesure, de les modifier ou d'en définir de nouvelles. A ce niveau, il est possible de visualiser et d'éditer les paramètres associés à la scène (météo, caractéristiques radar ...) et à chacune des zones définies (nature , végétation, humidité ...).

Le logiciel "SHOWM" calcule pour chaque zone diverses caractéristiques de rétrodiffusion : l'image étant en densité de réflectivité, on obtient le σ_0 sur chacun des canaux de polarisation par la moyenne du module carré de cette réflectivité. Le comportement polarimétrique est exprimé par les termes complexes ρ, β et ε déduits de la matrice Σ des covariances polarimétriques mise sous la forme:

$$\Sigma = \sigma_{hh} * \begin{vmatrix} 1 & \beta.\sqrt{\delta} & \rho.\sqrt{\gamma} \\ \beta^*.\sqrt{\delta} & \delta & \varepsilon.\sqrt{\delta.\gamma} \\ \rho^*.\sqrt{\gamma} & \varepsilon^*.\sqrt{\delta.\gamma} & \gamma \end{vmatrix}$$

L'hétérogénéité de la zone est calculée par le rapport écart type sur moyenne du module de la réflectivité. L'incertitude sur la statistique est évaluée par une fonction du nombre d'échantillons et de l'hétérogénéité de la zone. Une zone peut être extraite pour constituer un motif de texture et utilisée dans le générateur d'images sources radar SOURSAR [8].

A chaque scène est associée une photographie globale et des photographies de détail des différentes zones. Ces photographies sont numérisées et référencées. L'ensemble des caractéristiques sont stockées dans une entête de fichier au format SEND.

Le logiciel d'exploitation graphique peut également être utilisé pour l'extraction de données à partir d'images de radar SAR aéroportés ou satellisés.

La base de données radar (Planche 1)

Les entêtes des images exploitées sont placées dans une base de données ORACLE. Le logiciel d'interrogation "OMNISAR" permet alors, depuis un ordinateur personnel ou une station de travail :

- de consulter les données stockées (avec visualisation des images radar et photos associées)
- de comparer les résultats à des données de provenances diverses (données publiées , données aéroportés ou satellite , base de données du CESBIO de Toulouse [9]...)
- de réaliser des extractions pour représentation sous EXCEL ou pour export vers les applications de simulation d'images radars.

3 - RÉSULTATS D'UNE CAMPAGNE DE MESURE AVEC "SARAPE"

La base de mesure "SARAPE" est en fonctionnement depuis 1995. Les mesures sont effectuées en plusieurs périodes, échantillonnées tout au long de l'année en fonction de la saison et du rythme de croissance de la végétation. L'un des principaux objectifs de mesure, est d'alimenter en données sources le logiciel "SIROS"[5] qui permet de simuler un radar SAR. A partir de ces mesures on peut extraire les valeurs de rétrodiffusion moyenne du fouillis (σ_0), les informations de texture radar. Par ailleurs, le caractère polarimétrique de l'instrument de mesure permet d'obtenir la matrice de rétrodiffusion complète du fouillis.

Les scènes mesurées

L'installation définitive de SARAPE sur un camion aura lieu en 1996. En 1995, la station de mesure a fonctionné sur une plate forme mobile provisoire. Pour cette raison, la campagne comprend seulement des scènes mesurées à l'intérieur du site du CELAR:

- herbe (courte , haute , sèche , grasse ...)
- bitume (route , parking ...)
- chemins (terre , graviers ...)
- labours (diverses orientations de sillons)
- végétation basse (colza)

Les sorties typiques produites par SARAPE

La planche 2 illustre un résultat typique d'une mesure SARAPE. Elle comprend en haut une photographie de la scène et en bas les quatre images radar correspondant aux quatre canaux de polarisation HH, HV, VH et VV.

L'exploitation des images radar

La planche 3 montre le résultat du traitement par le logiciel graphique SHOWM d'une scène composée de deux types de labour. Comme on peut le constater sur les photographies, il s'agit sur la partie gauche d'un cas de labour parallèle et sur la partie droite perpendiculaire à la visée radar. Le logiciel a calculé, pour chacune des deux zones de la scène, les paramètres de rétrodiffusion:

- σ_0 dans chaque polarisation.
- ρ, β et ε les paramètres inter corrélation polarimétriques.

Mesure des hauteurs par interférométrie

La planche 4 montre les résultats d'une expérience d'interférométrie réalisée par SARAPE. La scène est composée (photographie du haut de la planche 4) de trois rangées de calottes sphériques placées à différentes hauteurs grâce à des supports transparents aux ondes électromagnétiques. Deux mesures SARAPE ont été effectuées à deux sites d'observation différents: 43° et 45°. Le graphique en bas de la planche 4 montre que l'exploitation des phases des pixels des deux images permet de reconstituer la hauteur des sphères au dessus du sol.

Mesure sur un calibrateur réparti

La planche 5 montre une expérience permettant de confronter les résultats d'une mesure SARAPE de σ_0 avec celui d'un calcul d'un calibrateur réparti [10] développé par le CERT.

Le calibrateur réparti est constitué de plusieurs calottes sphériques assemblées sur un grillage métallique assimilable à un plan conducteur. Ce type de calibrateur a fait l'objet d'une étude qui permet de prévoir la valeur du σ_0 en fonction des paramètres géométriques.

La planche 5 montre la photographie de la scène, une image radar haute résolution et un histogramme des σ_0 obtenus sur des images radar dont la résolution est de l'ordre de 8 m^2. Ces images faiblement résolues ont été produites en utilisant une bande de fréquence et un déplacement d'antenne de faible amplitude. L'histogramme a été produit à l'aide de 180 mesures de σ_0.

L'écart entre les mesures expérimentales avec SARAPE et le calcul théorique est inférieur à 2 dB.

Analyse des paramètres de rétrodiffusion

influence du site sur la rétrodiffusion moyenne (figure N°4)

Le graphique ci-dessous présente les résultats de diverses expériences de mesure de σ_0:

- Des mesures SARAPE
- Des mesures diverses réalisées ou collectées par le CESBIO [9] dans le cadre d'un travail effectué en collaboration avec le CELAR.
- Les résultats d'un modèle statistique donné par ULABY et DOBSON [11]

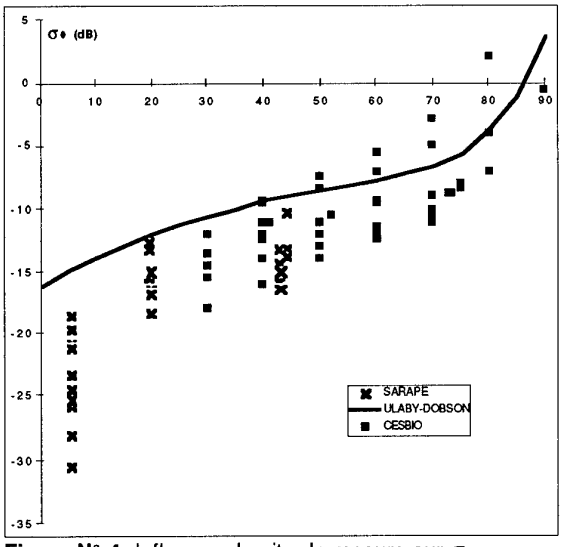


Figure N° 4: Influence du site de mesure sur σ_0

Coefficient d'intercorrélation polarimétrique de trois types de terrain

L'intercorrélation polarimétrique est définie par les 3 valeurs complexes suivantes:

- ρ : Corrélation entre les canaux HH et VV*
- β : Corrélation entre les canaux HH et HV*
- ε : Corrélation entre les canaux HV et VV*

Le tableau ci-dessous donne la moyenne et le rapport moyenne sur écart type pour les modules de ρ, β et ε. Ces valeurs sont comparables à celles publiées par FREEMAN [12] et elles sont directement utilisables dans le module de corrélation polarimétrique du logiciel SOURSAR de création d'images sources radar [8].

	Herbe (36 mesures)		Labour (21 mesures)		Terre (9 mesures)	
	M	E/M	M	E/M	M	E/M
$\lvert\rho\rvert$	0,45	0,15	0,68	0,10	0,58	0,11
$\lvert\beta\rvert$	0,07	0,56	0,08	0,54	0,06	0,71
$\lvert\varepsilon\rvert$	0,08	0,60	0,08	0,51	0,09	0,69

M : Valeur moyenne
E/M : Rapport écart type sur moyenne

De la même manière à partir des mesures SARAPE il est possible d'extraire les statistiques des phases de ρ, β et ε . Le tableau suivant donne la moyenne et l'écart type pour la phase de ρ. Les phases de β et ε sont proches d'une distribution uniforme.

	Herbe (36 mesures)		Labour (21 mesures)		Terre (9 mesures)	
	M	E	M	E	M	E
Phase de ρ	22°	12°	9°	10°	32°	15°

M : Valeur moyenne
E : Écart type

4 - CONCLUSIONS

L'instrumentation SARAPE actuellement opérationnelle au CELAR permet d'effectuer toutes les mesures nécessaires à la caractérisation du fouillis radar: mesure radar et mesure des paramètres d'environnement. Elle est associée à des outils logiciels permettant de traiter et d'archiver les données ainsi mesurées. Ces données sont disponibles sous une forme informatique. La base comprend à ce jour plus de 4500 valeurs de σ_0 et une partie de ces données possèdent des informations sur les caractéristiques polarimétriques.

SARAPE est aujourd'hui utilisable dans la bande 2-18 GHz. D'autres besoins importants commencent à apparaître. Ils devraient conduire à doter le système d'une capacité de mesure simultanée pour différents types de capteurs. Des actions prenant en compte le domaine des ondes millimétriques et de l'infrarouge sont actuellement en cours d'études.

SARAPE est une instrumenation complémentaire des moyens de mesure aéroportés. Cette station présente deux avantages importants: Le coût de mise en oeuvre est faible et la mesure permet d'accèder facilement aux paramètres d'environnement de la scène.

BIBLIOGRAPHIE

[1] S. EL ASSAD et J. SAILLARD, Reconstitution d'images radar à partir d'un moyen de mesures à visée latérale, IRESTE laboratoire S2HF NANTES

[2] O. CROP Description des conditions d'implantation d'une instrumentation de mesure pour imagerie un visée latérale, CELAR BRUZ

[3] C. BREUILLE et R. LOUVEL High resolution measurement of radar backscattering matrix of targets, Second international workshop on radar polarimetry, IRESTE, NANTES

[4] K. SARABANDI, F.T. ULABY, M.A. TASSOUDJI, Calibration of polarimetric radar systems with good polarisation isolation, IEEE Trans. Geosci. Remote Sensing, vol 28, pp. 70-75, January 1990.

[5] M. BUSSON, J.P HALBERT, J.C. MAGNAN, J.C. MOTET Un outil de simulation pour spécifier les systèmes spatiaux d'observation et de reconnaissance, Colloque AGARD de Colorado Springs du 17 Octobre 1989.

[6] N.MARCE : LEADER version 3.0. Document descriptif et justificatif des algorithmes. Manuel d'utilisation .CELAR BRUZ

[7] R.GUILLEREY : SARAPE, présentation des fichiers au format SEND. CELAR BRUZ

[8] C. COCHIN , J.C. MOTET , E. POTTIER
SOURSAR, génération de scènes polarimétriques opérationnelles pour un simulateur de radar SAR.
3ième JIPR - p 667 - Nantes Mars 1995

[9] H. LAUR, T. LE TOAN, A LOPEZ, E. MOUGIN, R. TOUZI
Rapport final de la convention CESR/CELAR N° 12.41.87 - Avril 1989

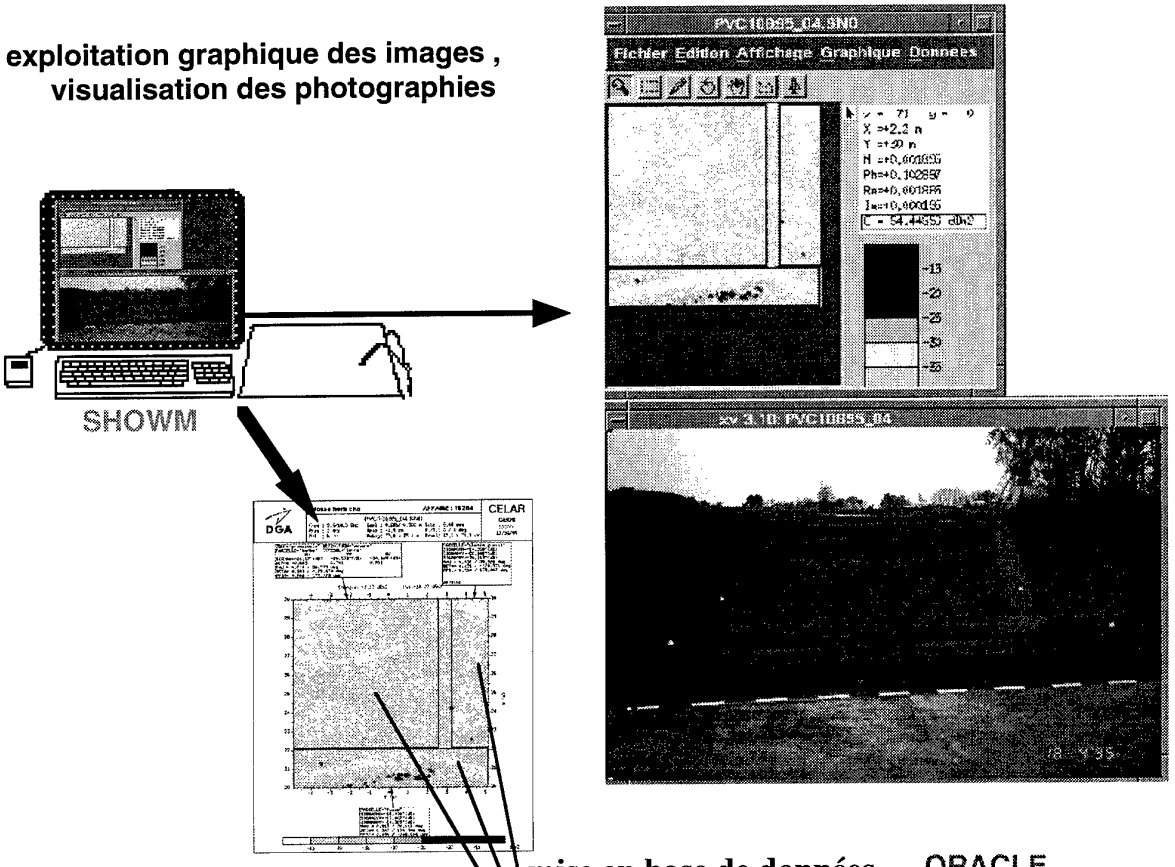
[10] H. MOREAU
Étude du couplage électromagnétique entre calottes sphériques métalliques. Application à la calibration de radar imageurs.
Thèse de l'université Paul Sabatier - Toulouse

[11] ULABY, DOBSON, Handbook of Radar Scattering statistics for terrain

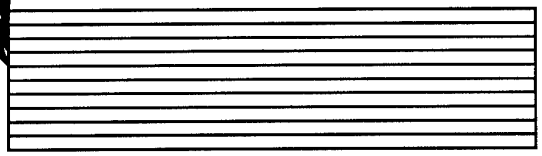
[12] A. FREEMAN, Classification of multi-frequency, multi-temporal polarimetric SAR images of natural vegetation.
2ième JIPR - Page 272 - Nantes Septembre 1992

PLANCHE 1
outils d'exploitation des images

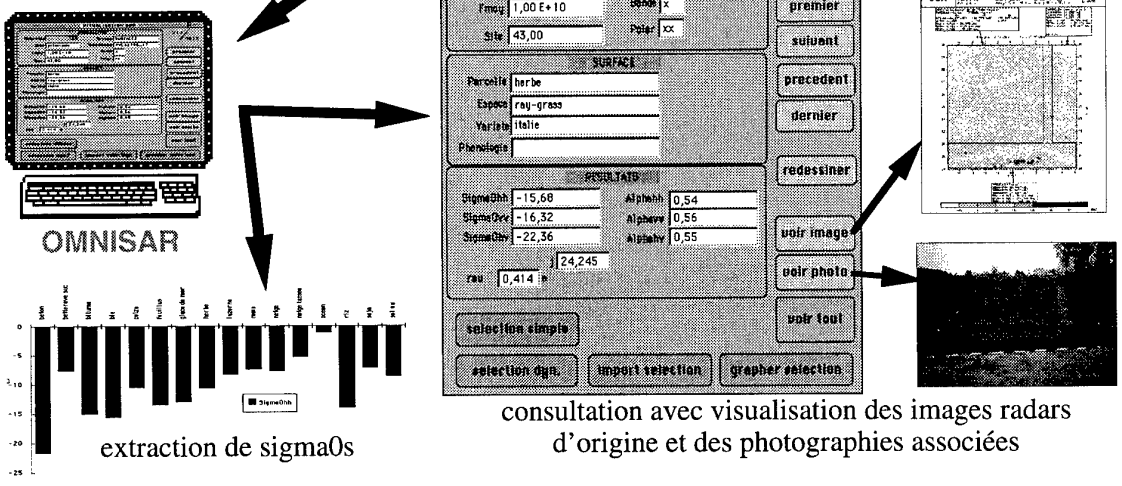
exploitation graphique des images ,
visualisation des photographies



mise en base de données **ORACLE**



interrogation de la base
de données



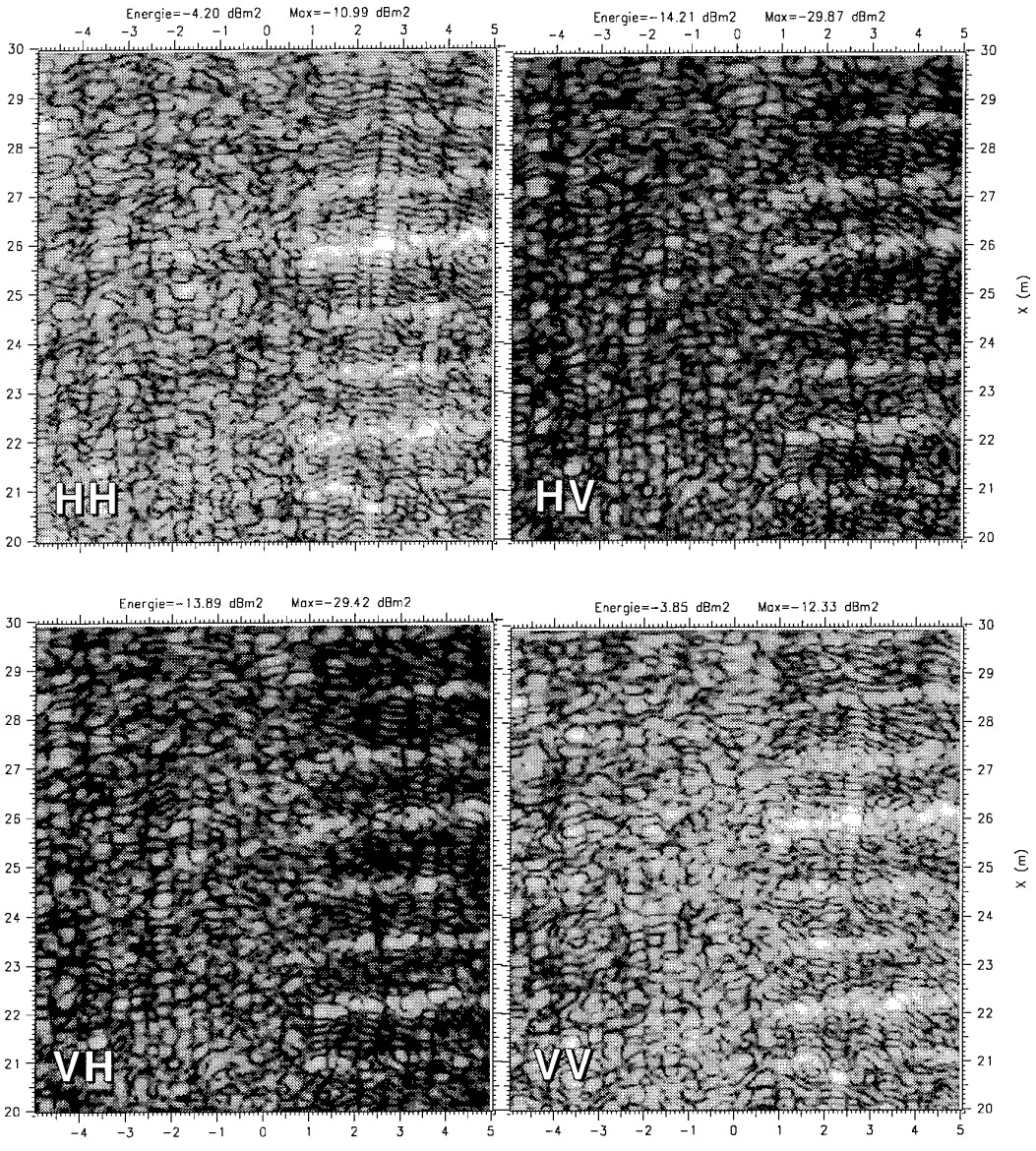
extraction de sigma0s

consultation avec visualisation des images radars
d'origine et des photographies associées

PLANCHE 2
imagerie d'une zone de labours



photographie de la scène



images radar

PLANCHE 3
imagerie d'une zone de labours



photographie de la scène

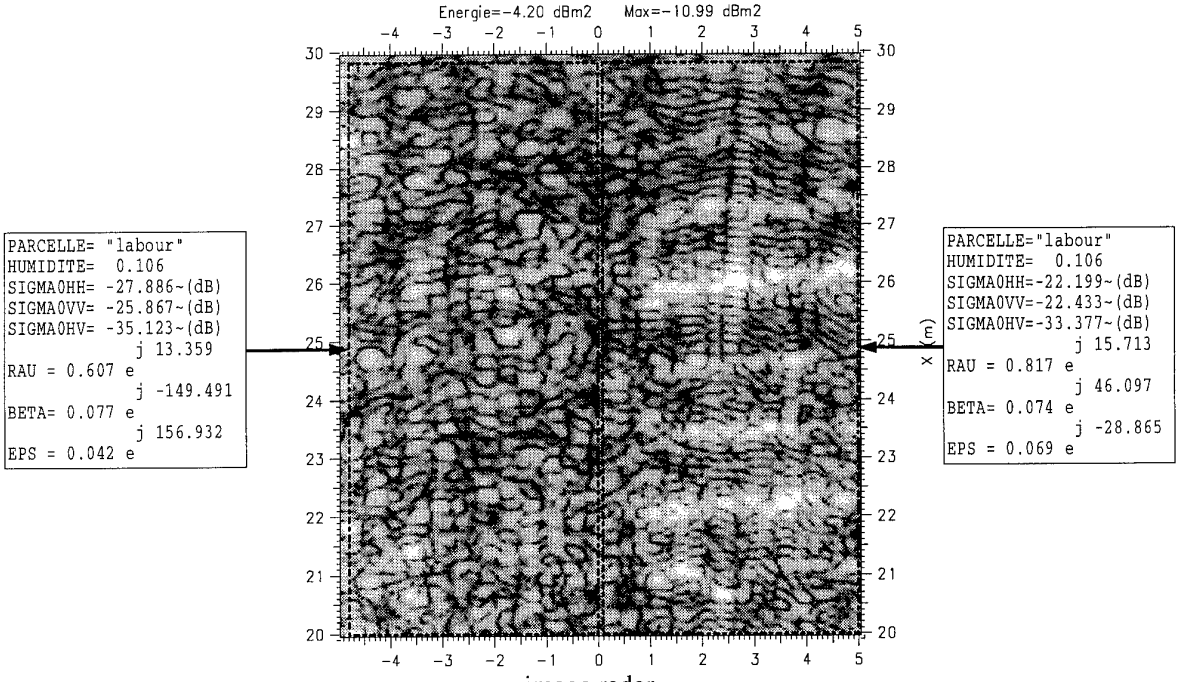
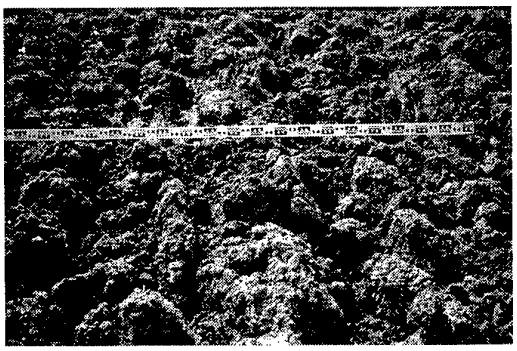


image radar

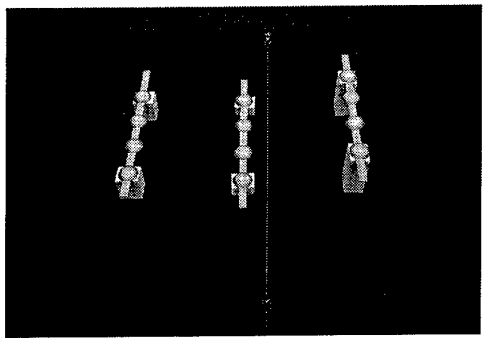


labour parallèle à la visée



labour perpendiculaire à la visée

PLANCHE 4
expérience d'interférométrie



photographie de la scène

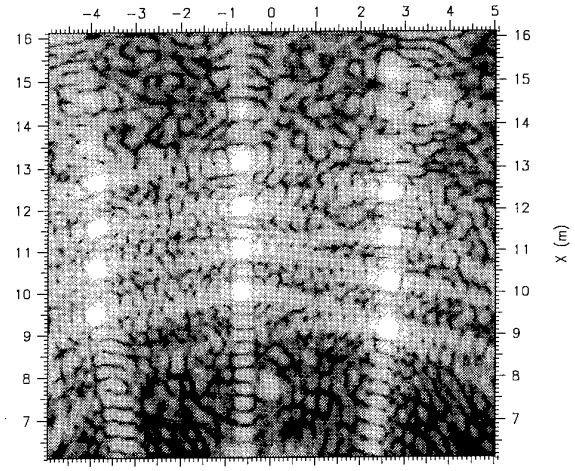


image radar

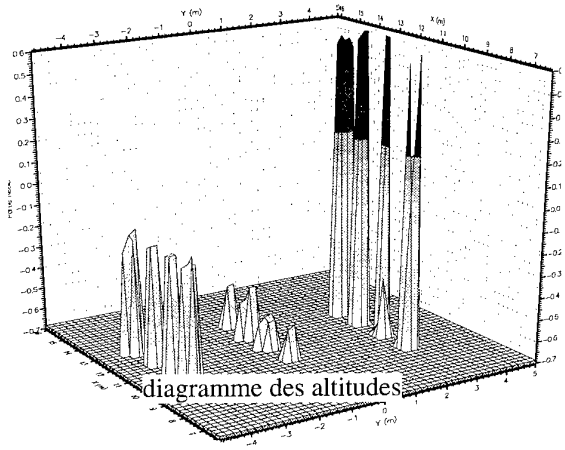
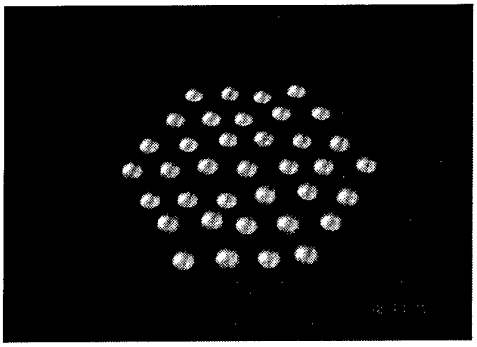


diagramme des altitudes

PLANCHE 5
mesure sur calibrateur réparti



photographie de la scène

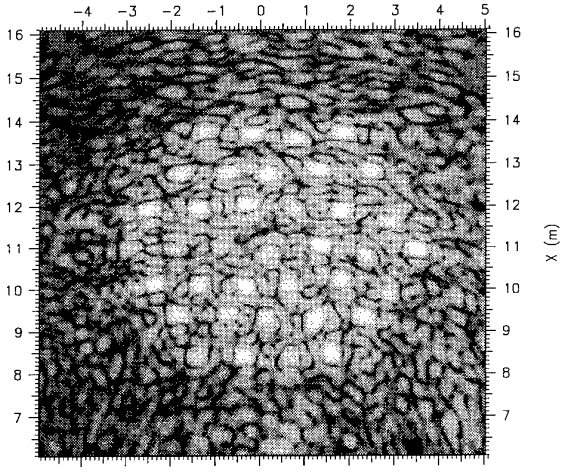
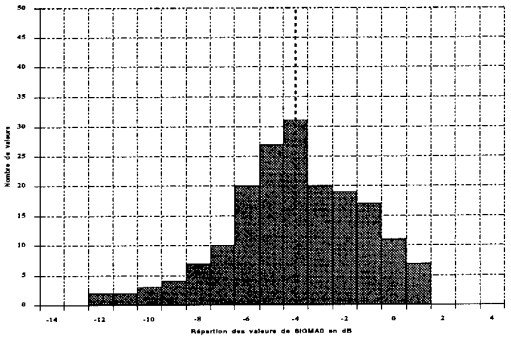


image radar



histogramme

Remote Sensing by Active and Passive Optical Techniques

C. R. Philbrick, M. D. O'Brien, D. B. Lysak, T. D. Stevens and F. Balsiger
Electrical Engineering Department, and
Remote Sensing Department, Applied Research Laboratory
Penn State University
State College, PA 16804 USA

1. SUMMARY

Active optical remote sensing techniques based on lidar have been mostly limited to ground based and aircraft applications because of rather restrictive $1/R^2$ dependence of signal, low power efficiency (~1%) of flash lamp pumped lasers and large physical size required by the power system. The advances in diode pumped lasers and other electro-optical instruments make it possible to consider new possibilities. Opportunities for future aircraft, RPV and satellite platform measurements over large spatial scales have prompted us to evaluate current capabilities for ground-based lidar measurements with a view toward future compact lidars. We have examined the Raman and DIAL lidar techniques to determine their capability for various remote sensing applications. The Raman techniques, even with inherent sensitivity disadvantages of small scattering cross-sections, have proven to be highly successful for ground-based remote sensing because they easily provide quantitative results, and these techniques will still be valuable from aircraft altitudes. Examples of ground-based lidar measurements of structure properties and minor species profiles have been used to provide confidence in the model calculations of expected performance from flight platforms. Use of acousto-optical tunable filters (AOTF) with lidar techniques provide special capabilities for measurement of Raman and fluorescent lidar returns. Applications of SPR-DIAL techniques for high altitude platforms would be useful for detection of chemical pollution in the atmosphere and waterways. Capabilities of current ground-based lidar systems for profiling meteorological properties, such as: density, temperature, water vapor, refractivity, chemical species, aerosol extinction and particle characteristics, provide the base from which we may consider future applications. Our recent reports have described the capability of lidars to measure species and temperature profiles, this paper focuses on the ability of lidar to measure and quantify aerosol properties.

2. INTRODUCTION

Use of LIDAR sensor systems for atmospheric characterization is becoming an established technology. Ground-based lidar using Raman, DIAL (DIfferential Absorption Lidar) and Doppler lidar techniques have demonstrated capabilities for remote sensing of all of the meteorological properties measured by standard radiosonde balloons. The flight of the NASA LITE experiment on the US space shuttle in 1994 (Ref 1) and the Russian-French lidar on the Mir Station in 1996 (Ref 2) provided major steps toward the eventual use of lidar at satellite altitudes. There are now many laboratories involved in lidar development, and advances are rapidly being made in materials and hardware. At ARL/PSU, we have developed and demonstrated several ground-based systems capable of measuring water vapor profiles, aerosol distribution and temperature profiles in the troposphere, stratosphere and mesosphere (Ref 3, 4, 5, 6). We are currently developing an operational prototype of a lidar system, Lidar Atmospheric Profile Sensor (LAPS), which will provide real time profiles of RF refractivity from measured profiles of temperature and water vapor (Ref 7). Use of Raman scattering and DIAL measurements to detect the spectra of several atmospheric pollution species and various chemical species has been examined (Ref 8). A multi-wavelength lidar instrument has demonstrated measurements of atmospheric pollutants and gas species using the SPR-DIAL technique with about sixty of the CO_2 laser lines in the 10μ region of the spectrum (Ref 9).

Capabilities of lidar, combined with passive optical remote sensing techniques provide the basis for future sensor systems to measure atmospheric pollution episodes and for real time monitoring of industrial processes. Detection and profiling of chemical species in the atmosphere have been a long term goal of lidar techniques, however, manufacture of reliable and stable lasers, development of measurement techniques, and preparation of specialized filters have delayed progress toward the autonomous instrument which would be useful in many applications of lidar. The LAPS instrument program is one of the current activities which has the goal of developing an operational lidar system.

Paper presented at the AGARD SPP Symposium on "Remote Sensing: A Valuable Source of Information", held in Toulouse, France from 22-25 April 1996, and published in CP-582.

3. LIDAR BACKGROUND

During the past twenty years, researchers at several laboratories have demonstrated that lidar is capable of measuring many different properties of the atmosphere. The first Raman measurements of atmospheric properties with lidar were carried out in the late 1960's by Leonard (Ref 10) and Cooney (Ref 11). Two years later Melfi, et al. (Ref 12) and Cooney (Ref 13, 14) showed that it was possible to detect atmospheric water vapor using the Raman lidar technique. Inaba and Kobayasi (Ref 15) suggested several species that could be measured using vibrational Raman techniques. While these early tests showed that it was possible to measure the water vapor with limited range and accuracy, recent investigations have demonstrated significant improvements. Particularly, the investigations of Vaughan et al. (Ref 16), Melfi et al. (Ref 17), Whiteman et al. (Ref 18) and Philbrick et al. (Ref 19) have demonstrated rather convincingly that the Raman technique has an excellent capability for making accurate water vapor measurements, both during the day and night. A most useful review of the Raman and DIAL lidar techniques applied to water vapor measurement has been prepared by Grant (Ref 20).

4. LIDAR INSTRUMENT DEVELOPMENT

4.1 LAMP Lidar

The LAMP (Lidar Atmospheric Measurements Program) lidar, which was developed at PSU in 1990, has been used during the past several years for atmospheric measurements using Raman vibrational and rotational scattering (Ref 19). Two detector systems have been prepared for the instrument. Initial measurements used a high altitude detector designed to cover the altitude range from 2 to 80 km. During the past 4 years, measurements have been focused on the troposphere with emphasis from the surface to 5 km, and several campaigns have been carried out which demonstrate the performance of Raman lidar techniques compared with standard rawinsonde balloon measurements. Investigations have emphasized water vapor and molecular nitrogen profiles which are determined from the 1st Stokes vibrational Raman transitions from laser wavelengths of 532 nm, 355 nm and 266 nm. Profiles of the N_2 vibrational Raman scatter provide true extinction measurements in the lower atmosphere, provided that the optical depth is sufficiently small that multiple scattering can be neglected. Water vapor profiles are obtained from the ratio of signals measured at the following wavelength pairs: 660/607, 407/387 and 295/284. The fact that the profile is determined from a ratio removes most of the errors in the profiles and makes the technique preferred for many applications. Temperature structure has been measured using rotational Raman scattering in the anti-Stokes region between 526 and 532 nm. Table I summarizes the LAMP lidar characteristics. Several reports (for example, Ref 3, 4, 7, 19) have shown examples of water vapor and temperature profiles obtained using the LAMP lidar. Measurements have proven the capability of Raman lidar to obtain profiles of atmospheric structure properties and water vapor in the lower atmosphere during night conditions. By comparing the molecular profiles of the N_2 Raman and rotational Raman with the neutral atmosphere density gradient, which can be determined from the temperature profile, the optical extinction profile can be obtained. Daytime measurement capabilities have been demonstrated using the "solar blind" region from the 266 nm fourth harmonic of Nd:YAG. Daytime measurements of water vapor are determined using the Raman scatter ratio of 295/284 ultraviolet wavelengths. A correction must be applied to this ratio because of the absorption due to tropospheric ozone. The ratio of the O_2 to N_2 vibrational Raman measurements (277/284) on the slope of the Hartley band of ozone provides a DIAL measurement of the ozone profile in the lower atmosphere, from the surface to altitudes between 2 and 3 km.

Table 1. LAMP Lidar characteristics

Transmitter	Continuum NY-82 -- 20 Hz 5X Beam Expander	400 mj @ 532 nm 80 mj @ 266 nm
Receiver	41 cm Diameter Telescope	Fiber optic transfer
Detector	Eight PMT channels Photon Counting	528 and 530 nm -- Rotational Raman Temperature 660 and 607 nm -- Vibrational Raman Water Vapor 295 and 285 nm -- Daytime Water Vapor 277 and 284 nm -- Raman/DIAL Ozone
Data System	DSP 100 MHZ	75 meter range bins

4.2 LAPS Lidar

The Lidar Atmospheric Profile Sensor (LAPS) instrument is currently undergoing tests of its automated operation to determine its performance under a wide range of meteorological conditions. The instrument measures water vapor profiles based on the vibrational Raman scattering and the temperature profile based on the rotational Raman scattering. These measurements are used to calculate real-time profiles of RF refractivity. Profiles are obtained each 5 minutes with a vertical resolution of 75 meters (efforts are underway to improve the resolution to 15 meters) from the surface to 7 km. The prototype instrument, which includes several sub-systems to automate and monitor the operation, has been designed to provide the real-time measurements of profiles. The instrument includes an X-band radar which detects aircraft as they approach the beam and automatically protects a 6 degree cone angle around the beam. The instrument includes calibration, performance testing and built-in-tests to check many functions. The larger laser and telescope of the LAPS instrument yield a signal approximately ten times that of the LAMP instrument.

4.3. Acousto-optical Tunable Filter

Remote sensing systems with high spectral resolution in the visible and near-IR regions can provide the sensitivity and discrimination against interferents and monitor the hazardous chemical vapors in the background environment. Due to their compact size, ruggedness and relatively narrow spectral resolution, Acousto-Optical Tunable Filters (AOTF) provide the opportunity to select the Raman line of a species of interest. Current state of the art AOTF instruments are capable of a spectral resolution of better than 10 cm^{-1} at selected visible/uv wavelengths (0.3 nm at 500 nm). The advantages in using the Raman signals to measure chemical species which exist in significant concentrations in the lower atmosphere are clear. In particular, Raman lidar measurements provide the accurate concentration measurements of species from the signal ratio when simple corrections are made for the optical extinction, which is also derived from the measurements. Examples of the many species that could be measured, using a filter with a narrow band-pass, typically 0.2 nm, are indicated in summary references (Ref 15, 21). An AOTF (Ref 22, 23) is being combined with Raman lidar to enable the operator to select any vibrational Raman line to measure the spectral signatures of atmospheric chemical species. Only the first Stokes vibrational states are considered, since the simple molecules have large vibrational energy state separation and the anti-Stokes lines are not normally populated at atmospheric temperatures. The Raman signal of a molecular species measured as a ratio to the N_2 signal provides a profile which is proportional to the species vapor concentration. The N_2 fraction of the atmospheric profile is known, thus the atmospheric profile of most any species, present in sufficient concentration, can be obtained. The error caused by the extinction differences between the backscatter wavelengths is small (few percent) and can be corrected using the results from multiple wavelengths.

Developments of new AOTF materials should result in extension of the useful range of the filters to the infrared spectrum where environmental applications are envisioned. The advantages in light through-put, frequency agility, ruggedness and resolution have drawn attention to this technique. The rich spectral signatures of many chemical species of interest in the infrared region from 8 to 12 μm make it very attractive to develop the AOTF capability in this wavelength region. The AOTF instrument would provide the flexible wavelength selection needed in a lidar detector to accompany development of OPO lasers in this same wavelength region.

Table 2. LAPS lidar characteristics.

Transmitter	Continuum 9030 -- 30 Hz 5X Beam Expander	600 mj @ 532 nm 130 mj @ 266 nm
Receiver	61 cm Diameter Telescope	Fiber optic transfer
Detector	Seven PMT channels Photon Counting	528 and 530 nm -- Rotational Raman Temperature 660 and 607 nm -- Vibrational Raman Water Vapor 295 and 284 nm -- Daytime Water Vapor 277 and 284 nm -- Raman/DIAL Ozone
Data System	DSP 100 MHZ	75 meter range bins (to be upgraded to 15 m)
Safety Radar	Marine R-70 X-Band	protects 6° cone angle around beam

5. ARL/PSU LIDAR MEASUREMENTS

5.1 Water Vapor Profiles

The measurement of water vapor has been described earlier (Section 4.1). Present capabilities for water vapor measurement under nighttime conditions are thoroughly demonstrated and are fully capable of describing the water vapor profiles in the troposphere. One example of a water vapor profile is shown in Figure 1 and many others are in the earlier referenced material. This example shows one of the ways in which the lidar excels over the radiosonde balloon. In the dry layer at 2 km the balloon sensor goes into cut-off and shows a late recovery at about 3.5 km on assent. Measurements of water vapor during the daytime have been demonstrated using 4th harmonic of the Nd:YAG laser.

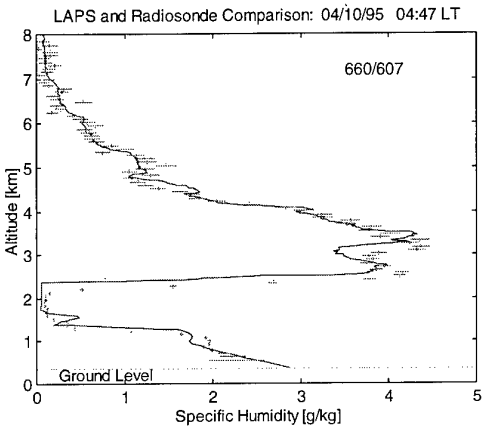


Figure 1. An example water vapor profile measured by the LAPS is lidar compared with a radiosonde balloon on 10 April 1995 at State College PA. The lidar profile shows the standard deviation of the water vapor based upon the count statistics during a 30 minute integration with no smoothing of the 75 meter measurement intervals.

At UV wavelengths, the measurements of N_2 and H_2O are contaminated by the difference in absorption at the two wavelengths due to O_3 in the lower troposphere (see Ref 24). However, we have found that an adequate correction can be obtained from the use of the measured Raman signals of the N_2 and O_2 compared to the known mixing ratio. This analysis results in a measured profile of O_3 from the surface to 3 km.

5.2 Temperature Profiles

Rotational Raman measurements of temperature were reported by Cooney (Ref 25). A double grating monochrometer was used by Arshinov, et al. (Ref 26) to measure the rotational Raman spectrum. Hauchecorne, et al. (Ref 27) and Nedeljkovic, et al. (Ref 28) demonstrated the capability to measure the temperature using narrow-band filter technology in the upper troposphere and lower stratosphere. We have demonstrated the first useful profiles of tropospheric temperature from the rotational Raman line envelop using narrow band filters (Ref 19). The molecular species of the atmosphere, principally N_2, O_2 contribute to the envelope of lines on either side of fundamental laser wavelength (Ref 29). In our case, several lines of the rotational states are measured at the 530 and 528 nm filter bands. The envelop shape is determined by the population distribution of the rotational states for the temperature of the gas in the scattering volume, which is illuminated by the doubled Nd:YAG laser at 532 nm. Rotational lines are available on both the long and short wavelength sides of the fundamental exciting frequency, however we have chosen the short wavelength side of the distribution to eliminate any excitation from fluorescent transitions. The ratio between the intensities of the two filter bands is used to directly determine the temperature. An example of the profile derived from the LAPS rotational Raman results is shown in Figure 2. Comparison of a rawinsonde balloon measurement is used to develop a calibration curve based upon an empherical fit. It is possible that the measured temperature profile and a ground based measurement of the surface pressure be used to generate the profiles of the structure properties, density, pressure and temperature based on calculations using the hydrostatic equation and the ideal gas law. The calculated profile of density can be used to obtain the N_2 profile which will then place an absolute density on the water vapor from the ratio of vibrational Raman signals.

5.3 Optical Extinction Profiles

Optical extinction can be measured based on the molecular profile from the Raman scatter, or from a proper description of the particle size distribution using a bistatic lidar measurement. During the CASE (Coastal Aerosol Scattering Experiment) and the EOPACE (Electro-Optical Propagation Assessment in the Coastal Environment) field campaigns, the optical extinction and the scattering phase function have been

measured as a way of describing the particle size distribution and density. The first tests of this measurement technique were conducted in September 1995, during the CASE I on Wallops Island Virginia. This location was chosen for its humid and misty, coastal/marine environment, and for an unobstructed 3.28 km horizontal path over a salt marsh. The Penn State LAMP monostatic lidar was operated on both a horizontal path and pointed vertical. When operating on a horizontal path the LAMP lidar collected rotational Raman temperature profiles, vibrational Raman water vapor profiles, vibrational Raman nitrogen profiles, and bistatic aerosol/molecular scattering profiles at the fundamental laser wavelength of 532 nm.

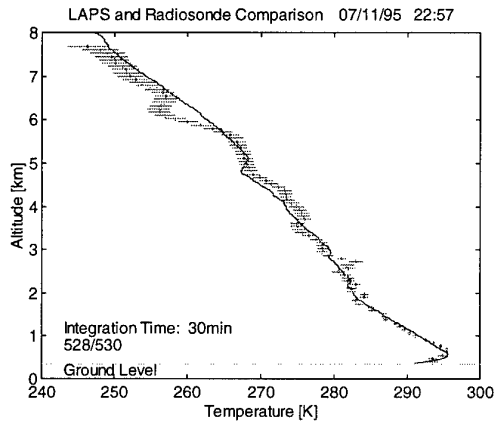


Figure 2. The temperature profile derived from the rotational Raman scatter measured by the LAPS lidar is shown with the standard deviation of the count statistics indicated.

5.31 *Optical extinction from molecular profiles determined from Raman scatter*

The previous techniques for measurements of the optical properties were limited to point sensors or path integrated transmission. The ARL/PSU lidar has been used to demonstrate that true extinction profiles can be obtained from molecular Raman scatter profiles. The approach provides a way of remotely sensing a vertical profile from the surface to 10 km. By measuring these meteorological properties together with the aerosol backscatter and extinction, the growth or dissipation phases of clouds and aerosols can be investigated.

Measurements of optical extinction are based upon gradients in the molecular profiles, using the N_2 vibrational Raman or a band of the rotational Raman lines. We have been able to demonstrate that the true optical extinction can be determined for clouds and aerosol layers. In order to calculate extinction due to aerosols from Raman lidar measurements, we take the following steps:

1) Apply the range correction to the backscatter Raman shifted signal.

2) Correct for the attenuation due to molecular scattering on the way out at 532 nm and the Raman shifted wavelength on the way back to the lidar. The molecular attenuation coefficients for 532 and 607 nm are known from the cross-sections for N_2 and O_2,

$$\alpha_{532} = 0.013\ km^{-1} \qquad \alpha_{607} = 0.007\ km^{-1}$$

3) For the 528 and 530 nm data we can calculate the extinction value for 532 nm by fitting a least square algorithm to the following function,

$$I(R) = I_0\, e^{-2\alpha_{532} R}$$

to the part of the profile that is not affected by the telescope form factor, beyond about 1 km.

4) The extinction of the 607 nm is obtained by fitting the function,

$$I(R) = I_0\, e^{-(\alpha_{532}+\alpha_{607}) R}$$

to the 607 nm profile. This algorithms can only determine the sum of the two extinctions but with the extinction determined in 3), the extinction for 607 nm can be calculated.

Figure 3 shows the variation of the optical scattering which can occur over the 3.3 km path as the air mass changes during a three hour period. The initial clear air measured at 22:34 on 17 September 1995 is replaced by an aerosol scattering air mass which is observed over about half of the path to the target board at 23:39. The increased optical scattering air mass is fairly uniform over the region by 00:39 on 18 September.

5.32 *Optical extinction from bistatic lidar*

The spherical particle scattering phase function calculated from the Mie scattering theory has obvious features which indicate high sensitivity to angle scattering function, particularly in the angle range of 150 to 175 degrees. The question of whether the distribution of realistic sizes mixed together would

remove the uniqueness of the phase function dependence upon size has been investigated. Also, the question of whether the Mie theory should apply to atmospheric measurements requires an assumed spherical shape for the particles. The coastal region, with its high humidity, is the natural region to investigate the applicability of the bistatic lidar measurement. Calculations that have been performed by Stevens (Ref 30) show that the unique character of the phase function dependence upon size is maintained when the calculation includes a realistic distribution of particle sizes. It is possible to ignore the absolute signal measurement and the sensitivity dependence on angle by using the polarization ratio. The ratio of the scattered intensity parallel and perpendicular to the measurement plane is used. By rotating the polarization plane and recording the signal on each pixel, it is possible to obtain information on the phase function dependence upon particle size based upon this ratio. The success in making these measurements has been demonstrated in recent CASE tests.

The bistatic receiver uses a linear photodiode array to image the radiation scattered from a high power CW or pulsed laser system. By observing the angular scattering variation along a horizontal path, information contained in the scattering angle phase function can be obtained. A technique has been developed to estimate particle size and distribution widths (of spherical scatters) by comparing the scattered return of laser beams polarized parallel and perpendicular to the scattering plane. A polarizer is used on the receiver to measure the cross polarization for determining the amount of multiple scattering and non-sphericity of the particles in the scattering volume.

The bistatic lidar uses the ratio obtained from the image of the scattering of two perpendicular polarization components with respect to angle, one in the scattering plane and one perpendicular to it. The use of a ratio cancels the effects of many problems, including non-linearity across the field of view of the receiver and extinction differences due to different path lengths for each scattering angle. The bistatic receiver simultaneously measures the scattering by forming an image of the backscattered radiation over an angle range of 155° to 180°. The assumption of a uniform

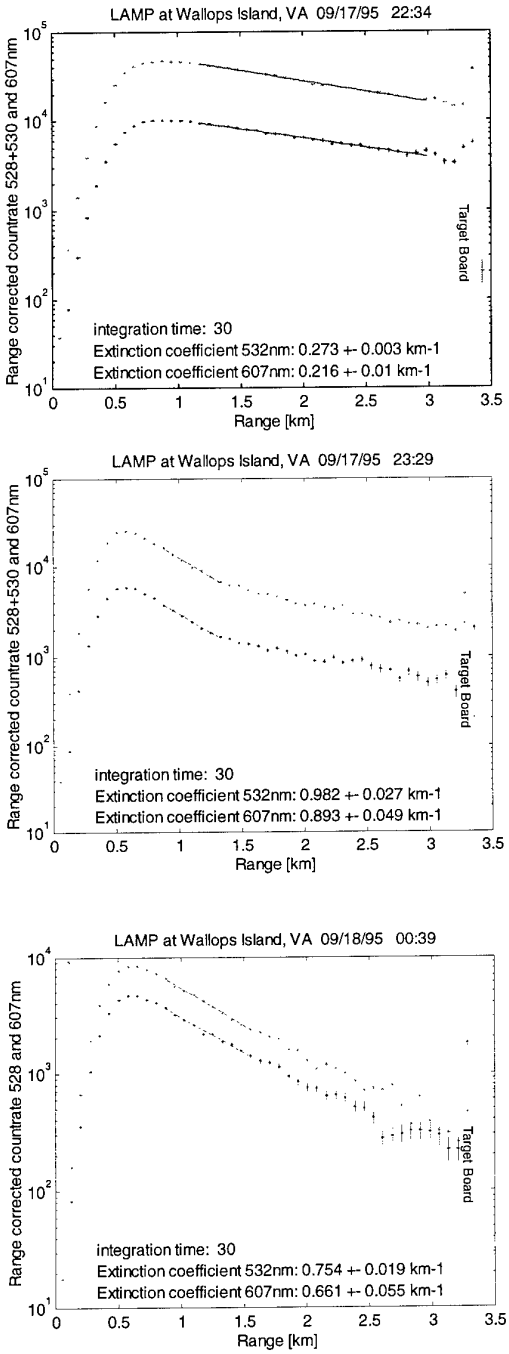


Figure 3. Optical extinction measured over a 3.3 km path at NASA Wallops Island VA on the night of 17/18 September 1995. An air mass change occurs near 23:29 when the aerosol scattering increases.

horizontal path is verified with horizontal extinction profiles from the monostatic Raman lidar during each data set. Although this polarization ratio contains the information needed to characterize the scatterers, a model is needed to relate this ratio to the desired parameters, such as extinction, and particle size distribution. The tri-modal aerosol distribution is indicated in Figure 4.

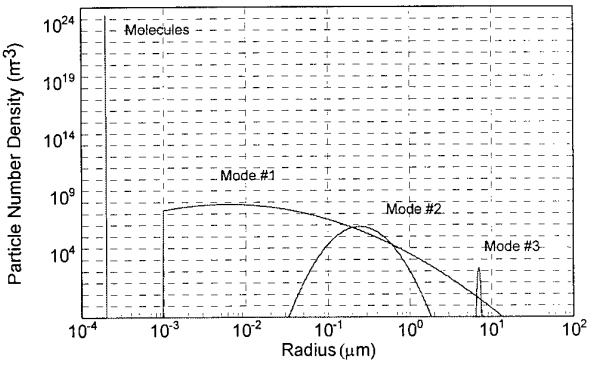


Figure 4. A plot of a trimodal lognormal particle size distribution using the bistatic lidar model. The addition of 2.54×10^{25} molecules per m^3 is shown at the "equivalent Mie radius" of 0.198 nm.

An example of this measurement technique is shown in Figure 5a, with the best fit of the model, and the corresponding trimodal distribution in Figure 5b. During this night, temperature decreased while the extinction increased until the wind brought a new air mass with different scatterers in the area. The second mode of the distribution was seen to narrow its distribution width as it grew from a radius of 0.166 μm to 0.237 μm. At the same time, the third mode increased from a radius of 6.46 μm to 8.91 μm. This data set is most striking because the model follows almost every contour in the data between 155° and 180° as seen in Figure 5a. Not only does the model fit the data almost perfectly, but the calculated extinction coefficients are also the same as those measured by the Raman lidar, see extinction data in Figure 5a. It is important to note that the extinction calculated from the Raman molecular extinction agrees very well with the extinction calculated for the size distribution determined from the bi-static lidar analysis, compare the values given in Figure 5a.

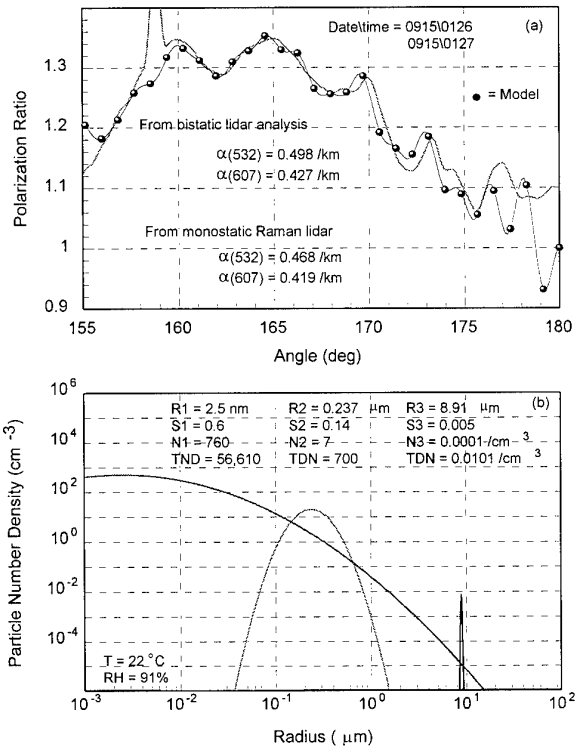


Figure 5. (a) A most convincing example of this bistatic lidar measurement technique, showing the best fit of the spherical model, using the distribution from (b), with the data. The radiation fog mean radius has grown from 6.46 μm to 8.91 μm in only 2.5 hours. Note the comparison of the extinction coefficients calculated from the Raman technique and from the particle size distribution which are listed in the figure.

6. SUMMARY

The application of laser remote sensing to measurement of the atmospheric structure properties has been demonstrated using the rotational Raman technique. Many chemical species in the atmosphere, whether these result from pollution sources, accidental release of hazardous vapor from industrial processes, or from the release of toxic chemicals in the battlefield environment, can be measured by vibrational Raman and DIAL techniques. The capabilities for active sensing using vibrational Raman scattering, DIAL (DIfferential Absorption Lidar), SPR-DIAL (Spectral Pattern Recognition-DIAL) have been demonstrated. Future developments with fluorescence and resonance Raman techniques should add to these capabilities. Lidar fluorescence techniques will have applications to measurements of contamination of surfaces, water columns and for

optical extinction by analyzing the molecular profiles from Raman scattering. Use of the scattering phase function with bi-static lidar measurements has shown that information on the particle size and distribution can be obtained. The progress in development of ground based techniques and the initial demonstrations of lidar on aircraft and satellites suggests that future lidars should make significant contributions to remote sensing from space based platforms, which will provide a global dimension to measurements and models.

7. ACKNOWLEDGMENTS

Special appreciation for the support of this work goes to J. Richter of NCCOSC NRaD, G. Schwemmer of NASA GSFC and SPAWAR PMW-185. The efforts of Savy Mathur, Paul Haris, Tom Petach, Bob Smith and Glen Pancoast have contributed much to the success of this project.

8. REFERENCES

1. M. P. McCormick, D. M. Winkler, E. V. Browell, J. A. Coakley, C. S. Gardner, R. M. Hoff, G. S. Kent, S. H. Melfi, R. T. Menzies, C. M. R. PLatt, D. A. Randall, and J. A. Reagan, "Scientific Investigations Planned for the Lidar In-Space Technology Experiment," Bulletin of the Am. Meteoro. Soc. 74, 205-214, 1993.

2. G. F. Tulinov, "Investigation of the Atmosphere by Ground-based and Space-based Lidars on the `PRIDODA' Module," Report of Institute of Applied Geophysics, Moscow, Feb. 1995.

3. C. R. Philbrick, D. B. Lysak, T. D. Stevens, P. A. T. Haris and Y.-C. Rau, "Atmospheric Measurements Using the LAMP Lidar during the LADIMAS Campaign," 16th International Laser Radar Conference, NASA Publication 3158, 651-654, 1992.

4. C. R. Philbrick, D. B. Lysak, T. D. Stevens, P. A. T. Haris and Y.-C. Rau, "Lidar measurements of middle and lower atmosphere properties during the LADIMAS campaign," Proceedings of the 11th ESA Symposium, ESA-SP-355, 223-228, 1994.

5. T. D. Stevens, P. A. Haris, Y.-C. Rau and C. R. Philbrick, "Latitudinal Lidar Mapping of Stratospheric Particle Layers," Advances in Space Research, 13, (9)193-198, 1994.

6. P. A. T. Haris, T. D. Stevens, S. Maruvada and C. R. Philbrick," Latitude Variation of Middle Atmosphere Temperatures," Advances in Space Research, 13, (9)83-87, 1994.

7. C. R. Philbrick and D. W. Blood, "Lidar Measurements of Refractive Propagation Effects," in Propagation Assessment in Coastal Environments, NATO-AGARD CP 567, Paper 3, 13 pg, 1995.

8. C. R. Philbrick and P. H. Kurtz, "Remote Sensing of Atmospheric Chemical Contamination," Proceedings of Institute of Nuclear Materials Management, 36th Annual Meeting, July 1995.

9. D. A. Leonard, R. N. Bigelow, I. K. Munson, N. T. Nomiyama and H. E. Sweeney, "Remote Chemical Agent Detection and Measurement System Using a CO2 Laser Radar," 1984 IRIS Speciality Group on Active Systems, Johns Hopkins University, Laurel MD, November 1984.

10. D. A. Leonard, "Observation of Raman scattering from the atmosphere using a pulsed nitrogen ultraviolet laser," Nature 216, 142-143, 1967.

11. J. A. Cooney, "Measurements on the Raman component of laser atmospheric backscatter," Appl. Phys. Lett., 12, 40-42, 1968.

12. S. H. Melfi., J. D. Lawrence Jr. and M. P. McCormick, "Observation of Raman scattering by water vapor in the atmosphere," Appl. Phys.Lett., 15, 295-297, 1969.

13. J. A. Cooney, "Remote measurement of atmospheric water vapor profiles using the Raman component of laser backscatter," J. Appl. Meteor., 9, 182-184, 1970.

14. J. A. Cooney, "Comparisons of water vapor profiles obtained by rawinsonde and laser backscatter," J. Appl. Meteor., 10, 301, 1971.

15. H. Inaba and T. Kobayasi, "Laser-Raman Radar," Opto-electronics, 4, 101-123, 1972.

16. G. Vaughan, D. P. Wareing, L. Thomas and V. Mitev, "Humidity measurements in the free troposphere using Raman backscatter," Q. J. R. Meteorol. Soc., 114, 1471-1484, 1988.

17. S. H. Melfi, D. Whiteman and R. Ferrare, "Observations of atmospheric fronts using Raman lidar moisture measurements," J. Appl. Meteo., 28, 789-806, 1989.

18. D. N. Whiteman, S. H. Melfi and R. A. Ferrare, "Raman lidar system for the measurement of water vapor and aerosols in the Earth's atmosphere," Appl. Optics 31, 3068-3082, 1992.

19. C. R. Philbrick, "Raman Lidar Measurements of Atmospheric Properties," in Atmospheric Propagation and Remote Sensing III, SPIE Volume 2222, 922-931, 1994.

20. W. B. Grant, "Differential absorption and Raman lidar for water vapor profile measurements: a review," Optical Engineering, 30, 40-48, 1991.

21. R. M. Measures, Laser Remote Sensing, Krieger Publishing, Malabar FL, 1992.

22. V. I. Pustovoit and V. E. Pozhar, "Collinear Diffraction of Light by Sound Waves in Crystals: Devices, Applications and New Ideas," Photonics and Optoelectronics 2, 53-69, 1994.

23. F. L. Vizen, Yu. K. Kallinikov and V. I. Pustovoit, "Quartz Acousto-Optic Filter," Scien. Inst. and Exp. Tech. 6, 170, 1979.

24. D. Renaut, J. C. Pourny and R. Capitini, "Daytime Raman-Lidar Measurements of Water Vapor," Opt. Lett. 5, 233, 1980.

25. J. Cooney, "Measurement of atmospheric temperature profiles using the Raman component of laser backscatter," J. Appl. Meteo. 11, 108-112, 1972.

26. Yu. F. Arshinov, S. M. Bobrovnikov, V. E. Zuev and V. M. Mitev, "Atmospheric temperature measurements using a rotational Raman lidar," Appl. Optics 22, 2984-2990, 1983.

27. A. Hauchecorne, M. L. Chanin, P. Keckhut, and D. Nedeljkovic, "Lidar monitoring of temperature in the middle and lower atmosphere," Appl. Phys. B55, 29-34, 1992.

28. D. Nedeljkovic, A. Hauchecorne and M. L. Chanin, "Rotational Raman lidar to measure the temperature from the ground to 30 km," IEEE Trans. Geos. Remote Sens. 31, 90-101, 1993.

29. P. A. T Haris, "Pure Rotational Raman Lidar for Temperature Measurements in the Lower Troposphere," Dissertation, Department of Electrical Engineering, Penn State University, 1995.

30. T. D. Stevens, "Bistatic Lidar Measurements of Lower Tropospheric Aerosols," PhD Dissertation, Department of Electrical Engineering, Penn State University, 1996.

Diffusion des micro-ondes par des hydrométéores non sphériques et inhomogènes. Simulations, observations et restitutions durant l'expérience TOGA-COARE.

E. OBLIGIS et C. KLAPISZ

Centre d'Etude des Environnements Terrestre et Planétaires (CETP)
92131 Issy-Les-Moulineaux
France

1. INTRODUCTION

L'inversion d'un modèle de transfert radiatif permet de restituer des paramètres géophysiques à partir des données radiométriques aéroportées obtenues le 6 Février pendant l'expérience TOGA-COARE. Ces paramètres sont les précipitations liquides, les précipitations en phase glace et l'eau liquide nuageuse. Il est malheureusement impossible de valider ces résultats. Toutefois la comparaison avec les taux précipitants obtenus à partir des données radar montre que les grandes structures de précipitations sont convenablement restituées.

Une amélioration du modèle de transfert radiatif consisterait à introduire une couche de mélange ou couche de fonte, pour prendre en compte la présence d'une bande brillante en situation stratiforme. Les théories de Mie, de la Tmatrice et des sphères enrobées sont utilisées pour obtenir des profils verticaux des coefficients de diffusion et d'extinction dans cette partie de l'atmosphère. Les variations de ces coefficients dans cette zone sont d'autant plus importantes que la fréquence est basse et la densité de la neige faible. Il semble donc necessaire de prendre en compte cette couche mixte dans le modèle de transfert radiatif.

2. RESTITUTION DE PARAMÈTRES GÉOPHYSIQUES PAR RADIOMÉTRIE HYPERFRÉQUENCE AÉROPORTÉE ET SATELLITAIRE

2.1 L'expérience TOGA-COARE et l'instrumentation utilisée

2.1.1 *L'expérience TOGA-COARE*

L'expérience TOGA-COARE s'est déroulée de Novembre 92 à Février 93 dans le Pacifique Ouest [1]. L'objectif de cette campagne était d'obtenir de nouvelles informations pour décrire et comprendre les processus responsables du fort couplage entre l'océan et l'atmosphère dans le "Warmpool" du Pacifique Ouest.

2.1.2 *Le cas du 6 Février 1993*

Le 6 Février 1993 s'est développé au sud-est des îles Salomon un système convectif intense lié au cyclone Oliver. Ce système a été très bien échantillonné par les 4 avions présents pendant la campagne TOGA-COARE. Une grande quantité de données sont donc disponibles pour ce cas d'étude.

2.2 Description de l'instrumentation

Toutes les données aéroportées utilisées proviennent du DC8, un avion de la NASA, volant à une altitude d'environ 10 km, et à une vitesse de 220 m/s, et qui a échantillonné pendant plusieurs heures une bande convective du cyclone tropical Oliver. A bord de cet avion se trouvaient les 2 radiomètres micro-ondes (l'Airborne Multichannel Microwave Radiometer et l'Advanced Microwave Moisture Sounder) ainsi qu'un radar (l'Airborne Rain Mapping Radar) (Fig. 1).

L'AMMR est un radiomètre micro-ondes fonctionnant aux 3 fréquences 18.7, 21 et 37 GHz et visant au nadir, alors que l'AMMS scanne perpendiculairement à la trajectoire de l'avion et réalise des mesures à 92, 183.3+/-2, 183.3+/-5 et 183.3+/-9 GHz. On peut donc obtenir au nadir de l'avion des points pour lesquels on a simultanément des températures de brillance à 18.7, 21, 37 et 92 GHz.

L'ARMAR a été développé par le département JPL de la NASA dans le cadre du projet de radar spatial TRMM (Tropical Rain Measuring Mission) qui devrait voler en 1997. Ce radar opère à 13.8 GHz et mesure des réflectivités dans un plan perpendiculaire à l'avion. Seules les données obtenues au nadir seront ici utilisées.

Nous avons également utilisé pour cette étude les données des dropsondes larguées par le DC8 tous les quarts d'heure, et qui permettent de connaître les profils de température, de pression et d'humidité sous l'avion.

2.3 Le modèle de transfert radiatif et son inversion

2.3.1 *Le modèle de transfert radiatif*

Le modèle de transfert radiatif utilisé est le modèle de Kummerow [2] qui permet de calculer les températures de brillance émergeant au sommet de l'atmosphère. Il est basé sur l'approximation d'Eddington et utilise une atmosphère stratifiée en couches homogènes horizontalement [3].

Pour un angle de visée θ, l'équation décrivant le transfert radiatif d'une onde monochromatique de fréquence υ à travers une couche d'épaisseur dz est :

$$\frac{dT_{B\upsilon}(z,\theta)}{dz} = -k_{e\upsilon}(z)\Big[T_{B\upsilon}(z,\theta) - \big(1-a_\upsilon(z)\big)T(z) - \frac{a_\upsilon(z)}{4\pi}\int_{-1}^{1} P(\theta,\theta')T_{B\upsilon}(z,\theta')d(\cos\theta')\Big]$$

$T_{B\upsilon}(z,\theta)$ est la température de brillance à l'altitude z, se propageant vers le haut. T, k_e, et a_υ sont respectivement la température, le coefficient d'extinction et l'albédo de simple diffusion à l'altitude z. $P(\theta,\theta')$ est la matrice de phase reliant les paramètres de Stockes de l'onde incidente aux paramètres de Stockes de l'onde diffusée. Les températures de brillance et les éléments de la matrice de phase sont développés en série de Legendre et en fonctions de Legendre associées. Les sections efficaces de diffusion et d'extinction sont obtenues par la théorie de Mie (les particules sont supposées sphériques) puis intégrées sur une distribution de tailles des gouttes appropriée pour obtenir les coefficients volumiques en m^{-1}. Pour cette étude, l'atmosphère est stratifiée en 2 couches séparées par l'isotherme 0°C, ce qui permet de séparer les composants "glace" des composants "eau liquide". La surface océanique de température supposée connue T_S est considérée lambertienne avec une émissivité ε_υ.

Paper presented at the AGARD SPP Symposium on "Remote Sensing: A Valuable Source of Information", held in Toulouse, France from 22-25 April 1996, and published in CP-582.

2.3.2 *L'inversion*

L'inversion du modèle est faite avec une méthode quasi-Newton. Un certain nombre de paramètres peuvent être fixés de manière à obtenir un meilleur ajustement. Ces paramètres sont les poids des différentes fréquences, les valeurs initiales des inconnues, ainsi que leurs domaines de variation. Un choix approprié de ces paramètres permet d'accélérer l'inversion et de minimiser les risques de divergence.

2.4 Les paramètres restitués

Les paramètres que l'on cherche à restituer sont les précipitations en phase liquide en mm/h, les précipitations en phase glace en mm/h, ainsi que l'eau liquide nuageuse (g/m^3).

Les données aéroportées utilisées dans cette étude ont été obtenues le 6 février de nuit pendant une période approximative de 10 minutes. La trajectoire de l'avion d'environ 130 km de long a permis d'échantillonner de l'air clair, des zones convectives et une zone stratiforme. La figure 2 représente la trajectoire de l'avion sur l'image infrarouge correspondante.

2.4.1 *A partir des données radar*

Les sections verticales de réflectivités observées par l'ARMAR montrent d'abord une région d'air clair, puis une région convective avec des coeurs de fortes réflectivités puis une zone stratiforme où apparaît nettement la bande brillante.

Les taux de pluie correspondant sont obtenus à partir de l'algorithme KZS [4] qui corrige les réflectivités de l'atténuation en utilisant l'écho de surface. Ces taux de pluie sont ensuite moyennés dans la phase liquide entre le sol et 4.7 km.

2.4.2 *A partir des données radiométriques*

La combinaison des températures de brillance mesurées par l'AMMR et par l'AMMS permet d'avoir 4 températures de brillance pour chaque point. L'inversion devant être surdéterminée, on peut donc restituer 3 paramètres avec ces 4 températures. La figure 3 représente les 4 températures de brillance mesurées par les radiomètres pendant la même trajectoire, avec des faibles températures de brillance basses fréquences dans la zone d'air clair, des températures à 92 GHz très faibles dans les cellules convectives dues à une forte diffusion par les cristaux de glace, et enfin des températures à peu près constantes dans la zone stratiforme.

L'atmosphère utilisée pour l'inversion est stratifiée en 2 couches homogènes. Certains paramètres sont fixés : les paramètres de surface (vent et température) sont obtenus à partir de données de bouées dérivantes et les profils de températures et d'humidité à partir des données des dropsondes larguées par le DC8.

Les 3 paramètres restitués sont les précipitations "glace", les précipitations "liquide" et l'eau liquide nuageuse. Les précipitations "liquide" peuvent être directement comparées aux taux pluies restitués à partir des mesures de réflectivités radar (figure 4). Les taux de pluie restitués à partir des données radiométriques sont très bien corrélés à ceux obtenus à partir des données radar. Les principales structures (air clair, stratiforme, et convectif) sont bien restituées. Les pics de taux de pluie très élevés sont un peu sous-estimés par notre méthode, sans doute à cause de la résolution au sol, plus mauvaise pour les données radiométriques.

3. DIFFUSION PAR DES HYDROMÉTÉORES NON SPHÉRIQUES ET INHOMOGÈNES

La deuxième partie de notre étude a consisté à simuler radiométriquement la bande brillante. Les radaristes appellent bande brillante la zone de fortes réflectivités observées juste en dessous de l'isotherme 0°C dans le cas de situations stratiformes, où les particules de neige sont en train de fondre et s'enrobent d'eau. Nous avons étudié les variations en altitude des coefficients d'extinction et de diffusion juste en-dessous de l'isotherme 0°C. On considère le cas de particules fondantes sphériques traitées par la théorie de Mie ou ellipsoïdales traitées par la théorie de la Tmatrice. Les coefficients obtenus peuvent ensuite être introduits dans un modèle de transfert radiatif.

3.1 Description de la couche de mélange

La partie inférieure de l'atmosphère est maintenant divisée en 3 couches. Dans la couche supérieure, les particules sont des flocons de neige constitués d'un mélange d'air et de glace. Le processus de fonte commence lorsque les particules atteignent le sommet de la couche de fonte fixé par l'isotherme 0°C. A l'intérieur de la couche inférieure, toutes les particules sont fondues même les plus grosses.

3.1.1 *Vitesse de chute des particules*

On définit V_s, V_m, et V_r comme les vitesses de chute (m/s) comptées positivement des particules de neige (rayon a_s), des particules fondantes (rayon a_m) et des gouttes d'eau (rayon a). La vitesse de chute des particules fondantes peut s'écrire :

$$V_m = \frac{V_r}{\left\{Q\left(1-\frac{\rho_w}{\rho_s}\right)+\frac{\rho_w}{\rho_s}\right\}^{1/3}}\left[\frac{Cd(1,a)}{Cd(Q,a)}\right]^{1/2}$$

où Q est la proportion de neige fondue en masse, ρ_s et ρ_w sont les densités respectives de la neige et de l'eau liquide pour les particules fondantes et C_d est le coeficient de pénétration dans l'air, fonction linéaire du taux de fonte Q :

$$C_d(Q,a) = C_d(0,a) + \left[C_d(1,a) - C_d(0,a)\right]Q$$

Pour des particules sphériques de rayon a :

$$V_r = 9.65 - 10.3\exp(-12a)$$

$$\text{avec } A = \frac{C_d(1,a)}{C_d(0,a)} = \left[\frac{V_s}{Vr}\frac{a_s}{a}\right]^2$$

La valeur moyenne de la vitesse de chute des particules est obtenue de [5]:

$$V_s = \frac{2+16.1(\rho_s-0.05)}{(0.85+1.14\rho_s)}\left\{1-\exp\left[-a_s(9.76+1.36\,a_s\right]\right\}$$

et celle des particules fondantes est :

$$V_m = \frac{V_r}{\left\{Q\left(1-\frac{\rho_w}{\rho_s}\right)+\frac{\rho_w}{\rho_s}\right\}^{1/3}}\left[\frac{A}{1+(A-1)Q}\right]^{1/2}$$

V_r, V_s, V_m sont corrigées de la densité de l'air :

$$\left[\frac{\rho_a(z_b)}{\rho_a(z=z^*)}\right]^{0.4} = \exp\left(-\frac{0.4(z^*-z_b)}{9500}\right)$$

où z*(m) est l'altitude de l'isotherme 0°C et z_b(m) est la distance en-dessous de ce niveau comptée positivement.

Dans le cas de sphéroïdes aplatis avec leur axe de rotation parallèle à la direction verticale, on a d'après Oguchi and Hosoya :

$$a_h(Q=1) = \frac{a}{(1-0.9a)^{1/3}}$$

où a_h est le demi grand axe, et a est maintenant le rayon de la particule fondue sphérique de même volume. La loi utilisée pour l'applatissement des gouttes est la suivante :

$$\frac{c_h}{a_h} = 1 - a_m Q$$

On considère que les particules gelées sont des sphères (quand Q=0, $a_h=c_h=a_s$) et que la déformation est maximum quand la particule est entièrement fondue (Q=1).

3.1.2 *Constante diélectrique dans la couche de mélange*

La constante diélectrique de la neige sèche ε_s est obtenue en utilisant la fonction de Maxwell-Garnett pour des inclusions ellipsoïdales d'air orientées aléatoirement dans une matrice de glace [6]. Si on considère que la particule fondante est composée d'inclusions de neige dans une matrice d'eau, alors sa constante diélectrique moyenne ε_m est également décrite par la loi de mélange de Maxwell-Garnett appliquée une seconde fois aux inclusions de neige dans une matrice d'eau.

3.2 Diffusion par les particules fondantes

La théorie de Mie [7] est utilisée pour des particules de neige fondantes sphériques, et dans le cas de particules sphéroïdales on utilise l'"Extended Boundary Condition Method" [8] développée par Waterman. Cette méthode consiste en une approche matricielle du problème de la diffusion, dans laquelle le diffuseur est caractérisé par sa propre matrice de transition. Les champs électriques incidents et diffusés sont développés en séries de vecteurs d'harmoniques sphériques. Les éléments de la matrice ne dépendent pas de l'onde électromagnétique incidente mais seulement de la taille, de la forme, de l'indice de réfraction, de l'orientation de la particule diffusante, ainsi que de la fréquence étudiée.

Dans le repère sphérique choisi, les notations "i" et "s" désignent les ondes incidente et réfléchie (fig.5). Les vecteurs d'onde des ondes incidentes et réfléchies sont respectivement $\mathbf{k_i}$ et $\mathbf{k_s}$. Les vecteurs unitaires de polarisations de l'onde incidente sont $\mathbf{v_i}$ (contenu dans le plan d'incidence Oxz) et $\mathbf{h_i}$ ($\mathbf{k_i}=\mathbf{v_i}\times\mathbf{h_i}$). De même pour l'onde diffusée, on a $\mathbf{k_s}=\mathbf{v_s}\times\mathbf{h_s}$. L'origine du repère choisi est pris au centre de la particule diffusante d'axe de symétrie selon z.

Les coefficients volumiques de rétrodiffusion, de diffusion et d'extinction k_b, k_s et k_e sont fonction des sections efficaces correspondantes σ_b, σ_s et σ_e.

$$k_j(z) = \int_0^{a_{max}} \sigma_j(a_m) N_m(a_m) da_m \qquad j = b,s,e$$

où N_m est la distribution de tailles de particules pour les hydrométéores fondants, qui tient compte de l'aggrégation dépendant de z. Le coefficient d'absorption est $k_a=k_e-k_s$. Sous l'hypothèse que le taux précipitant est constant dans la couche de mélange et égal au taux de pluie en bas de cette couche :

$$\int_0^\infty \rho_s a_s^3 V_s N_s(a_s) da_s =$$

$$\int_0^\infty \rho_m(Q) a_m^3 V_m N_m(a_m) da_m = \int_0^\infty \rho_w a^3 V_r N_r(a) da$$

N_s, N_m et N_r sont respectivement les distributions de tailles des particules pour des hydrométéores gelés, fondants et liquides. De même les densités ρ_s, $\rho_m(Q)$ et ρ_w sont les densités des particules gelées, fondantes et liquides. Le facteur de réflectivité équivalent $Z_{eq}(z)$ est alors défini par :

$$Z_{eq}(z) = 10\,lg\left[\frac{\lambda^4}{\pi^5}\left|\frac{\varepsilon_w+2}{\varepsilon_w-1}\right|^2 k_b(z)\right]$$

3.3 Résultats et discussion

3.3.1 *Vitesse de chute et distribution de tailles des gouttes*

La figure 6 représente le rapport a_m/a pour 2 valeurs de densité de la neige ρ_s=0.05 g/cm^3 et ρ_s=0.2 g/cm^3 et pour différentes valeurs de z_b(m) comptées positivement à partir du sommet de la couche fondante. Les particules précipitantes de plus faible densité fondent plus rapidement que celles de densité plus élevée, et les petites particules plus vite que les grosses. La distribution utilisée pour calculer les différents coefficients volumiques dépend de z et tient compte de l'aggrégation :

$$N_m(z,a_m) = N_0(z) exp\left[-\Lambda_m(z,R)\, a_m\right]$$

3.3.2 *Particules sphériques*

La figure 7 présente les profils verticaux des coefficients de rétrodiffusion pour différentes densités et à la fréquence de 14 GHz. Quand la densité de la neige décroît, la bande brillante est plus proche de l'isotherme 0°C, plus étroite et plus marquée.

3.3.3 *Particules sphéroïdales*

Dans le cas de particules aplaties de forme sphéroïdales, le rapport c_h/a_h est fonction du taux de fonte et vérifie $a=(a_h^2 c_h)^{1/3}$. Les valeurs de k_s obtenues dans le cas de sphères et de sphéroïdes diffèrent d'environ 20 % dans la bande brillante. Mais l'effet de l'applatissement des gouttes en phase liquide est négligeable.

4. CONCLUSIONS

Un modèle de transfert radiatif approprié et convenablement inversé permet de restituer de manière satisfaisante les principaux contenus en eau dans l'atmosphère. Les résultats de cette méthode ont pu être comparés à ceux obtenus à partir de réflectivités radar. Cette méthode est également utilisée pour restituer des paramètres géophysiques à partir de données radiométriques satellitaires.

La présence de la bande brillante conduit à des fortes valeurs des coefficients de diffusion et d'extinction. Ce signal est d'autant plus fort que la densité de la neige et la fréquence sont faibles. L'étude en cours consiste donc à introduire ces résultats dans le modèle de transfert radiatif en utilisant une couche intermédiaire correspondant à la couche de fonte, et voir l'effet de l'introduction de la bande brillante sur les sorties du modèle.

REFERENCES

[1] P. J. Webster and R. Lukas, *Bull. Amer. Meteor. Soc.,* vol. **73**, No. 9, Sept. 1992
[2] C. Kummerow, *J. Geo. Res.*, vol. **98**, No. 2, pp 2757-2765, Feb. 1993
[3] R. Wu and J. A. Weinman, *J. Geo. Res.*, vol. **89**, No. 5, pp 7170-7178, Aug. 1984
[4] M. Marzoug and P. Amayenc, *J. Atmos. Ocean. Technol.,* **11**, pp 1480-1505, Dec. 1994
[5] Schols and Weinman, *Atmos. Res.*, **34**, pp 329-346, 1994
[6] C. F. Bohren and L. J. Battan, *J. Atmos. Sci.*, **39**, pp 2623-2628, 1982
[7] C. F. Bohren and D. R. Huffman, *John Wiley & Sons*, 1983
[8] P. C. Waterman, *Proc. I.E.E.E.*, **53**, pp 805-812, 1965

FIGURES

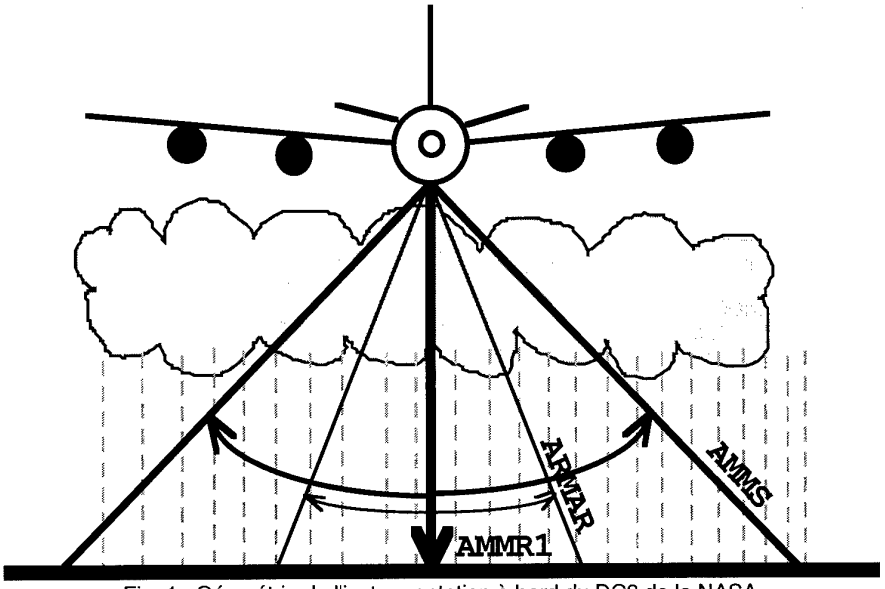


Fig. 1 : Géométrie de l'instrumentation à bord du DC8 de la NASA.

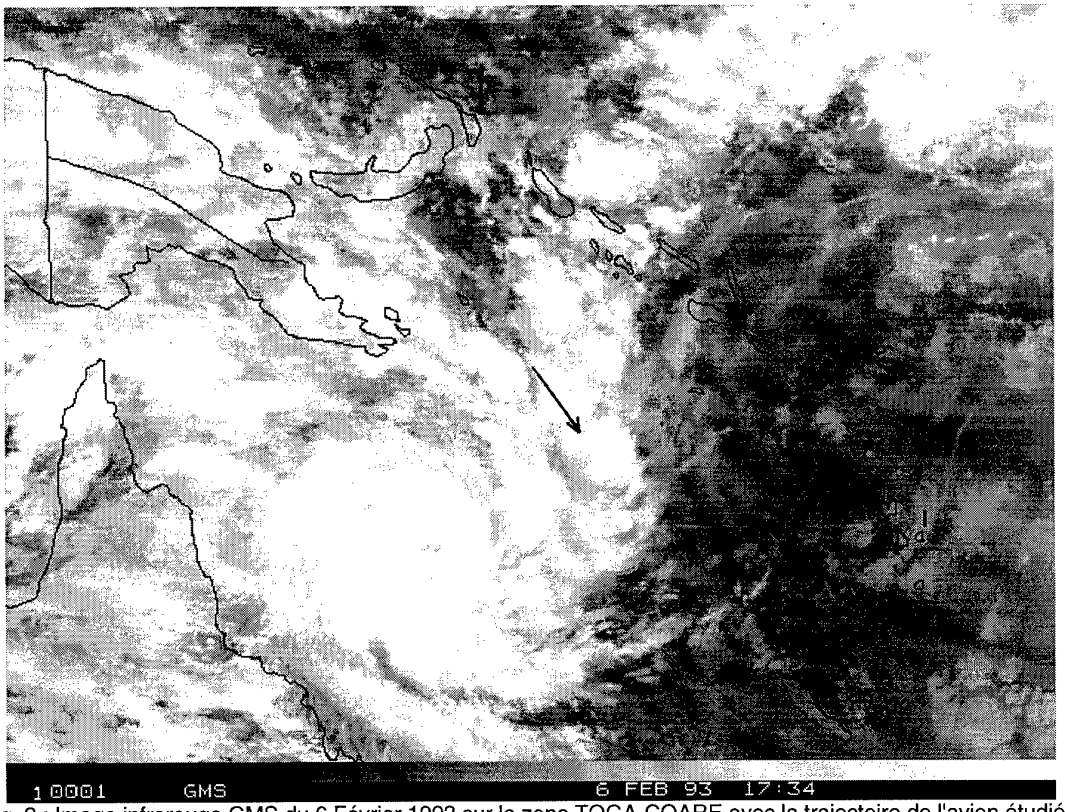


Fig. 2 : Image infrarouge GMS du 6 Février 1993 sur la zone TOGA-COARE avec la trajectoire de l'avion étudiée.

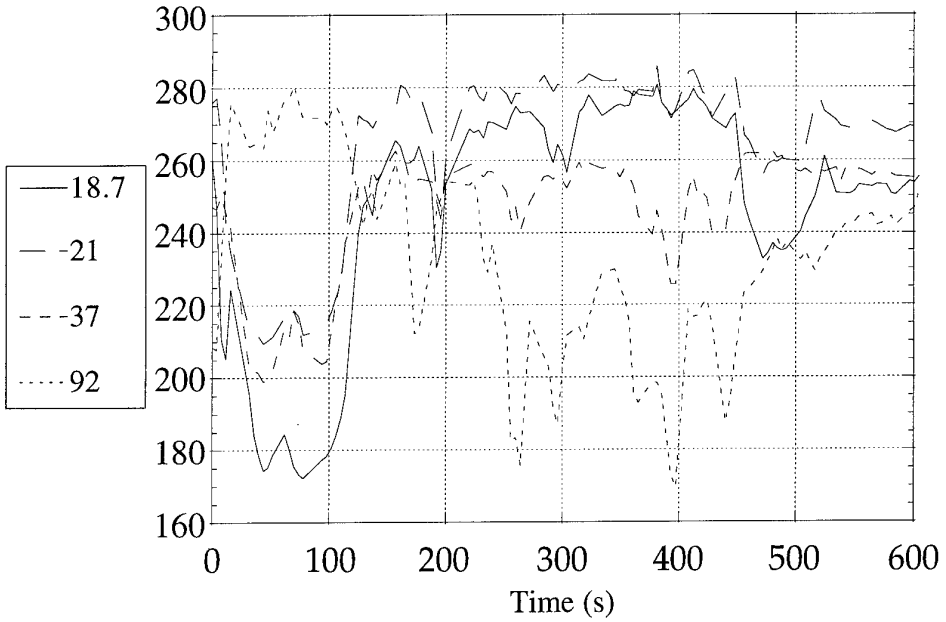


Fig. 3 : Températures de brillance en K à 18.7, 21, 37 et 92 GHz mesurées pendant le leg étudié par les radiomètres AMMR1 et AMMS.

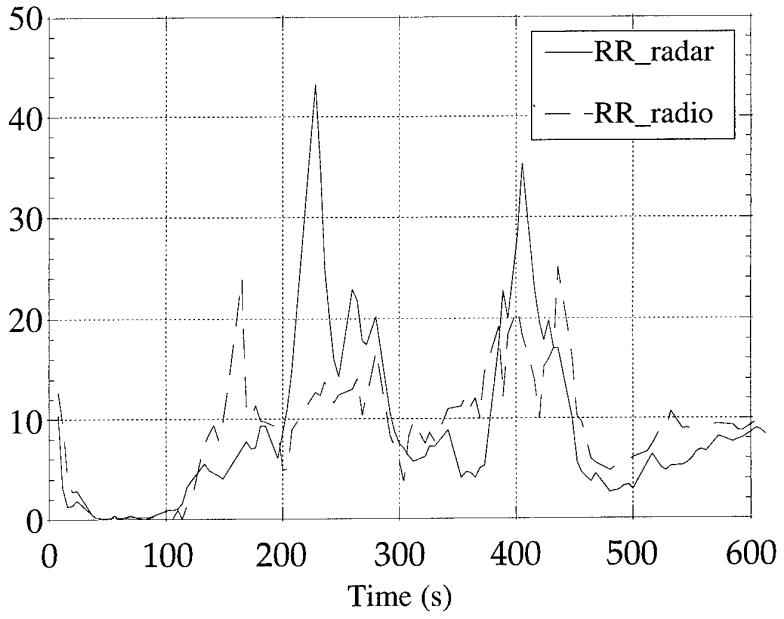


Fig. 4 : Taux de pluie en mm/h. Comparaison des restitutions obtenues à partir des données radiométriques et à partir des données radar.

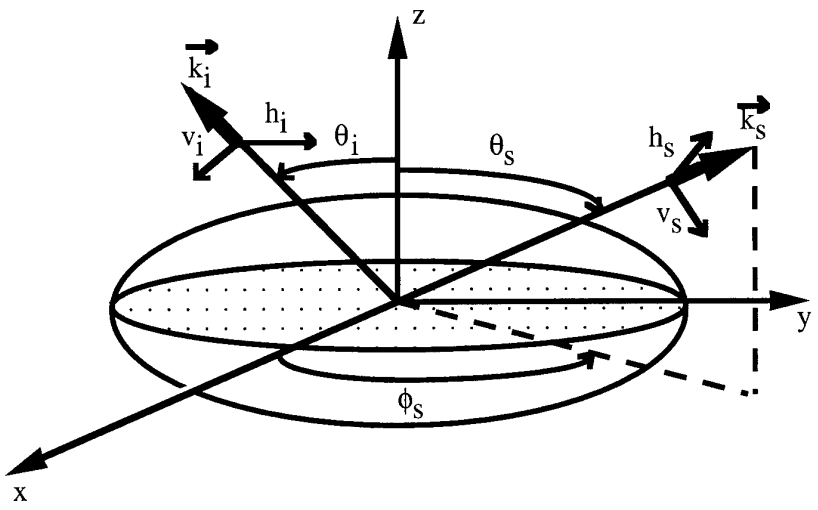
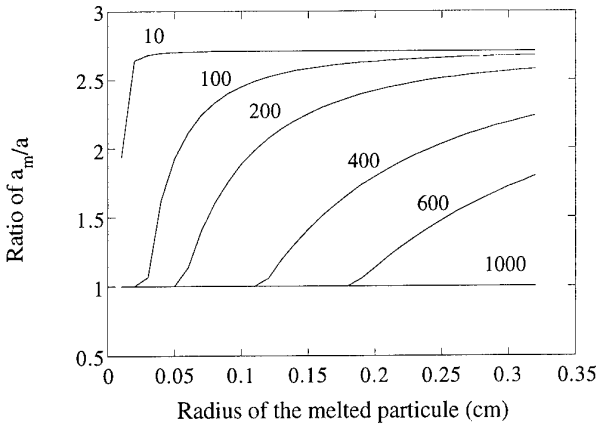


Fig. 5 : Coordonnées cartésiennes et sphériques pour une particule sphéroïdale.
$\mathbf{k_i}$ et $\mathbf{k_s}$ sont les vecteurs d'onde des ondes incidentes et diffusées respectivement.

a



b

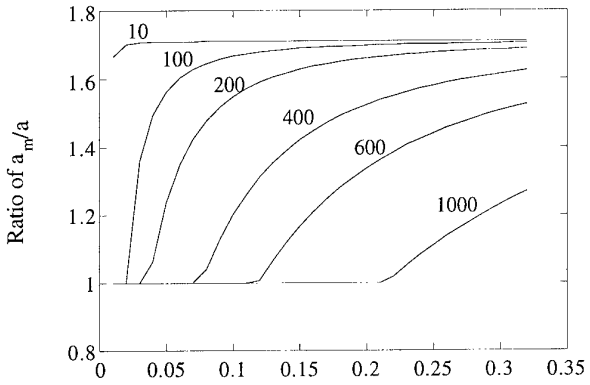


Fig. 6 : Rapport a_m/a pour des particules fondantes en fonction du rayon a (cm).
La distance au sommet de la couche de fonte z_b(m) est un paramètre.
(a) ρ_S=0.05 g/cm3 - (b) ρ_S=0.2 g/cm3.

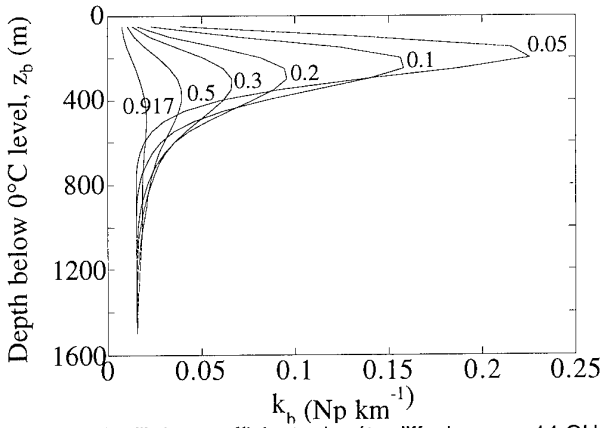


Fig. 7 : Profil des coefficients de rétrodiffusion pour 14 GHz.
Le paramètre est la densité de la neige ρ_S en g/cm^3.

Paper 9

E. Schweicher (BE)

1) Comment avez-vous fixé le seuil des X^2 lors de la minimalisation?

2) Je m'étonne que la résolution en température des radiomètres du DC8 soit meilleure aux basses fréquences (ou elle est de 1K). Comment expliquez-vous cela?

[1. How did you fix the threshold for X^2 during minimization?

2. I am surprised that the temperature resolution of the radiometers of the DC8 is better at low frequencies (where it is 1K). How do you explain this?]

Author's reply:

1) Subroutine de NUMERICAL-RECIPIES: dans l'espace des paramètres, les solutions sont representées par des "creux" et un calcul de pente dans les différentes directions permet de s'en rapprocher: lorsque toutes les pentes ont le même signe, on suppose qu'on a trouvé la solution.

2) Les 2 radiomètres AMMR1 and AMMS sont 2 radiomètres différents avec des géométries différentes - les résolutions en température de brillance ne sont donc pas comparables.

[1. NUMERICAL-RECIPIES subroutine: in the space for parameters the solutions are represented by the "dips" and a calculation of the slope in different directions allows you to come close: when all the slopes have the same sign then you assume you have found the solution.

2. The AMMR1 and AMMS are two different radiometers with different geometries. The brilliance temperature resolutions are therefore not comparable.]

Paper 9

J. Barckicke (FR)

Le modèle de transfert radiatif (présence de nuages) semble assez réaliste. Pourrait-il trouver une application dans les modèles de prévision numérique météteorologique (temps de calcul et disponsibilité des paramètres d'entrée, qui ne sont pas nécessairement des variables prognostiques des N.W.P modèles)?

[The radiation transfer model (presence of clouds) seems fairly realistic. Could it find an application in meteorological numerical forecasting models (calculation time and availability of input parameters, which are not necessarily prognostic variables of N.W.P. models?]

Author's reply:

Oui: ce modèle tourne assez vite (1 restitution par seconde). Il pourrait servir à initialiser certains modèles de prévision numérique.

[Yes: this is a fast model (1 restitution per second). It could be used to initialize certain numerical forecasting models].

An Overview of EOPACE (Electrooptical Propagation Assessment in Coastal Environments), Including In Situ and Remote Sensing Techniques

Kathleen M. Littfin
Douglas R. Jensen
Propagation Division
NCCOSC RDTE DIV 883
49170 Propagation Path
San Diego, CA 92152-7385 USA

SUMMARY

EOPACE is a five year multi-national effort to characterize electrooptical propagation in coastal environments. Existing EO performance prediction codes incorporate models that were developed for open ocean conditions. These codes are not applicable to coastal environments where small surface craft and sea-skimming missiles pose threats. Coastal environments may differ significantly from open ocean conditions, and need to be fully characterized. The objectives of EOPACE are threefold: (1) to investigate coastal aerosols by studying surf production, coastal air mass characterization, and near ocean surface transmission characteristics; (2) to develop mesoscale and data assimilation models; and (3) to evaluate EO systems performance by studying targets and backgrounds, polarization techniques, performance of forward looking infrared (FLIR) and infrared search and track (IRST) systems, and tactical decision aids.

During EOPACE, in situ and remote sensing techniques will be used to infer the impact of surf-generated aerosol, air mass parameterization required for propagation codes, near ocean surface infrared transmission properties, and IRST/FLIR systems performance in coastal environments. This paper gives an overview of the EOPACE effort and discusses the in situ and remote sensing techniques used.

1. INTRODUCTION

EOPACE is a measurement and analysis program to improve performance assessment for EO weapon and sensor systems operating in coastal environments. The majority of the EOPACE measurements, which started in January, 1996, are being conducted in the California coastal region (see Figure 1) by participants from the US and NATO countries. The climate of the region is similar to that of other regions of interest and represents an excellent example of a littoral environment. Follow-on measurement campaigns in high-latitude or tropical areas will be conducted as necessary.

The primary objectives of EOPACE are to:

- Quantify effects of coastal aerosols on EO propagation extinction
- Develop mesoscale and data assimilation models
- Evaluate EO systems performance.

One key feature of EOPACE is to conduct long term observations of one to two years, with several intensive operational periods lasting two to three weeks each. For the coastal aerosol study, measurements will support the investigation of surf production of aerosols, characterization of coastal air masses, and characterization of near ocean surface transmission. EOPACE will provide the database for model development and evaluation.

Paper presented at the AGARD SPP Symposium on "Remote Sensing: A Valuable Source of Information", held in Toulouse, France from 22-25 April 1996, and published in CP-582.

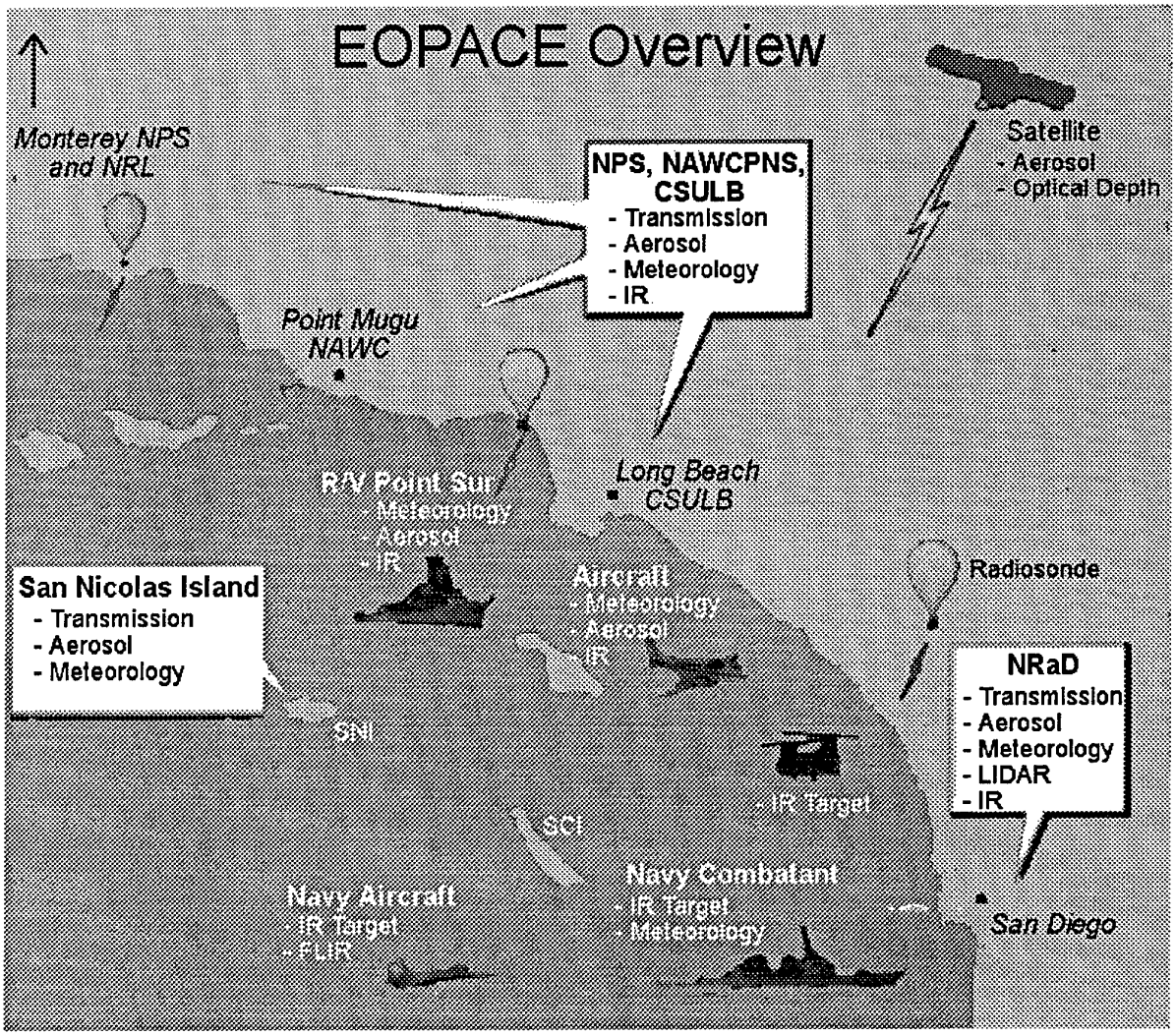


Figure 1. Overview of the southern California EOPACE area.

The performance of EO systems will evaluated from measurements and study of targets and backgrounds, polarization effects, IRST and FLIR performance, and tactical decision aids. Only long term observations assure encountering the full range of atmospheric conditions.

2. COASTAL AEROSOLS

Aerosols in the coastal marine environment are becoming an increasingly important concern for the modern Navy. Diverse threats range from speeding patrol boats in shallow waters to missiles approaching from over the horizon. IR systems are an important complement to existing radar systems for the surveillance, detection, and identification of these threats, but these systems are sensitive to coastal aerosols which are not as well understood as those in the open ocean. The coastal environment is far more complex than that of the open ocean.

Anticipated results will be used to (1) develop an effective coastal aerosol model which will represent the aerosol and its optical effects based on measurable parameters, and (2) the investigation of possible remote sensing techniques for assessing these aerosols.

2.1 Surf Production of Aerosol

The objectives of the surf production effort are twofold: (1) to determine the impact of surf-generated aerosols on visual and infrared (IR) extinction in coastal environments, and (2) to evaluate the measurable meteorological and physical oceanographic parameters of a surf zone by which surf aerosol production may be estimated.

The first intensive surf aerosol measurement period had just been conducted at the Scripps Pier in San Diego, California from January 22 to February 9, 1996. The pier extended seaward beyond the surf line, allowing for a set of measurements to be taken both in and out of the surf for comparisons. Some very interesting phenomena were observed during the Scripps lidar measurements (see Figure 2). The beam had a vertical angle spread, but no horizontal angle, so a vertical "slice" of the surf and the area above it could be seen. Dynamic, plume-like structures of surf aerosol could be seen above the surf. They appear in the figure as vertical lines in the light beam. The bright lights on the right side of the figure are lights at the ocean end of Scripps Pier.

The second surf aerosol measurement period was concluded at the Moss Landing Pier in Monterey, California from March 4 - 15, 1996, and a third is planned at Duck, North Carolina, in 1997.

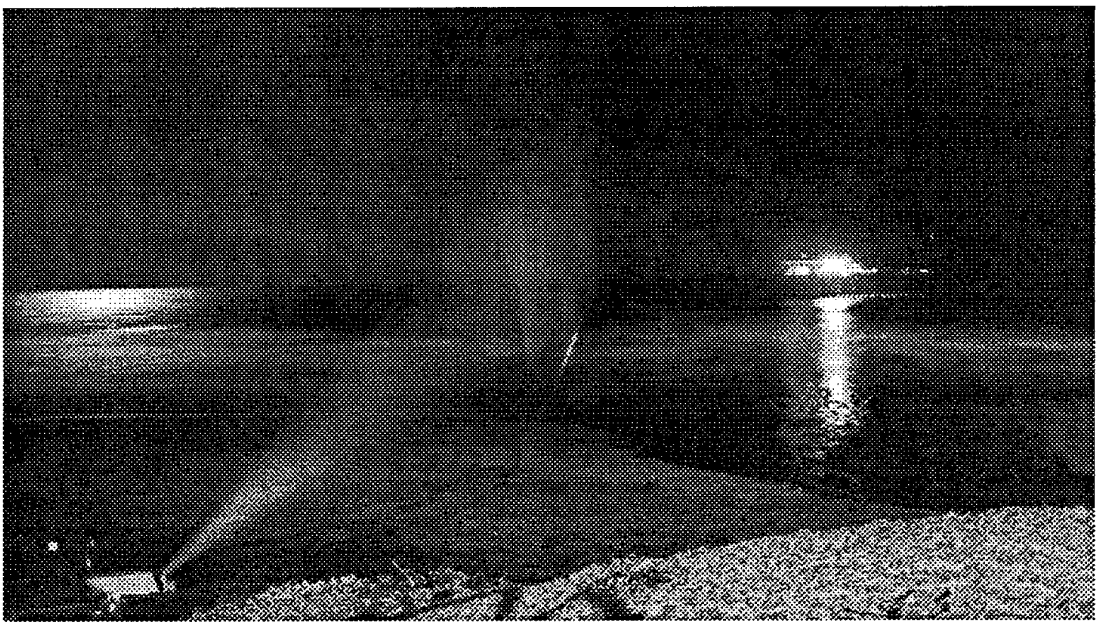


Figure 2. LIDAR beam shot from shore to the ocean over surf line. Vertical plumes of aerosol produced by the surf can be seen in the light beam. The bright lights right of center are on the ocean end of Scripps Pier in San Diego, California. Photo is courtesy of Dr. C. R. Philbrick, Pennsylvania State University.

2.2 Coastal Air Mass Characterization

The objectives of the air mass characterization effort are: 1) to establish the variability of aerosol concentrations and composition for coastal air masses, 2) to determine if the air mass parameters in various coastal locations can be derived from remotely sensed satellite imagery, 3) to evaluate the optimum satellite-derived air mass parameters for Navy real-time assessment and dynamic aerosols models, and 4) to provide a database for initializing and testing the mesoscale coastal aerosol models being developed.

Characterization of the air mass in southern California will be made using satellite optical depths, surface-based sunphotometers, lidar, condensation nuclei, radon, soot carbon loading, aerosols, aerosol chemistry, radiosondes, air trajectories, meteorological measurements and air pollution monitoring through a cooperative effort with the California Air Resources Board. Equipment will be installed in two coastal areas (San Diego and Point Mugu), a coastal island (San Nicolas Island), a ship (the Point Sur research vessel), and an instrumented aircraft. Measurements from these platforms will be compared to satellite information.

The optical and infrared (IR) properties of coastal aerosol are dependent on both the particle concentration and the index of refraction. It is

important to understand the coastal aerosol chemistry in order to estimate the index of refraction, and this chemistry can be complex and variable, ranging from sea salt to various man-made pollutants and sources.

2.3 Near Ocean Surface Transmission

The objective of the near ocean surface transmission characterization is to quantify IR propagation characteristics for the 3-5 and 8-12 micron wavebands for near ocean transmission, and to determine the measurable meteorological parameters near the ocean surface by which IR transmission may be estimated.

Infrared transmission properties will be determined by establishing low level over-the-water IR transmission links supported by IR imaging, refractivity, lidar, and on-path and end-path meteorological and aerosol measurements. Transmission paths will be located across the San Diego Bay and the Monterey Bay. Figure 3 shows the geographic location of the 7 km and 15 km transmission paths at San Diego, and Figure 4 shows the 22 km Monterey Bay transmission path.

During the intensive operational period ending February 9, 1996 in San Diego, a short ¼ km transmission path was used from the seaward end of Scripps Pier to a point on the coast. During the next intensive operational periods, the longer 7, 15, and 22 km paths will be supported continuously by on-path measurements. Buoys will be located mid-path to give necessary meteorological and wave data for modeling the near surface transmission. Turbulence information will be available from the buoy on the two longest paths.

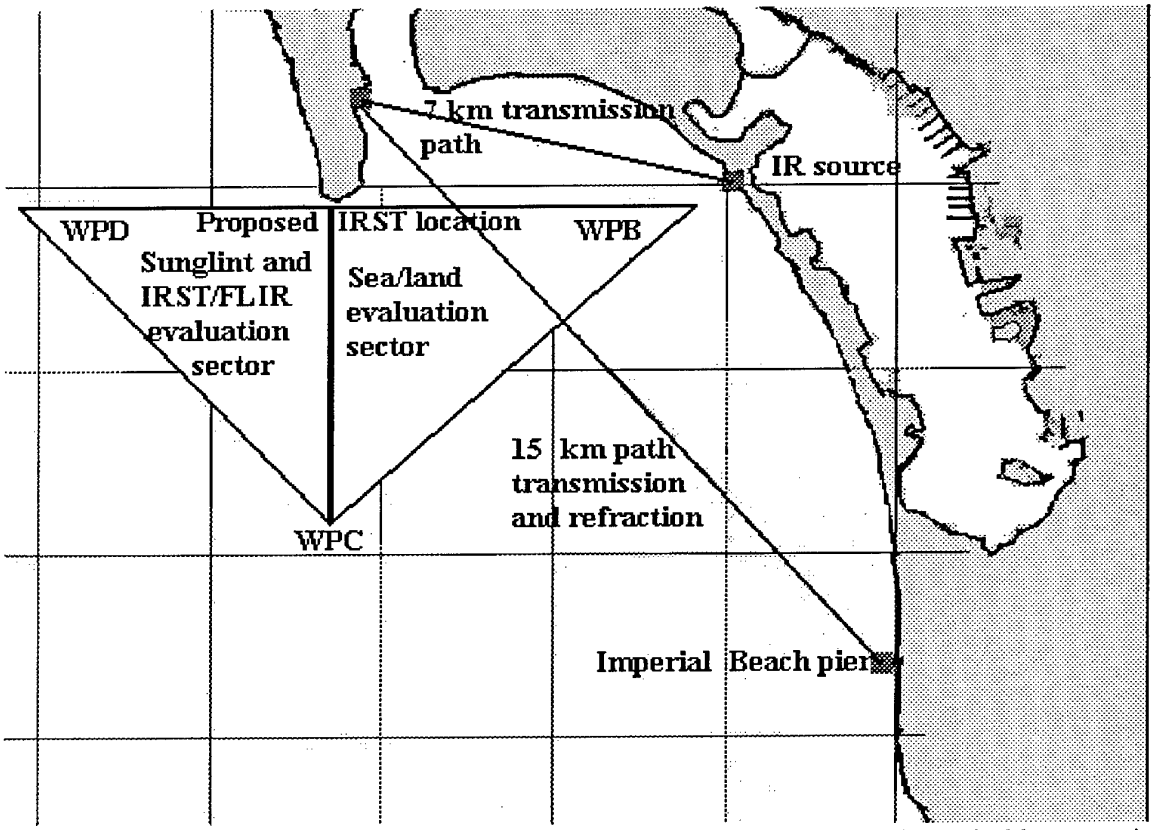


Figure 3. Map shows the area for the San Diego intensive operational period. The 7km and 15 km transmission paths are shown. The triangle defines the operational area for IRST and sensor imaging which will allow for a variety of backgrounds.

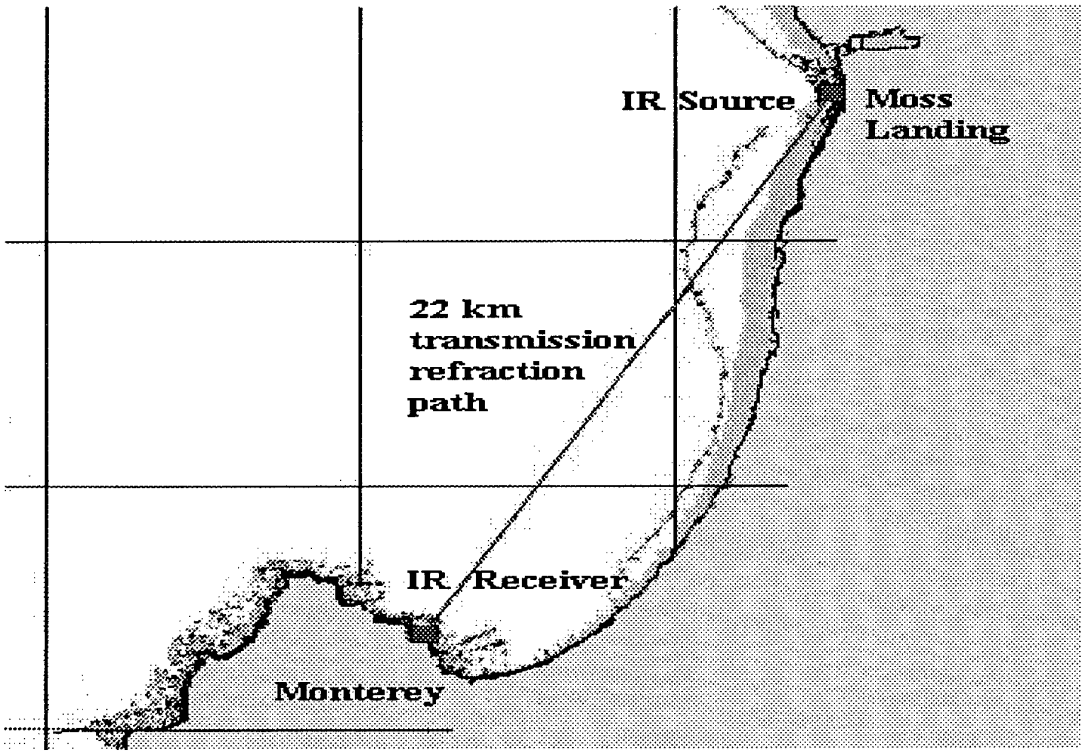


Figure 4. Map shows the area for the intensive operational period in Monterey, California. The 22km transmission path is shown. A buoy will be placed mid-path to collect meteorological information.

3. DEVELOPMENT OF MESOSCALE AND DATA ASSIMILATION MODELS

Both spatial and temporal dynamics of the mesoscale meteorological setting within which EOPACE is conducted must be carefully analyzed if the EOPACE results are to be synthesized and generalized. This will involve the collection of routine and special data, as well as the development of readily usable and accessible databases. Hourly meteorological surface charts from a wide variety of sources will be gathered, analyzed, and used for constructing trajectories and establishing aerosol transport. Frequent radiosonde launches will be conducted at a number of locations. Ship, aircraft, and satellite observations will add to this analysis process. Multi-wavelength images of the EOPACE area are routinely made. Mesoscale models will be used to assimilate this data in a dynamically consistent framework.

Satellite-derived estimates of aerosol loading, along with surface and aircraft aerosol concentration measurements will be used to map the aerosol distribution within the EOPACE domain. Mesoscale models will be used to forecast the transport and diffusion of aerosol. The term "forecast" here means that the models will use EOPACE data after-the-fact in a case study fashion to both initialize and validate models. Evaluation of the strengths and shortcomings of the models' abilities to forecast mesoscale meteorological conditions is a vital component of improving these models' capacity to support electrooptical and electromagnetic requirements. However, EOPACE takes a major step beyond traditional meteorological field experiments by providing the aerosol observations necessary to test the coupling of aerosol codes such as the Navy Aerosol Model (NAM) and the Navy Oceanic Vertical Aerosol Model (NOVAM) to the mesoscale codes .

4. EVALUATION OF EO SYSTEMS PERFORMANCE

The evaluation of infrared search and track (IRST) and imaging systems operating within a

coastal region for differing land, sea, and sky backgrounds will be addressed during EOPACE.

4.1 Targets and Backgrounds

The objectives of the target and background effort are twofold: 1) to assess the performance of EO systems when operating in a coastal environment with both land and sea backgrounds and to compare these results with the predictions of existing background models, and 2) to develop a comprehensive set of coastal and marine backgrounds, combined with a variety of targets and meteorological support for the refinement and development of marine and coastal background radiance models.

The San Diego site is well suited for a variety of backgrounds: land-based looking out to sea, sea-based looking into land, and land-based looking over water to land, with a variety of locations providing industrial, residential, and uninhabited backgrounds.

Targets include Navy combatants of opportunity, a dedicated boat instrumented with a calibrated target and meteorological station, a dedicated helicopter with a suspended calibrated target, Navy aircraft of opportunity, FLIR instrumented Navy aircraft, and a high-speed, low-profile, inflatable boat of the type from which a hand-held missiles could be launched. Portable calibrated targets will also be placed at chosen locations along the South San Diego Bay to utilize the variety of backgrounds described above. Comprehensive meteorological data will be collected.

The results of these observations will be compared with the existing background models and used to validate and improve them. The meteorological data will be used as inputs to the propagation codes.

4.2 Polarization Effects

The objectives of the polarization effort are:
1) to evaluate and validate potential polarization techniques for enhancement of IR target and background contrast in a coastal environment with land, sea, and sky radiance backgrounds,
2) to assess polarization techniques for extinction measurements, and 3) to gather environmentally characterized target and background data to support and validate inclusion of polarization in IRST and FLIR performance models.

4.3 IRST and FLIR Performance

The objectives of the IRST and FLIR performance effort are to measure IRST and FLIR performance and compare with present propagation prediction codes, and to establish the correlation between the IRST and FLIR performance and the underlying atmospheric driving forces such as transmission, refraction, mirages, turbulence, scintillation, and clutter effects.

The performance of IRST and FLIR systems will be assessed by operating them under all the varying target/background combinations previously mentioned. The results will be used as a data base for the development and evaluation of improved prediction codes which incorporate the coastal land/sea/sky backgrounds and improved coastal aerosol models.

5. TACTICAL DECISION AIDS

The objective of the Tactical Decision Aids (TDA) evaluation effort is: 1) to provide a database for evaluating the end-to-end Electrooptical Tactical Decision Aid (EOTDA) performance in a coastal environment, 2) to determine those modules within the EOTDA which need to be improved for Navy applications, and 3) to test the state-of-the-art land/sea/sky background models for improvement of the EOTDA performance.

An AGEMA 3-5 and 8-12 micron imaging system mounted aboard an instrumented Piper Navajo aircraft will collect calibrated IR images and detection ranges of marine targets and backgrounds. Military F/A-18 and P-3 FLIR systems will be used as available to determine detection ranges of calibrated targets for comparison with the EOTDA predictions.

6. PARTICIPANTS IN EOPACE

Several countries are involved in the EOPACE effort. From the United States, participants include the Office of Naval Research (ONR) in Washington, D.C.; the Naval Command, Control

and Ocean Surveillance Center, RDT&E Division (NRaD) in San Diego, California; the Naval Research Laboratories in Monterey, California and in Washington, D.C.; the Naval Air Warfare Center at Pt. Mugu, California; the Applied Research Laboratory at Pennsylvania State University; California State University at Long Beach; the Naval Postgraduate School in Monterey, California; the Naval Surface Warfare Center in Silver Springs, Maryland; and the California Air Resources Board.

Also participating are the Physics and Electronics Laboratory TNO from The Netherlands; the University of Manchester Institute of Science and Technology (UMIST)in the United Kingdom; the Defense Research Establishment Valcartier in Quebec, Canada; and the Defence Science and Tecnology Organisation/University of Western Australia in Perth, Western Australia.

7.0 CONCLUSION

The EOPACE effort is ambitious and well-focused. Electrooptical systems operating in coastal environments are subject to varied and complex atmospheric conditions which can change rapidly. The modeling of these conditions and improvement of existing transmission codes will greatly enhance our ability to predict the performance of electrooptical systems under a variety of conditions. The involvement of so many participants ensures the highest level of expertise in each area and a wide dissemination of results.

More specific information about EOPACE is available on the internet under the URL http://sunspot.nosc.mil/543/ , and then selecting "EOPACE experiment."

7.1 ACKNOWLEDGMENTS

EOPACE is sponsored by the Office of Naval Research, Code 322.

Paper 10

N. Farsaris (GR)

Have there been experiments in rough sea conditions?

Author's reply:

Turbulence is being measured on buoys along transmission paths, but is not a major part of the study. I would refer you to papers on the MAPTIP (Marine Aerosol Properties and Thermal Imaging Program) experiment in the North Sea in October1993, for more information on rough sea conditions.

Lateral Resolution Considerations of Refractivity Profiles Associated with Remote Sensing Measurements of Over-Water Coastal Regions

Julius Goldhirsh, G. Daniel Dockery
Applied Physics Laboratory, The Johns Hopkins University
Johns Hopkins Road, Laurel, Maryland 20732-6099, U.S.A.
email:julius_goldhirsh@jhuapl.edu

1. SUMMARY

A means to assess propagation conditions is to obtain sets of vertical refractivity profiles employing remote or in situ sensing methods, and to inject these profiles into propagation models from which we calculate propagation factor fields. Fundamental questions which are addressed in this effort are: (1) What errors are introduced when only a single measurement of refractivity profile is possible and the refractivity environment is assumed laterally homogeneous? (2) Where multiple profiles are capable of being measured, what errors are introduced for different profile spacings? Errors in the propagation factor field at 1 GHz, 3 GHz, and 10 GHz over broad areas of the lower troposphere are quantified in this paper as they relate to the above questions for two coastal regions; namely, the coastal regions of Wallops Island, Virginia and San Nicolas Island, California.

2. INTRODUCTION

Means may become available in the future to remotely sense the vertical refractivity profile near the earth's surface by using appropriate sensors on board a ship, satellite or aircraft [1-4]. The refractivity environment may also be determined employing dropsondes from manned or unmanned aircraft which telemeter pressure, temperature, and humidity to a remotely located receiver [5]. Refractivity profiles measured by these means may subsequently be used with an appropriate propagation model for assessment of propagation conditions [6-10]. For such scenarios, important questions which should be broached are: (1) What are the incurred errors in the propagation factor field by employing a single profile measurement assuming lateral uniformity of the refractivity environment? (2) What should the profile spacing be to achieve predesignated errors in the propagation factor fields? (3) How do the above errors depend on frequency?

The objective of this paper is to partially address the above questions by considering in detail two sets of helicopter measured refractivity profiles along the east and west coasts of the United States. In particular, the profiles selected correspond to coastal regions off of Wallops Island, Virginia and San Nicolas Island, California. Injecting these measurements into appropriate models, "truth" estimations of the propagation factor environment are obtained at frequencies of 1, 3, and 10 GHz. These are compared with propagation fields derived by "spoiling" the resolutions of the original set of measurements in the lateral dimensions.

A related paper was given by the authors at an AGARD meeting which described (a) propagation factor levels versus range for selected heights, and (b) propagation factor levels versus height at selected ranges for various lateral and vertical measurement resolutions [11]. The work presented here amplifies on reference [11] in that propagation factor error fields are examined over a designated contiguous low altitude area; where we focus our attention on the lateral profile spacing issue. Errors caused by spoiling the resolution are quantified via cumulative distributions which describe the percentage of the area over which errors exceed various levels.

Because only two cases are considered, these results are not meant to establish conclusive measurement resolution criteria, but are meant to (1) demonstrate a methodology by which errors may be quantified over an area, (2) establish the level of errors in the propagation factor fields for the cases considered, and (3) amplify the seriousness of the above questions broached, especially for those cases in which the refractivity field is laterally inhomogeneous. It is the intention of the authors in the future to consider a sufficient number of cases so that the above questions may be addressed more conclusively.

3.0 DATA BASE AND ACQUISITION METHODOLOGY

Since 1984, JHU/APL has collected numerous atmospheric profiles of pressure, temperature, and relative humidity during propagation experiments in support of U.S. Navy exercises employing a civilian helicopter equipped with sensors [5]. These measurements enable a determination of the refractivity N(h) given by [12]

$$N(h) = 77.6 \frac{P(h)}{T(h)} + 3.73 \times 10^5 \frac{e(h)}{T(h)^2} \quad (1)$$

and modified refractive index M(h) given by

$$M(h) \approx N(h) + 157\, h \quad (2)$$

where P(h), e(h), and T(h) are the total pressure (mb), the partial pressure of water vapor (mb), and the temperature (K) measured at the height h, respectively, where h in (2) is expressed in km.

Paper presented at the AGARD SPP Symposium on "Remote Sensing: A Valuable Source of Information", held in Toulouse, France from 22-25 April 1996, and published in CP-582.

Profiles of the modified refractive index described in this report have been acquired with the helicopter by flying "sawtooth" patterns, with measurements being recorded during the descending portions. Data were recorded at a 2 Hz rate resulting in typical vertical and lateral measurement resolutions of approximately 0.8 m and 15 m, respectively. The lateral interval over which a profile is recorded is dependent upon the beginning altitude and the wind conditions. For a nominal beginning altitude between 200 to 250 m, the helicopter travels 5 to 7 km during the descent. This procedure results in typical range separations (distance between profile centers)of between 7 and 9 km.

4.0 TEMPER PROPAGATION MODEL

Injection of helicopter profile data and evaporation duct information into high-fidelity propagation models have typically yielded low altitude signal predictions to within 5 dB for ranges as large as 75 km from the source [13, 14]. The propagation model used here is called Tropospheric Electromagnetic Parabolic Equation Routine (TEMPER). The model output is the two dimensional field of propagation factors (i.e., an array of propagation factors in height and range). The input is a set of height profiles of the refractive index or modified refractive index, where each subsequent profile is referenced to an increasing lateral distance from the transmitting source. This model is based on the Fourier split-step numerical solution of the parabolic wave equation, and has to date demonstrated robust performance for environments containing complex range of varying refractive profiles [10, 14].

4.1 LARRI

In advance of injecting the measured profiles into TEMPER, a preliminary processing of the data is required. The processing routine used here is called the Large-scale Atmospheric Refractivity Range Interpolator (LARRI) [15]. LARRI automatically performs the following steps: (1) smooths the individual refractivity profiles, (2) extrapolates the profiles to the ground and to a pre-selected height, (3) fits connecting line segments to the smoothed profile, (4) identifies profile features and categorizes them, (5) finds best possible matches between profile features of adjacent profiles, and (6) interpolates between measured adjacent profiles, producing additional profiles at predesignated lateral intervals which preserve the matched features of (5). The processed set of profiles are entered into TEMPER which performs a further interpolation of these data onto its internal calculation grid, which is typically much finer in both range and altitude. Resolution cell sizes of the propagation factor field generated by TEMPER are approximately 0.2 m in height and 200 m in range.

5. ANALYSIS METHODOLOGY

Results are presented here for frequencies of 1, 3 and 10 GHz, and for a fixed antenna at a height of 20 m above the ocean surface, having a beamwidth of 2°, and pointing at zero elevation. The investigation focuses on the determination of propagation factors at altitudes between 5 m and 300 m and ranges up to 90 km from the source. For purposes of simplification, evaporation duct effects are neglected. The LARRI software interpolates the measurements made above 3 m and extends the profile to the surface.

5.1 Definition of Benchmark Cases

The methodology employed here is to investigate the extent to which various profile spacings degrade the "benchmark" predicted propagation factor levels calculated using TEMPER. The benchmark cases are taken to correspond to the propagation factor fields generated from the original full set of measured profiles; these are regarded as the "truth" cases. As mentioned above, the "truth" cases were found to give valid signal predictions to within 5 dB.

5.2 Spoiling of Lateral Resolution

We frequently do not have the luxury of making a set of multiple, closely measurements of refractivity. In establishing the error fields, the cases considered here correspond to the following: (1) a single profile from the set assuming lateral homogeneity of the refractivity environment, (2) the first and last profiles of the set, (3) the first, middle, and last profiles of the set, (4) four and five profiles out of the set. For those cases where three or more profiles are considered, the average spacings between profile centers are noted.

5.3 Smoothing

Propagation factors at the above frequencies tend to exhibit sharp nulls and maxima due to interference phenomena caused by multipath from surface reflections and competing modes within surface ducts. The exact locations of the nulls and maxima are very sensitive to changes in the refractive environment and differences of tens of dB can be caused by slight shifts of the propagation measurement location or dynamic changes of the environment. The precise prediction of these maxima and minima is not considered important for most applications because they generally occur over a small region of space and are dynamically changing for any fixed measurement point.

At the above frequencies, TEMPER typically generates results on a grid of spacing 200 m in horizontal range and 0.2 m in height. In order to mitigate the effects of deep nulls and sharp maxima and make a comparison between different resolution cases more realistic, the TEMPER output grids of propagation factor were subsequently smoothed for all cases by averaging over approximately 15 m in height and 2 km in horizontal range.

5.4 Analysis of Error Fields

We quantify the measurement resolution errors by differencing the smoothed benchmark and smoothed degraded propagation factor fields. The differenced fields are depicted here by two dimensional "color modulated" representations which generally have the appearance of contoured regions. We further quantify errors by selecting appropriate range and height intervals and determining cumulative distributions which describe the percentage of area having an error magnitude greater than designated levels. In essence, these values may be roughly taken as the probability of exceeding various error levels relative to the benchmark case over a defined geometric region. Examples, of the contoured color modulated displays and associated distributions are given in the following sections.

5.5 Lateral Resolution Studies: Wallops Island Case

In Figure 1 is given a set of modified refractive index (M) profiles measured off the coast of Wallops Island, Virginia during August 24, 1994 during the time interval 07:57–08:15 (local time). The vertical scale is the altitude above the surface (m), the lower abscissa denotes an assigned profile

number. As an example, of the variability of M, the scale showing an interval of 340 to 390 for the first profile is given in the upper left part of the figure. Each profile is marked with the range interval over which the measurements were made relative to a simulated transmitting source located on a hypothetical ship at sea. The indicated distance intervals are in the direction towards the shore and range from 52 km (profile #1) to 14 km (profile #5) from the shore line The average distance between profile centers is approximately 8 km.

These profiles show radical changes over the indicated distance interval. For example, the first profile is representative of an approximate "standard" atmosphere case, whereas, the last profile corresponds to a surface duct propagation condition with a duct height of 100 m. These profiles are shown in Figure 2 after being smoothed by LARRI.

5.6 Error Effects for Single Profile Case: Lateral Homogeneity Assumption

Many environmental measurement systems such as those associated with coastal or shipboard radiosondes or rocketsondes rely on a single height-profile measurement and assume lateral uniformity to arrive at a propagation assessment. Remotely sensed profiles may also be limited to a single profile over a particular propagation assessment region. Although such systems are generally better than having no measurements at all, we examine the levels of errors that may arise through the assumption of lateral uniformity. It is assumed in our study that the profile closest to the ship represents the single measured profile, although any one of the indicated profiles could have been used.

TEMPER-derived color modulated displays of propagation factor in dB at f = 3 GHz for the benchmark and the single profile cases are shown in Figure 3 by the upper and lower left inserts, respectively. The indicated scale is given in the bottom of the page with bright red representing propagation factors of 5 dB and greater and green representing values between -10 and -20 dB. The range interval between 0 and 90 km (from the ship towards the shore) and height interval between 0 to 300 m are depicted. It is apparent that the single profile assumption gives rise to large errors in the propagation factor field; these errors will be quantified in the next section. Similar propagation factor fields (not shown) were generated at frequencies of 1 GHz and 10 GHz.

The corresponding errors associated with the single profile assumption are given in the upper insert of Figures 4, 5, and 6 for frequencies of f = 1 GHz, 3 GHz, and 10 GHz. These color modulated diagrams show absolute errors whose dB error scales are located at the bottom of the figures. These figures were derived by subtracting the dB values of the propagation factors from the same resolution cells of the benchmark and single profile fields (e.g., difference between the upper left and bottom left inserts of Figure 3). We note that errors of less than 5 dB exist over the range interval between 0 to 30 km and substantially large errors (greater than 20 dB) may exist elsewhere.

The above described errors are quantified in the cumulative distributions given in Figures 7–9 which correspond to errors over the area defined by the range interval 30–90 km and height interval 5–300 m. The 50% and 10% values are summarized in Table 1. For example, errors in excess of 6 dB, 12 dB, and 12 dB occur over 50% of the area at 1, 3, and 10 GHz, respectively. Over 10% of the area, errors in excess of 20 dB occur at all three frequencies.

5.7 First and Last Profiles: Spacing = 33 km

The 3 GHz propagation factor field associated with the first and last measured profiles (e.g., profile #1 and #5 of Figure 1) is shown in Figure 3 by the upper right insert. We note that this diagram begins to resemble the benchmark case (upper left insert) and should have less errors associated with it. The corresponding error diagrams are given by the center inserts in Figures 4–6, and the cumulative distributions are shown in Figures 7–9. The two-profile case (Table 1) shows that errors in excess of approximately 4 dB, 5 dB, and 12 dB occur over 50% of the area and 12 dB, 14 dB, and 20 dB occur over 10% of the area at 1, 3, and 10 GHz, respectively.

5.8 First, Middle, and Last Profiles: Spacing = 17 km

A third case considered assumes measurements of the first, middle, and last profiles (profiles #1, #3, and #5 of Figure 1), where the average spacing between profiles is approximately 17 km. The corresponding color modulated display of propagation factor field at 3 GHz is represented by the lower right insert of Figure 3. Of the previous three cases, this figure appears to best resemble the benchmark scenario. The error diagrams are given by the bottom inserts of Figures 4-6 and the cumulative distributions are shown in Figures 7–9. The 50% errors are in excess of approximately 3.5 dB, 4 dB, and 8 dB at 1, 3, and 10 GHz, respectively (Table 1). It is interesting to note that little noticeable change occurs at 1 GHz and 3 GHz between the two and three profile cases (Figures 7 and 8). For the 10 GHz case, significant differences occur at the smaller percentages (Figure 9).

5.9 Four Profile Case: Spacing = 11 km

As a demonstration that the errors may be dramatically reduced as more profiles are inserted into the system, we have considered a four profile case corresponding to profiles 1, 3, 4, and 5 of Figure 1. The average profile spacing for this case is 11 km. The corresponding cumulative distributions are given in Figures 7–9 at 1, 3, and 10 GHz. These figures show a dramatic reduction in the errors at 1 and 3 GHz and moderate reduction at 10 GHz. For example, at the 10% point, the errors approximately exceed 5 dB at all three frequencies (Table 1).

5.10 Dependence on Frequency

The effect of frequency on errors has been alluded to in the previous section. It may however be better observed through an examination of Figures 10-13, the tabulations in Table 1 at the 50% and 10% thresholds, and the vertical bar chart in Figure 14. Each of the figures describe a particular spacing scenario and show the cumulative distributions at 1 GHz, 3 GHz, and 10 GHz. For the single profile case assuming lateral uniformity (Figure 10), no frequency trend is observed. On the other hand, for the two profile case (Figure 11) in which the spacing is 33 km, a clear trend is observed where the larger frequencies show systematically greater errors. This is also observed in Figures 12 and 13 for the three and four profile cases, respectively; the 10 GHz distributions stand out in that they result in substantially greater errors. The smaller wavelength at 10 GHz causes the propagation factor to be more sensitive to small changes in the refractivity field, thus producing larger errors as the profile resolution is degraded. Figure 14 shows a bar chart of errors for the different profile spacing cases at the 10% of

area value. The black, cross hatch, and horizontal line bars represent 1 GHz, 3 GHz, and 10 GHz cases, respectively. It is apparent that the larger frequencies result in greater errors for the different profile spacings. For the single profile case assuming lateral uniformity, the errors are in excess of 20 dB at all three frequencies considered.

6.0 LATERAL RESOLUTION STUDIES: COAST OF SAN NICOLAS ISLAND

Figure 15 presents a set seven modified refractive index profiles measured off the coast of San Nicolas Island, California on August 19, 1988 during the period 09:59–10:22 (local time). As in Figure 1, the vertical scale is the altitude above the surface (m), the lower abscissa denotes the assigned profile number, and the variability of M for the first profile is denoted by the upper left scale showing an interval of 300 to 350. As in Figure 1, each profile is marked with the range interval over which the measurements were made relative to a simulated transmitting source located on a hypothetical ship at sea, where the indicated distance intervals are in the direction towards the shore. As was the case for the Wallops Island set of profiles, the average distance between profile centers is approximately 8 km.

Although the profiles appear dissimilar over smaller scales, each of the profiles have the common feature of being representative of a surface duct with an upper surface duct height at approximately 100 m. Profiles #1 and #2 exhibit a more complex "double surface duct" type of feature with the lower surface duct height occurring at approximately 30 m and the upper one near 100 m.

Following identical methods discussed for the Wallops Island set of profiles, cumulative distributions of errors in propagation factor were derived for the following cases: (1) Single profile assuming lateral homogeneity, (2) First and last profile (spacing = 50 km), (3) First, middle, and last profile (average spacing = 25 km), (4) Four (of seven) profiles (average spacing = 17 km), and (5) Five of (seven) profiles (average spacing = 13 km).

The results shown in Figures 16–23 correspond to cumulative distributions describing the percentage of area in which errors exceed the abscissa value for an area whose range is in the interval 15 km to 90 km, and height is between 5 m and 300 m. Results are also tabulated in Table 2 for the 50% and 10% points. Figure 24 is a bar chart of the 10% area case depicting the errors at the three frequencies.

6.1 Single Profile Case: Lateral Homogeneity Assumption

Assuming a measurement only of profile #1 concomitant with the assumption of lateral uniformity, cumulative distributions of errors generated are given in Figures 16-18 for frequencies of 1 GHz, 3 GHz, and 10 GHz, respectively (solid curve). The 10% fades are 19 dB, 14.5 dB, and 18.2 dB at the respective frequencies (Table 2). Although these errors may be significant, they are somewhat smaller than for the Wallops Island scenario in which the errors were all in excess of 20 dB at the 10% point.

6.2 Errors Versus Profile Separation

The common result for all three frequencies is that a profile spacing of 13 km (e.g., 5 profiles) is required to bring the 10% value down to 7 dB or less. The 1 GHz case (Figure 16) shows that a radical reduction in error occurs between spacings of 17 km (four profiles) and 13 km (five profiles). For example, the 10% error is 15.6 dB for the 17 km spacing versus 4.9 dB for the 13 km spacing. On the other hand, at 3 GHz (Figure 17), errors of 10.9 dB and 6.2 dB arise at the 10% case for the 17 km and 13 km spacing cases, respectively (Table 2). At 10 GHz (Figure 18), respective errors of 9.4 dB and 6.5 dB arise at the 10% point.

6.3 Errors Versus Frequency

Figures 19-23 give cumulative distributions of errors showing the effect of frequency for specific spacings. A general perusal of these figures and the bar chart results of Figure 24 indicates that except for the five profile scenario (average spacing = 13 km), no well defined error level versus frequency trend exists, unlike the Wallops Island case as depicted in Figure 14.

7. CONCLUSIONS

As previously mentioned, the results presented here are specific to the two sets of refractivity measurements discussed, as well as to the operational scenario under consideration. Specifically, the analysis pertains to the detection of a moving target within a defined region. The regions are 30 km to 90 km in range and 5 m to 300 m in altitude for the Wallops case, and 15 km to 90 km in range and 5 m to 300 m in altitude for the San Nicolas case.

Nevertheless, this effort has described and demonstrated a general methodology for the determination of errors in propagation factor over defined areas. Furthermore, application of this methodology to profile sets off the coast of Wallops Island and San Nicolas Island (east and west coasts, respectively) has revealed certain trends in propagation prediction accuracy versus lateral resolutions. In particular, analysis of the profile measurement cases established the following common features relative to the 10% error exceedance value: (1) Measuring a single profile at an at-sea location and assuming lateral uniformity of the refractivity produced errors in excess of 15–20 dB at 1 GHz, 3 GHz, and 10 GHz. (2) Measured profile spacings between 11 km to 13 km apart produced errors of less than approximately 6 dB at 1 GHz, 3 GHz, and 10 GHz. (3) A profile spacing of 17 km produced errors as high as 17 dB at 10 GHz for the Wallops coast case and 16 dB at 1 GHz for the San Nicolas coast case. Also, for the Wallops Island case, the error systematically increased with increasing frequency. This systematic increase of error with increasing frequency was not observed in the San Nicolas Island results.

8. ACKNOWLEDGEMENTS

The authors thank Bert Musiani for his assistance in developing software related to the results of this effort. This work was supported by the AEGIS Shipbuilding Program, PMS-400 under Navy Contract N00039-95-C-0002.

9. REFERENCES

1. Melfi, S. H., D. Whiteman, and R. Ferrare, "Observation of Atmospheric Fronts Using Raman Lidar Moisture Measurements," *J. Appl. Meteor.*, Vol. 28, Sept. 1989, pp. 789-806.

2. Grant, W. B., "Differential absorption and Raman lidar for water vapor profile measurements: a review," *Opt. Eng.*, Vol. 30, No. 1, 1991, pp. 40-48.

3. Gossard, E. E. and N. Sengupta, "Measuring gradients of meteorological properties in elevated layers with a surface-base Doppler radar," *Rad. Sci.*, Vol. 23, No. 4, 1988, pp. 625-639.

4. NATO Advisory Group for Aerospace Research and Development (AGARD) Conference Proceedings, No. CP-502, "Remote Sensing of the Propagation Environment," Cesme, Turkey, 30 September-4 October 1991 (Proceedings published February 1992, AGARD-CP-502).

5. Rowland, J. R., and S. M. Babin, "Fine-Scale Measurements of Microwave Refractivity Profiles with Helicopter and Low-Cost Rocket Probes," *Johns Hopkins APL Tech. Dig.*, Vol. 8, No. 4, 1987, pp. 413-417.

6. Hitney, H. V., J. H. Richter, R. A. Pappert, K. D. Anderson, and G.. H. Baumgartner, Jr., "Tropospheric Radio Propagation," *Proc. IEEE*, Vol. 73, No. 2, 1985, pp. 265-283.

7. Ko, H. W., J. W. Sari, and J. P. Skura, "Anomalous Microwave Propagation Through Atmospheric Ducts," *Johns Hopkins APL Tech. Dig.*, Vol. 4, No. 1, 1983, pp. 12-16.

8. Dockery, G. D., and E. R. Thews, "The Parabolic Equation Approach to Predicting Tropospheric Propagation Effects in Operational Environments," *AGARD Conference Proceedings No. 453: Operational Decision Aids for Exploiting or Mitigating Electromagnetic Propagation Effects,* AGARD-CP-453, 1989, pp. 18-1 through 18-9.

9. Reilly, J. P., and G. D. Dockery, "Influence of Evaporation Ducts on Radar Sea Return," *IEE Proceedings,* Pt. F, Vol. 137, No. 2, 1990, pp. 80-88.

10. Dockery G. D., "Modeling Electromagnetic Wave Propagation in the Troposphere Using the Parabolic Equation," *IEEE Trans. Ant. Propag.*, Vol. 36, No. 10, 1988, pp. 1464-1470.

11. Dockery, G. D., and J. Goldhirsh, "Atmospheric Data Resolution Requirements for Propagation Assessment: Case Studies of Range-Dependent Coastal Environments," *NATO Advisory Group for Aerospace Research and Development (AGARD) Conference*, Bremerhaven, Germany, 19–22 September 1994. (Proceedings published February 1995, AGARD-CP-567, pp. 7.1–7.13).

12. Bean, B. B., and E. J. Dutton, "Radio Meteorology," Dover Publications, New York, 1968.

13. Dockery, G. D. and G. C. Konstanzer, "Recent Advances in Prediction of Tropospheric Propagation Using the Parabolic Equation" *Johns Hopkins APL Tech. Dig.* Vol. 8, No. 4, 1987, pp. 404-412.

14. Kuttler, J. R. and G. D. Dockery, "Theoretical Description of the Parabolic Approximation/Fourier Split-Step Method of Representing Electromagnetic Propagation in the Troposphere," *Rad. Sci.*, Vol. 26, No. 2, 1991, pp. 381-393.

15. Dockery, G. D., and G. C. Konstanzer, "Recent Advances in Prediction of Tropospheric Propagation Using the Parabolic Equation," *Johns Hopkins APL Tech. Dig.*, Vol. 8, No. 4, 1987, pp. 404-412.

Profile Scenario	Frequency					
	f = 1 GHz		f = 3 GHz		f = 10 GHz	
	50%	10%	50%	10%	50%	10%
Single Profile (Lateral Uniformity)	6.2	>20	12.3	>20	11.8	>20
First and Last (Spacing = 33 km)	3.7	12.5	5.0	13.9	9.8	>20
First, Middle, Last (Spacing = 17 km)	3.6	10.8	3.9	11.6	8.2	17.3
Four Profiles (Spacing = 11 km)	1.5	5.0	1.5	5.1	1.6	6.1

Table 1: Error values in dB exceeded over 50% and 10% of selected region: August 24, 1994 measurements - Coastal region of Wallops Island

Profile Scenario	Frequency					
	f = 1 GHz		f = 3 GHz		f = 10 GHz	
	50%	10%	50%	10%	50%	10%
Single Profile (Lateral Uniformity)	5.1	19.0	5.6	14.5	7.2	18.2
First and Last (Spacing = 50 km)	4.9	16.0	6.0	14.3	4.5	>20
First, Middle, Last (Spacing = 25 km)	4.8	15.7	6.0	17.6	4.6	12.2
Four Profiles (Spacing = 17 km)	3.5	15.6	3.9	10.9	3.1	9.4
Five Profiles (Spacing = 13 km)	1.5	4.9	2.2	6.2	2.1	6.5

Table 2: Error values in dB exceeded over 50% and 10% of selected region: March 19, 1988 measurements - Coastal region of San Nicolas Island.

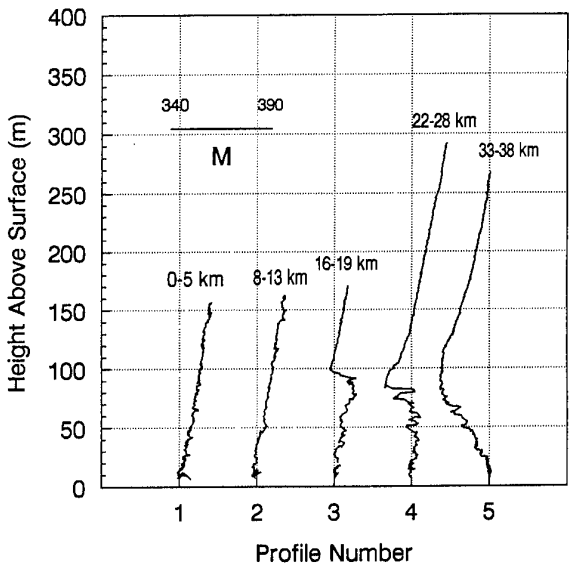


Figure 1: Modified refractive index profiles measured off the coast of Wallops Island, Virginia on August 24, 1994 during the period 07:57 to 08:15 (local time) over a distance 14 km to 52 km from shore. Indicated distance intervals are relative to a simulated ship at sea and are directed towards the shore.

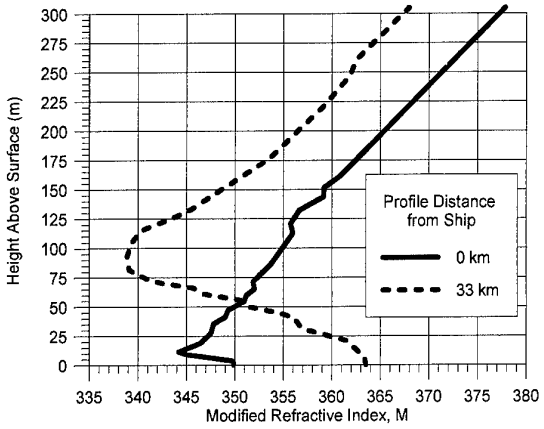


Figure 2: Smoothed versions of modified index of refraction profiles #1 and #5 from Figure 1 (coast of Wallops Island).

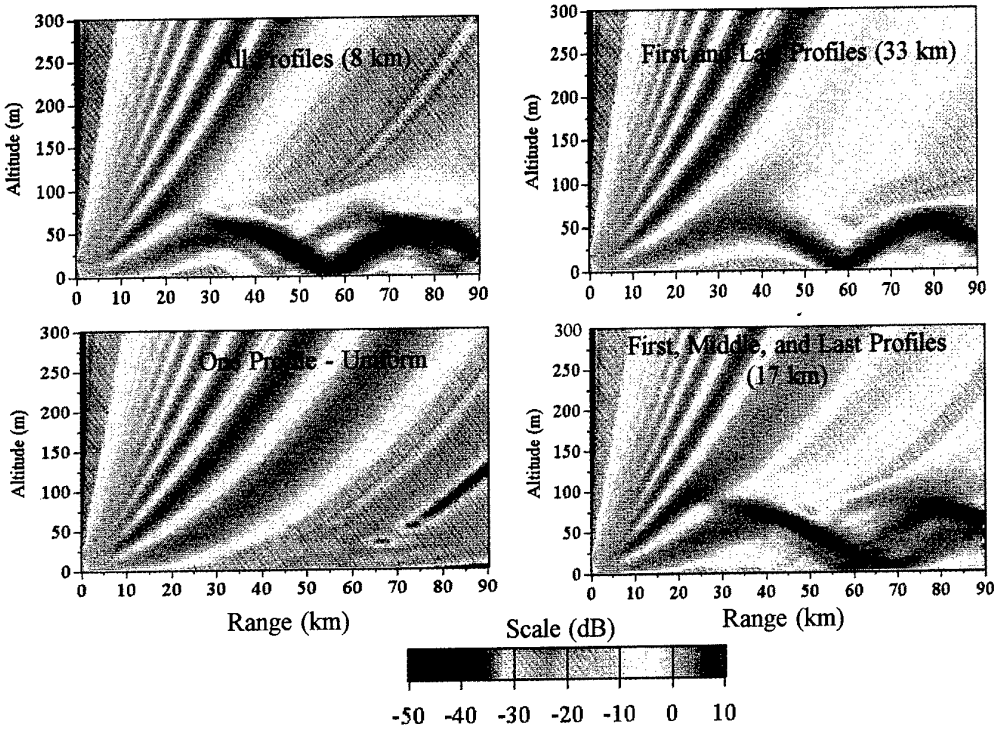


Figure 3: Color-modulated displays of the 3 GHz propagation factor fields calculated using TEMPER for various profile cases (coast of Wallops Island).

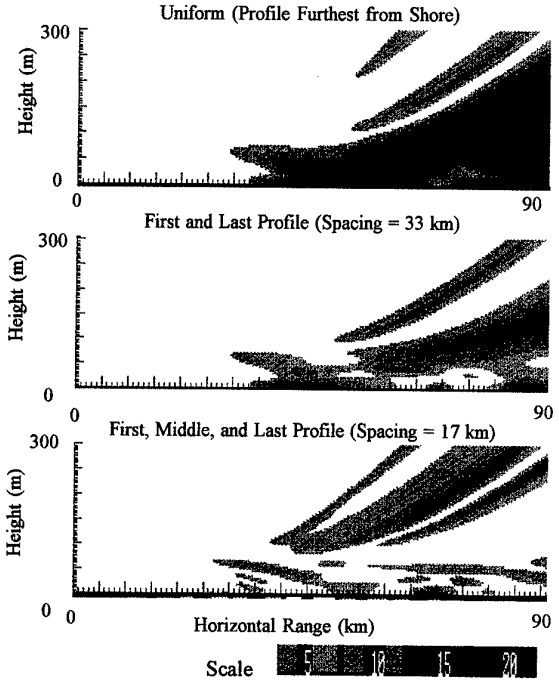


Figure 4: 1 GHz color-modulated diagrams of propagation factor errors in dB for various profile cases (coast of Wallops Island). Large tick mark intervals in range and height are 5 km and 5 m. White denotes errors smaller than 5 dB, green 5 to 10 dB, light blue 10 to 15 dB, red 15 to 20 dB, and dark blue greater than 20 dB.

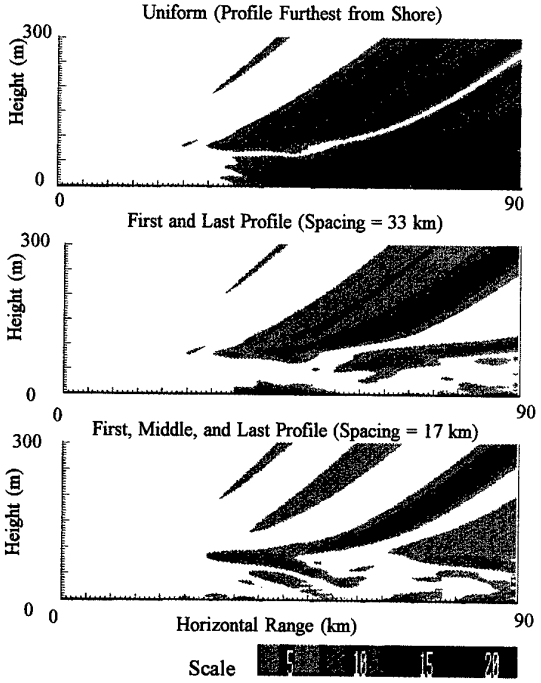


Figure 5: 3 GHz color-modulated diagrams of propagation factor errors in dB for various profile cases (coast of Wallops Island). Large tick mark intervals in range and height are 5 km and 5 m. White denotes errors smaller than 5 dB, green 5 to 10 dB, light blue 10 to 15 dB, red 15 to 20 dB, and dark blue greater than 20 dB.

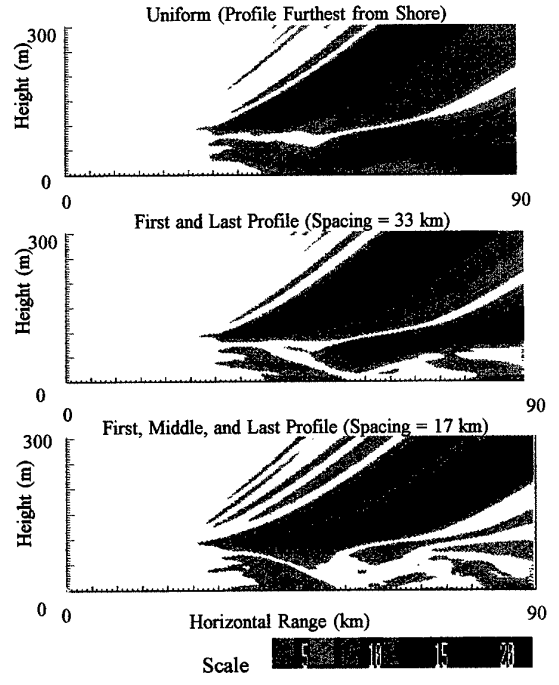


Figure 6: 10 GHz color-modulated diagrams of propagation factor errors in dB for various profile cases (coast of Wallops Island). Large tick mark intervals in range and height are 5 km and 5 m. White denotes errors smaller than 5 dB, green 5 to 10 dB, light blue 10 to 15 dB, red 15 to 20 dB, and dark blue greater than 20 dB.

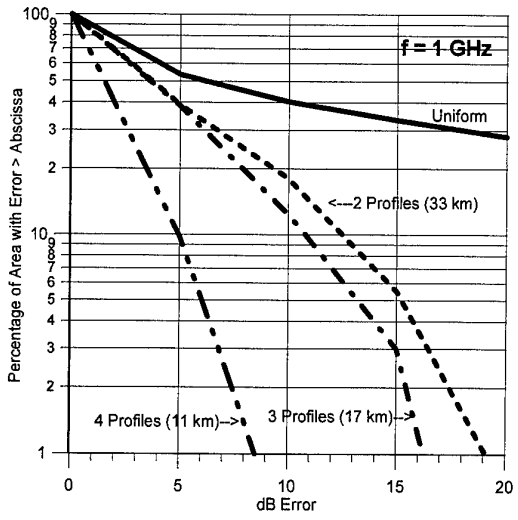


Figure 7: Cumulative distribution of errors at 1 GHz for different profile spacings. Area of consideration is defined by the range interval 30 km to 90 km and height interval 5 m to 300 m (coast of Wallops Island)

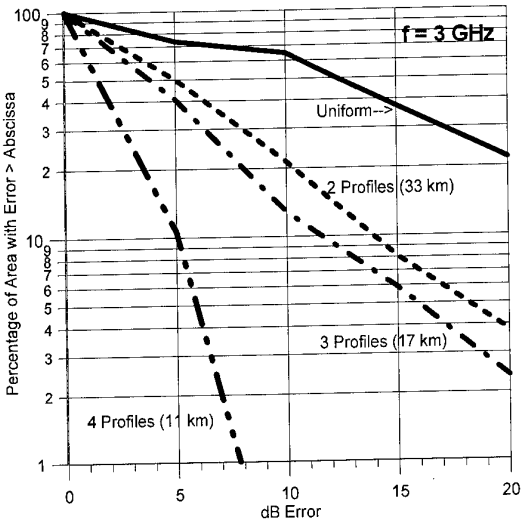


Figure 8: Cumulative distribution of errors at 3 GHz for different profile spacings. Area of consideration is defined by the range interval 30 km to 90 km and height interval 5 m to 300 m (coast of Wallops Island)

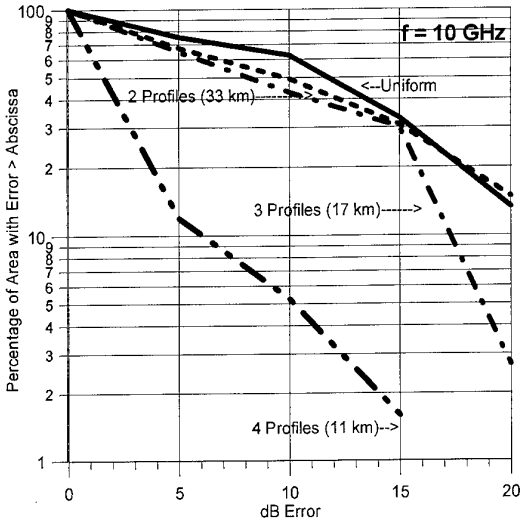


Figure 9: Cumulative distribution of errors at 10 GHz for different profile spacings. Area of consideration is defined by the range interval 30 km to 90 km and height interval 5 m to 300 m (coast of Wallops Island)

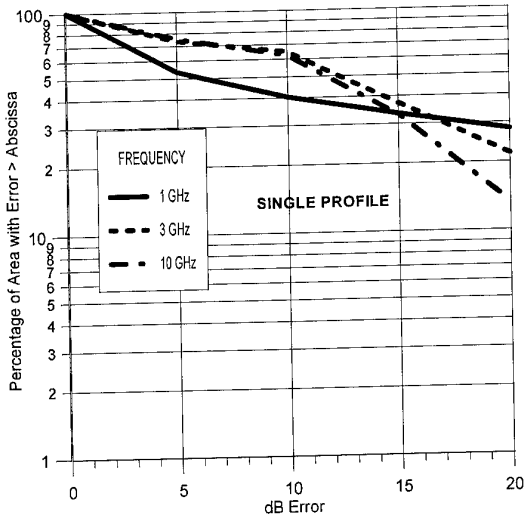


Figure 10: Cumulative distributions of errors at 1 GHz, 3 GHz, and 10 GHz for single profile case (coast of Wallops Island).

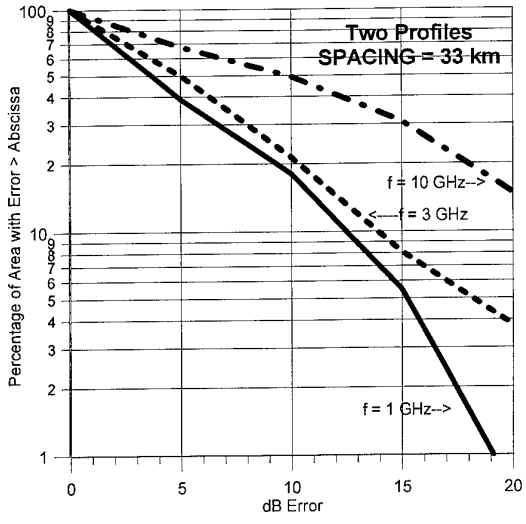


Figure 11: Cumulative distributions of errors at 1 GHz, 3 GHz, and 10 GHz for two profile case with a spacing of 33 km (coast of Wallops Island).

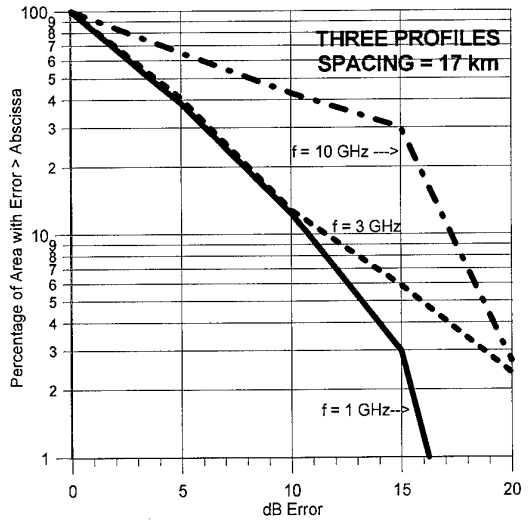


Figure 12: Cumulative distributions of errors at 1 GHz, 3 GHz, and 10 GHz for three profile case with an average profile spacing of 17 km (coast of Wallops Island).

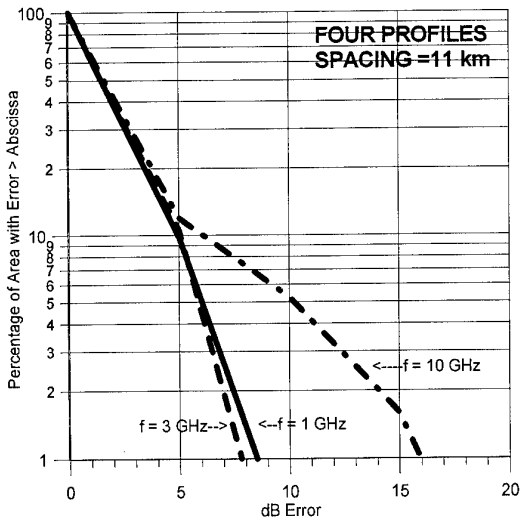


Figure 13: Cumulative distributions of errors at 1 GHz, 3 GHz, and 10 GHz for four profile case with an average spacing of 11 km (coast of Wallops Island).

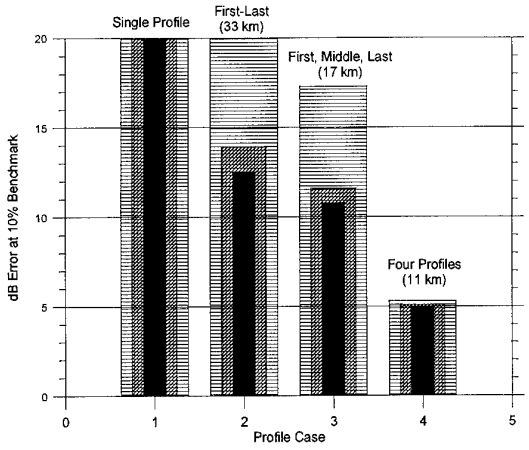


Figure 14: Vertical bar chart depicting errors exceeded over 10% of defined area for various profile cases. The 1 GHz, 3 GHz, and 10 GHz frequency cases correspond to the black, cross hatched, and horizontal line bars, respectively (coast of Wallops Island).

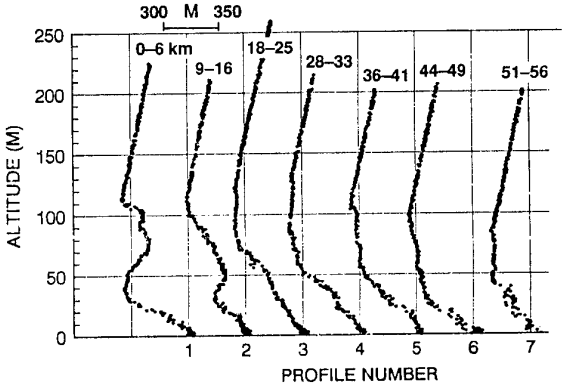


Figure 15: Modified refractive index profiles measured off the coast of San Nicolas Island during the period 09:59 to 10:22 (local time), March 19, 1988. Indicated distance intervals are relative to a simulated ship at sea and are directed towards the shore.

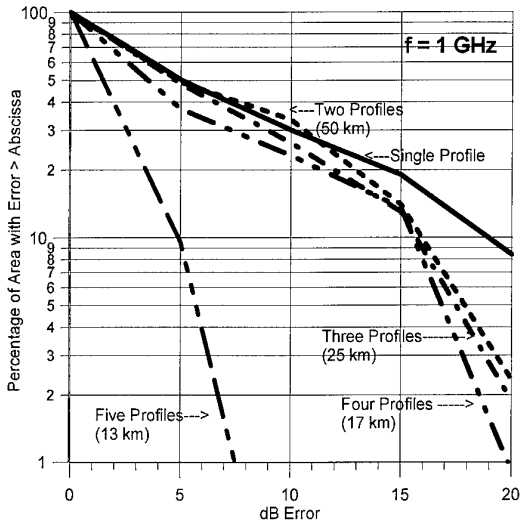


Figure 16: Cumulative distribution of errors at 1 GHz for different profile spacings. Area of consideration is defined by the range interval 15 km to 90 km and height interval 5 m to 300 m (coast of San Nicolas Island).

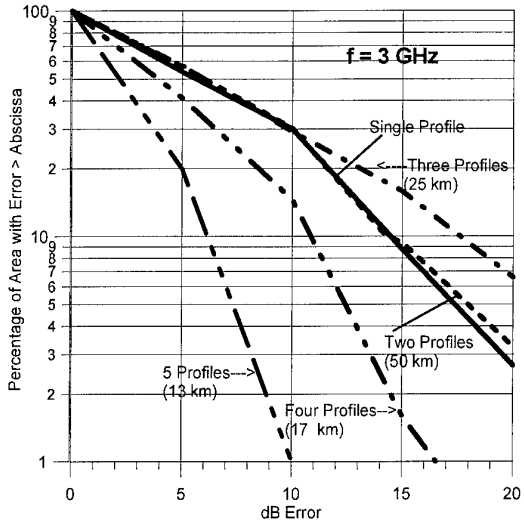


Figure 17: Cumulative distribution of errors at 3 GHz for different profile spacings. Area of consideration is defined by the range interval 15 km to 90 km and height interval 5 m to 300 m (coast of San Nicolas Island)

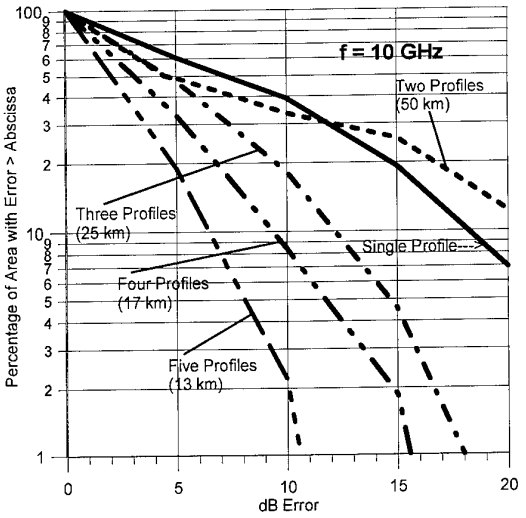


Figure 18: Cumulative distribution of errors at 10 GHz for different profile spacings. Area of consideration is defined by the range interval 15 km to 90 km and height interval 5 m to 300 m (coast of Wallops Island)

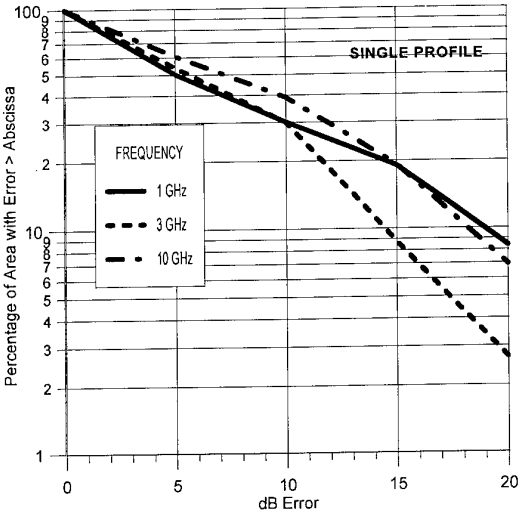


Figure 19: Cumulative distributions of errors at 1 GHz, 3 GHz, and 10 GHz for single profile case (coast of San Nicolas Island).

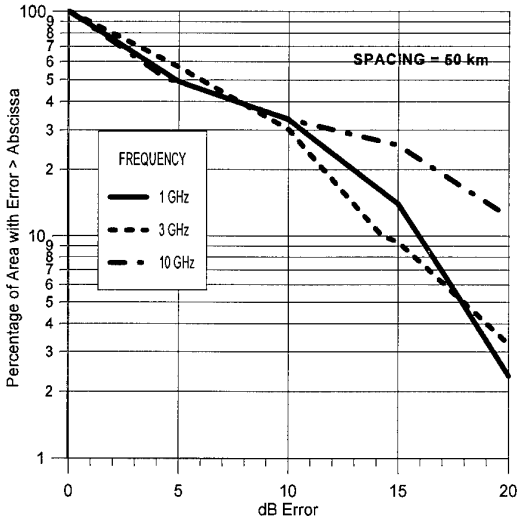


Figure 20: Cumulative distributions of errors at 1 GHz, 3 GHz, and 10 GHz for two profile case with a spacing of 50 km (coast of San Nicolas Island).

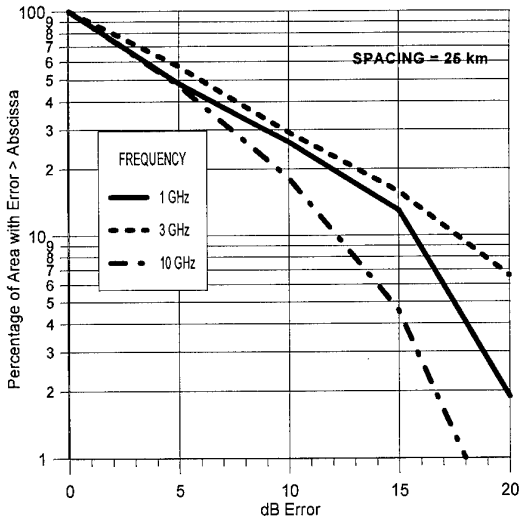


Figure 21: Cumulative distributions of errors at 1 GHz, 3 GHz, and 10 GHz for three profile case (first, middle, and last) with an average profile spacing of 25 km (coast of San Nicolas Island).

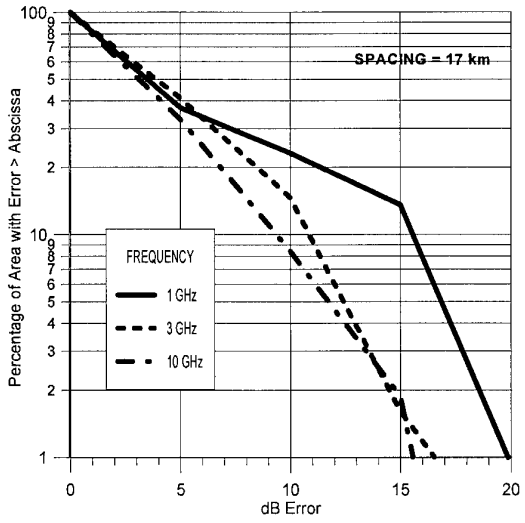


Figure 22: Cumulative distributions of errors at 1 GHz, 3 GHz, and 10 GHz for four profile case with an average spacing of 17 km (coast of San Nicolas Island).

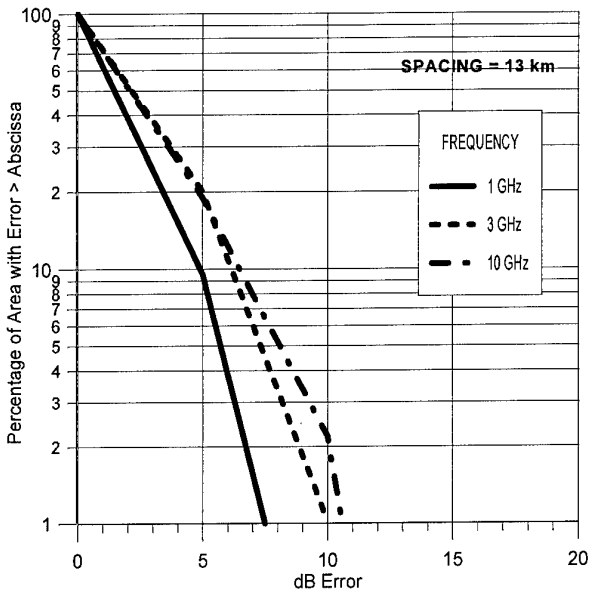


Figure 23: Cumulative distributions of errors at 1 GHz, 3 GHz, and 10 GHz for Five profile case with an average spacing of 13 km (coast of San Nicolas Island).

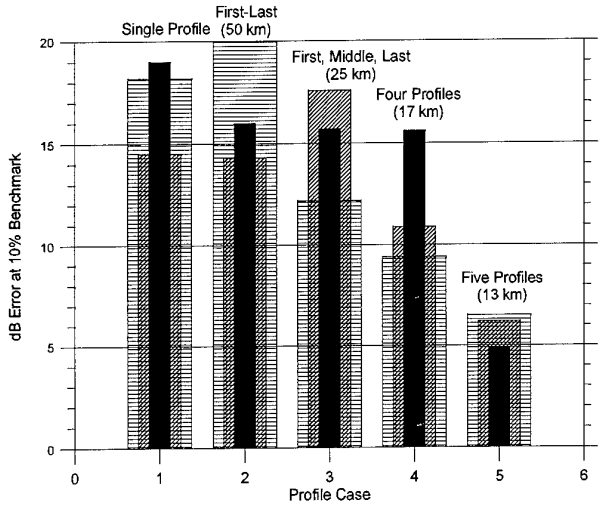


Figure 24: Vertical bar chart depicting errors exceeded over 10% of defined area for various profile cases. The 1 GHz, 3 GHz, and 10 GHz frequency cases correspond to the black, cross hatched, and horizontal line bars, respectively (coast of San Nicolas Island).

Paper 11

L.T. Rogers (US)

Comment # 1:

You are computing the probability that the homogeneous error ϵ_H exceeds some threshold value ϵ_o, conditioned upon the true refractive environment $\mathcal{F}$ being equal to the refractive environment for one of your two sample range varying environments, $\mathcal{F}_1$, or $\mathcal{F}_2$: i.e. you are determining $P(\epsilon_H > \epsilon_o / \mathcal{F} = \mathcal{F}_1)$ and $P(\epsilon_H > \epsilon_o / \mathcal{F} = \mathcal{F}_2)$.

To determine the total probability of ϵ_H exceeding ϵ_o, i.e. to determine $P(\epsilon_H > \epsilon)$, it is necessary to know the probability of the different realizations of the refractive environments, i.e. $P(\mathcal{F}_1)$, $P(\mathcal{F}_2)$ ..... $P(\mathcal{E}_H)$ so as to get:

$$P(\epsilon_H > \epsilon_o) = \sum_i P(\epsilon_H > \epsilon_o / \mathcal{F} = \mathcal{F}_i) P(\mathcal{E}_i)$$

Question # 1:

The major question in the interpretation of your results is just how often instances of horizontal inhomogeneity as extreme as those in your two sample cases actually occur.

Question #2:

Were your sample cases randomly chosen?

Comment #2:

Your sample case from southern California appears to me to be unusually inhomogeneous, based upon my examination of soundings from that area.

Author's reply to questions #1 and #2:

The individual cumulative distributions depicted in the paper only represent the percentage of the designated area of the region for which the errors exceed given levels for *the two given refractive environment cases.* The paper does not state the distributions correspond to absolute probabilites. In the conclusion to my talk I stated that in order to arrive at an absolute statistic a large number of cases must be considered. Of course, these cases must be randomly selected. The two specific cases were not random but were selected as a demonstration of how severe the errors can become assuming a single profile with lateral homogeneity. The second major point made in the paper was a determination of the required profile spacing associated with these two severe cases for reducing the errors to acceptable levels.

Direct Measurement of Microwave Propagation Effects

J. Stapleton
S. Kang
NAVSURFWARCENDIV
CODE F42
17320 DAHLGREN RD
DAHLGREN VA 22448-5100 USA

1. SUMMARY

The detection and tracking of low flying cruise missiles by shipboard radars is dominated by environmental effects. Significant temporal and spatial fluctuations in microwave propagation loss have been observed in this low altitude region. In order to better understand these fluctuations and to support the design and test of future naval radars, two microwave propagation measurement systems were developed which allowed the direct measurement of microwave propagation loss. Both systems concentrated on the low altitude region and made measurements at one frequency versus range and across a wide bandwidth for a fixed range. The test data presented in this paper gives a few examples which illustrate the variety of conditions sensed with these systems.

2. INTRODUCTION

Microwave propagation at low altitudes is strongly affected by the structure of the refractive index of the lower atmosphere. Microwave propagation in this region has long been known to vary with altitude, range, time and frequency. Under subrefractive conditions, the radar horizon may be significantly reduced. Under enhanced propagation conditions (ducting), deep nulls have been both predicted and observed with vertical extents less than 1 meter in altitude, and having varying, and inadequately defined, temporal and spatial characteristics. Testing of developmental sensor systems in this low altitude region is beset with the difficulty of achieving enough test runs over broad environmental conditions to adequately quantify system performance. The problem is compounded by the necessity of testing with surrogate targets, which often requires the scaling of results in altitude and range in this most difficult region. As a general rule, in previous analyses of test data, insight into microwave propagation during the tests was sought by executing computer models that used as inputs refractive index profiles (and surface conditions) calculated from measurements of surface water and air temperate, relative humidity, pressure, and wind speed. These microwave propagation models are unable to reliably predict the short-term dynamics of the environment because of practical limitations in the sample rate and range coverage of the refractive profile inputs. In view of these deficiencies, a task was undertaken to make direct measurements of microwave propagation rapidly and simultaneously over multiple one-way paths in the low altitude region where surface ships must defend against low altitude missile attacks. Initial goals were to instrument in a form suitable for support of planned land-based-sensor tests at the Naval Surface Warfare Center Dahlgren Division, Wallops Island Detachment, both to get greater immediate benefit from the tests and to further the understanding of propagation in the low altitude region for the benefit of future sensor design efforts.

3. APPROACH

Two measurement systems were designed for the purpose of conducting these tests. The first, and simpler, of the systems was called the Single Frequency System (SFS), and was comprised of a single frequency transmitter/receiver pair. The transmitter was mounted on a 10 meter tower and was moved up and down the tower by a computer controlled motor driven shuttle. This transmit tower was mounted on the stern of a boat capable of approximately 10 m/s. The receiver was mounted on the shoreline at heights commensurate with shipboard radar heights. The transmitter produced a single frequency continuous wave signal at 13.95 GHz. The boat was brought into position with a stern aspect approximately 11 km in front of the receiver. From this position, the boat proceeded outbound to approximately 37 km. The system was carefully calibrated so that pathloss or propagation factor relative to freespace propagation could be calculated from the received power levels. Global Positioning Satellite (GPS) data was collected to give the position of the boat during measurements.

The second system, called the Microwave Propagation Measurement System (MPMS), was comprised of 10 transmitters and 4 receivers which allowed the simultaneous measurement of 40 paths with a frequency coverage of 2 to 18 GHz. The transmitters switched between 16 discrete carrier frequencies at a 2 Hz rate which allowed coverage of the 2 to 18 GHz band in 16 discrete steps in 8 seconds. The original measurements were made with the 10 transmitters mounted on shore some 29 km from the 4 receivers which were mounted on an opposing shoreline. The transmitters were attached to a 10 meter tower spaced 1 m apart vertically in order to sample propagation variability with altitude, commensurate with those flown by low altitude missiles. Two of the receivers were mounted at shipboard radar heights and the other two were mounted at low and high extremes of the support structure.

Paper presented at the AGARD SPP Symposium on "Remote Sensing: A Valuable Source of Information", held in Toulouse, France from 22-25 April 1996, and published in CP-582.

4. TEST SITES AND TIME PERIODS

The SFS was used during two test periods at two separate locations. The first test period occurred at the Naval Surface Warfare Center Dahlgren Division in Dahlgren, Virginia during the months of June and July 1993; the boat was operated on the Potomac River Test Range at this site. The second test period occurred at the Naval Surface Warfare Center Dahlgren Division, Wallops Island Detachment on Wallops Island Virginia during March 1994. The boat was operated in the Atlantic off the coast of Wallops Island during these tests.

The MPMS was used between the months of January and April 1994 at the Naval Surface Warfare Center Dahlgren Division Wallops Island Detachment. The data was collected along a radial between Wallops Island, where the receive tower was located, and Parramore Island some 29 km south along Virginia's Atlantic coast, where the transmit tower was located. During the early weeks of this test period, the system was undergoing calibration and check-out; therefore, the highest quality data were collected during the months of March and April. The SFS was also operational during this test period, as stated above.

5. EXAMPLE MEASUREMENTS - SFS

Results of the data collected using the SFS will be discussed first. Several notable features were measured with this system and are discussed in the following sections.

5.1 Horizontal Fade

A phenomenon sometimes referred to as the "horizontal fade" has been both produced using propagation models (Ref 1) and measured experimentally. During testing conducted using the SFS at the NSWCDD Potomac River test location, clear evidence of the effect can be seen in the received power versus height and range data. Figure 1 illustrates the data in a coverage diagram format with propagation factor relative to freespace shown using a gray scale. The dark region in the upper subplot at approximately 2.4 m in altitude depicts the "horizontal fade" region. The lower subplot in Figure 1 shows propagation model output (The model PCPEM (Ref 2) was used for all modeled data in this report.) for a 28 m evaporation duct. Figure 2 shows the refractive profile used to model this duct. The modeled data also indicates the "horizontal fade" at the same approximate location. The "horizontal fade" was present in the data collected on several occasions. An attempt is being made at this time to develop a percent occurrence for this phenomenon using a larger data set collected with the MPMS.

5.2 Other Multi-Modal Propagation

Evidence of a surface based duct is reflected by the data shown in Figure 3. This data was collected at the Wallops Island test location. The lower subplot shows propagation model output generated from refractive profiles collected by the Johns/Hopkins University Applied Physics Laboratory meteorological measurement helicopter (Ref 3). The refractive data was collected along the same radial the boat traveled during the data collection and at approximately the same time; several profiles were measured over the range covered. Figure 4 shows the refractive data which was used as input to the propagation model. The refractive profiles shown in Figure 4 were generated using a program called LARRI (Large-Scale Atmospheric Refractivity Range Interpolator) which smooths and prepares measured meteorological data for use by propagation models. LARRI was developed by the Johns-Hopkins University Applied Physics Laboratory for use with their helicopter based meteorological measurements. The character of the modeled data is representative of the measured propagation factor data shown in the upper subplot; however, the model overall tends to reflect a stronger ducting environment than was measured by the microwave system. The reason for this discrepancy is not known, but it does point out that even with good temporal and spatial correlation between meteorological data collection and microwave measurements, there can be significant differences in the propagation factors generated from each.

5.3 Subrefraction

On several occasions the SFS data collected at the Wallops Island location indicated a subrefractive environment. Figure 5 shows an example, and the lower subplot shows the modeled propagation factor using another set of refractive profiles collected by the helicopter. The refractive profiles are shown in Figure 6. The black portions of the upper subplot in Figure 5 are areas where the boat was not keeping its alignment with the receiver well enough that the antenna beam pattern losses could be considered less than 1 dB. Since the boat's heading was collected using a digital compass and the boat's bearing was available from the GPS data, a filter could be used on the data to reject any measurements made when the antenna pattern loss was expected to be worse than 1 dB. The modeled and measured data match reasonably well in this case. The refractive data was collected versus range on nearly the same radial the boat traveled and at nearly the same time.

5.4 SFS Data Summary

A wide variety of conditions have been sampled using this system to directly measure the microwave propagation versus range over the low altitude region. The "horizontal fade" was measured on two days at NSWCDD Potomac River test location and perhaps once to a lesser degree at Wallops Island. Three other days at the Potomac River test location also showed some evidence of the "horizontal fade." Other evidence of multi-mode propagation was also measured during periods where surfaced based ducts seemed to dominate the propagation environment. During these periods, a high degree of variability in the propagation factor versus range and altitude was sensed by the system. A total of 17 test days were covered at the two test locations, during two seasons of the year (6 of the test days were at the Potomac River Test Range, and 11 were at Wallops Island). Some degree of ducting or enhanced propagation was exhibited during all but three of the test days. On these three days, a subrefractive environment was

indicated. All of the subrefractive days occurred at Wallops Island. The data set is not large enough to draw statistical conclusions about the percent occurrence of ducting versus subrefraction. The fact that a wide variety of different refractive environments were sensed indicates the dynamic nature of the propagation environment.

In general, the model predictions made using the best refractive profile data with respect to time and radial coincidence with the microwave measurements, yielded reasonable agreement at least with regard to trends. In the case of some of the more complicated refractive environments, exact matches seem out of reach, perhaps due to the temporal variability present in the refractive environment.

6. EXAMPLE MEASUREMENTS - MPMS

Data from the MPMS was collected, as stated earlier, for a fixed range over a water path along the coast of Virginia. This path was parallel with the coastline and therefore subject to the land-sea boundary over its entire length as opposed to a path which traverses this boundary and proceeds to open water (a path perpendicular to the coastline). While a coastal path of this type is of interest to the US Navy when considering the littoral environment, it is recognized that a path parallel to the coast would be subject to a significantly different refractive environment than a path perpendicular to the coast.

6.1 Temporal Variability

Both a long and a short term temporal variability were observed during the data collection period. The long term variability was marked primarily by transitions from ducting to subrefraction and vice versa. Figure 7 shows an example of the system output during one such transition for one receiver and 4 of the 16 frequencies used. The 4 horizontal bands in Figure 7 depict the measured propagation factor at the 4 frequencies. The labels at the left of the graph indicate the receiver and frequency by number, and the legend at the bottom gives the values for each of the frequencies and the height of each receiver. The transmitter heights are covered from lowest to highest in each of the 4 bands from bottom to top. The variable width vertical lines in medium gray denoted periods when the system was not collecting data. The larger width stripe near 0600 hours is the longest period in the data shown where the system was offline. The horizontal time axis is noted at the bottom. Approximately 2 hours pass during the transition between ducting and subrefraction as seen using the gray scale at the bottom of the graph. Values larger than 10 dB relative to freespace were measured for some frequency and height combinations indicating enhanced propagation in the earlier time period shown in the Figure. This environmental condition gradually evolved to a strongly subrefractive situation versus transmitter height and time as indicated by the gradual change to darker gray and finally black. Upon examination of the data before and after this transition, it was discovered that the general ducting or subrefractive conditions lasted over 24 hours. The other 3 receivers and 12 frequencies behaved in a similar fashion versus time following predicted trends for the differing geometries and wavelengths. This type of transition is typical of the long term temporal variability evident in the data set.

The short term variability occurred primarily during periods when a ducting environment was indicated. Figure 8 shows a display created by the MPMS in which 4 subplots are used, one for each of the 4 receivers. The data key to the right of the subplots gives the receiver number and frequency number used in each subplot; the data key list is given for the subplots from left to right. The legend at the left of the graph shows the height in feet for each of the receivers by number and the gigahertz frequency value for each of the frequency numbers. One-way propagation factor versus transmitter height is shown in each subplot. (Please note that in Figures 8 through 10 the vertical units are in feet not meters.) In the case of Figure 8, frequency 2 was selected which corresponds to 3.6 GHz. The cluster of lines were created by over plotting 8 minutes of data; since each frequency was revisited every 8 seconds, 60 samples are displayed. Receiver 1 (which was at an altitude of 5.7 m) shows the greatest amount of variability over the 8 minute interval. About 30 dB of variability is evident at 3 m. This is a large amount of change over a short period of time especially considering that the frequency is relatively low. The higher altitude receivers do not experience quite as much variability and the amount of variability differs versus transmitter height as well. Figure 9 shows the same type of plot in all respects including the time interval covered except that the data was collected at frequency 7 or 8.475 GHz. Again receiver 1 shows the most variability, but all receivers show more variability at this frequency than at 3.6 GHz. There are more nulls evident at this frequency as well, and overall the short term variability is more pronounced at this frequency than at the previous. Figure 10 shows the same time period again but for frequency 14 (15 GHz). The trend established with increasing frequency is supported again here. More null structure is evident in Figure 10 than in the previous two figures and a strong degree of short term variability is present, more than 30 dB at some transmitter heights. The fact that in all three figures the propagation factor was centered around freespace indicates that some amount of ducting was present during the data collection. No supporting meteorological data was available at the time of the data collection since it occurred at about 2 a.m. local time (The MPMS was designed to collect data continuously around-the-clock unmanned and was able to collect a large amount of data during periods not typically covered by other test events.)

6.2 Height-Gain Null Position Versus Frequency

One important feature not clearly evident in the previous three Figures is that the placement of the nulls versus transmitter height was shifted versus time and frequency. In fact, the nulls do not occur at the same height at the same time for frequencies sufficiently spaced in the 2 to 16 GHz band. Previous studies (Ref 1) have illustrated using propagation models that sufficient bandwidth can be used to fill nulls in the propagation structure versus both height

and range. There is some debate as to the percentage bandwidth necessary to perform this job adequately. The data collected by the MPMS shows that this effect does occur, and the data could be used to help define the bandwidth necessary to fill nulls at least for the environments encountered during data collection. While this analysis has not yet occurred, it is a follow on study planned for the future use of the MPMS data.

6.3 MPMS Data Summary

A wide range of conditions were encountered over the test period at Wallops Island. While this region of the world is perhaps more dynamic in some respects than others, it is representative of a littoral environment in which the US Navy is interested in operating. A large amount of both short and long term temporal variability as well as fairly fine scale vertical variability was measured. The long term variability is fairly well understood meteorologically (Ref 4); however the meteorological changes necessary to produce some of the significant short term variability experienced is being investigated.

7. CONCLUSIONS

The data discussed in this report has shown that the propagation environment in the low flyer region is highly dynamic, giving rise to 10's of dB variation in signal level. These variations can occur on both a long and short term basis. The long term variability occurred over several hours with more than 40 dB change in one-way propagation factor over that period. The short term variability occurred over 10's of seconds with 15 dB change in one-way propagation factor at some altitudes. The height-gain structure of multi-modal propagation lobes was also spatially dynamic during these short term variations. The placement of a null in the height-gain profile moved 2 m vertically on a similar time scale.

The comparison of the directly measured propagation with modeled propagation calculated using simultaneously collected in-situ meteorological measurements was favorable in general. However, it is evident that some of the short term temporal and spatial variability cannot be predicted by a propagation model operating from a set of meteorological data collected no more frequently than once every half hour. The models do appear sensitive to having temporally and spatially coincident data for good comparison with directly measured pathloss. Even when care is taken in this regard, exact matches are rare, and 5 to 10 dB differences are likely.

The data discussed has been and continues to be a resource used in the following activities: ground truth measurements for radar test observations, propagation model validation, meteorological measurement requirement assessment, support data for radar design trade-off studies, and remote sensing in terms of the development of refractivity inversion techniques. Several on-going efforts are making use of the data as a basis for analysis of radar performance and another paper in these proceedings entitled, "Using Remote Refractivity Sensing to Predict Tropospheric Refractivity from Measurements of Microwave Propagation," by D. Boyer, et. al., discusses use of the data as a driver for a refractivity inversion technique.

8. REFERENCES

1. Dockery, G. D., "Propagation Fade Characteristics in Low-Altitude Surface Ducts," The Johns Hopkins University Applied Physics Laboratory, Technical Report, NATO AAW Task 3-1-19, Radar Performance Analysis, October 20 1988.

2. PCPEM [Personal Computer Parabolic Equation Model] © 1989, 1990 of Signal Science Limited.

3. Rowland, J. R., and Babin, S. M., "Fine-Scale Measurements of Microwave Refractivity Profiles with Helicopter and Low-Cost Rocket Probes," Johns Hopkins APL Technical Digest, Volume 8, Number 4 (1987).

4. Goldhirsh, J. and Dockery, G. D., "Propagation Measurements and Modeling at C band for Over-the-Water, Line-of-Sight Propagation Links in the Mid-Atlantic Coast," Radio Science, Vol. 26, Number 3, May-June 1991.

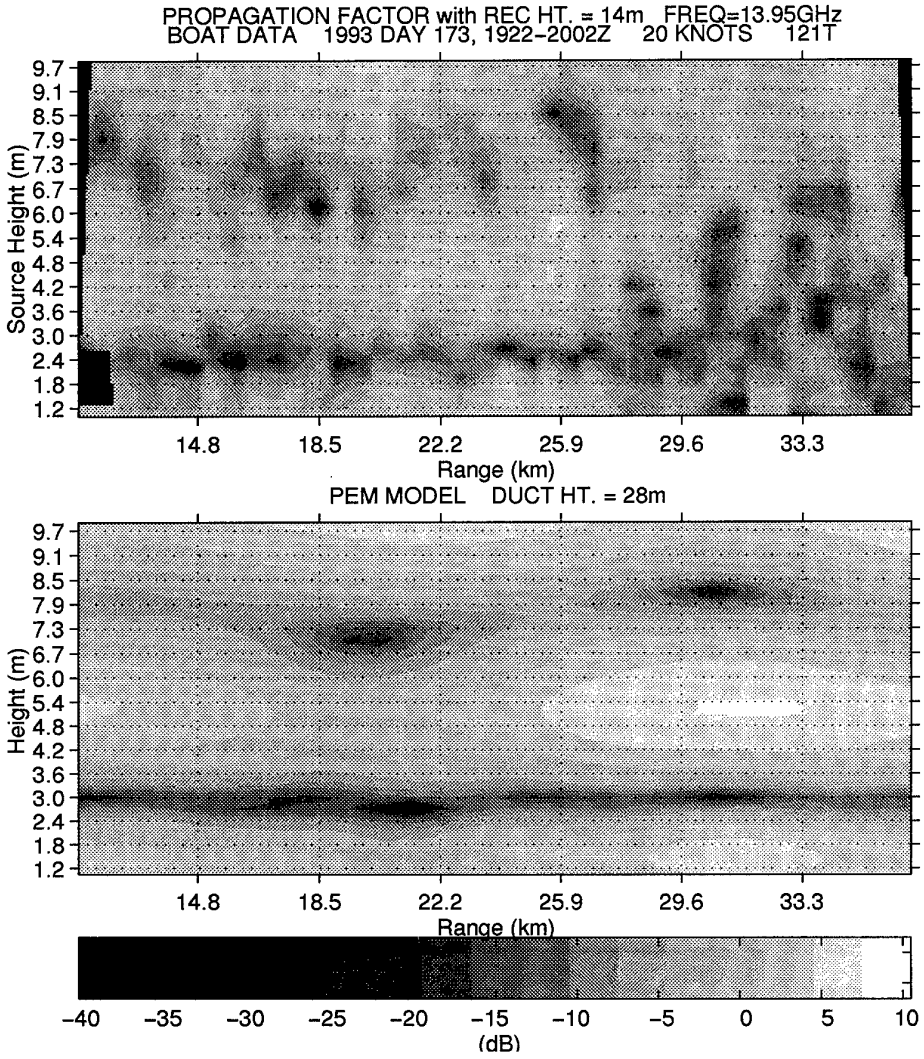


Figure 1. Horizontal Fade

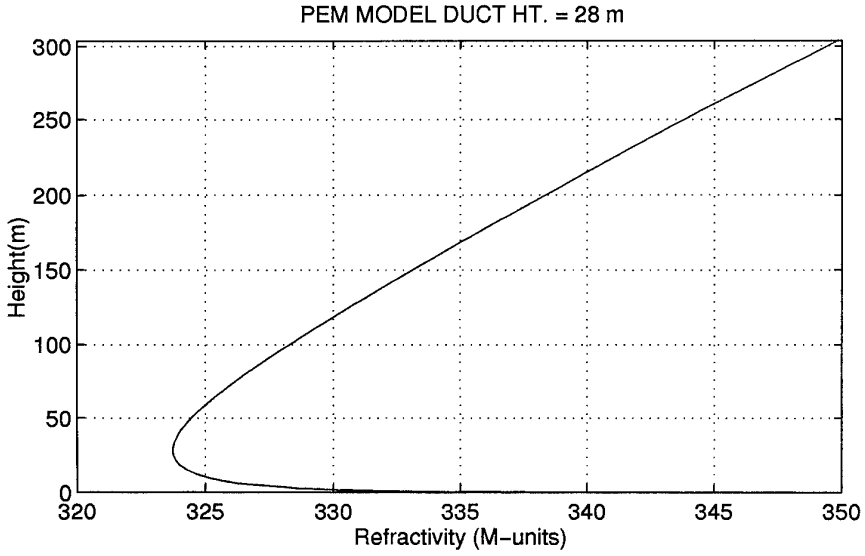


Figure 2. Refractive Profile for a 28 m Evaporative Duct

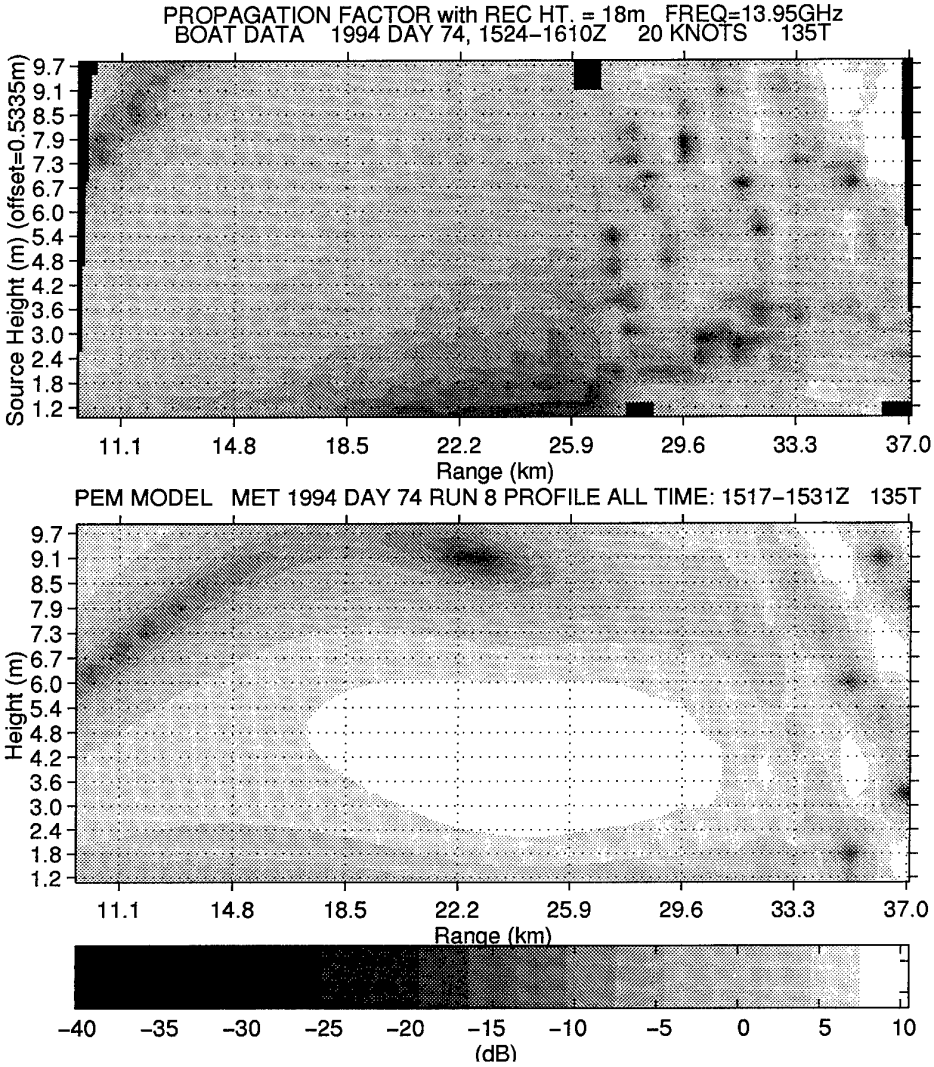


Figure 3. Propagation Due to a Surface Based Duct

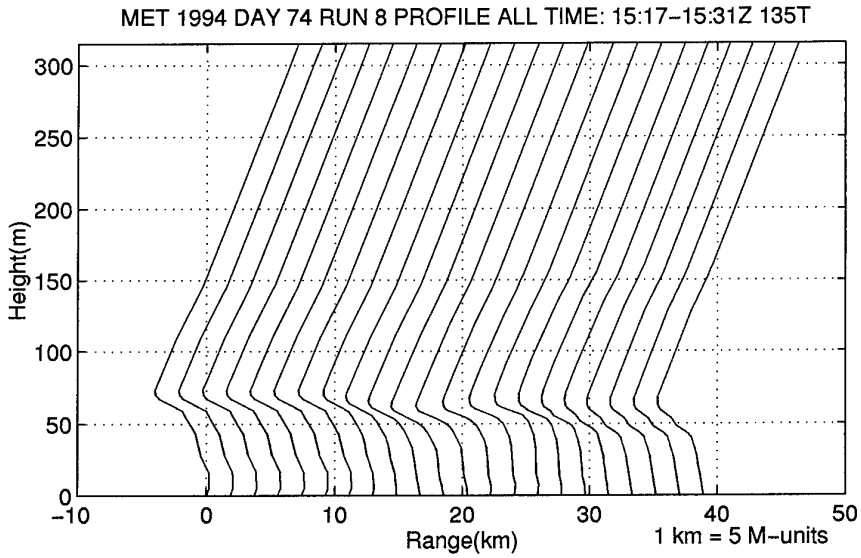


Figure 4. Measured Refractive Profiles Versus Range

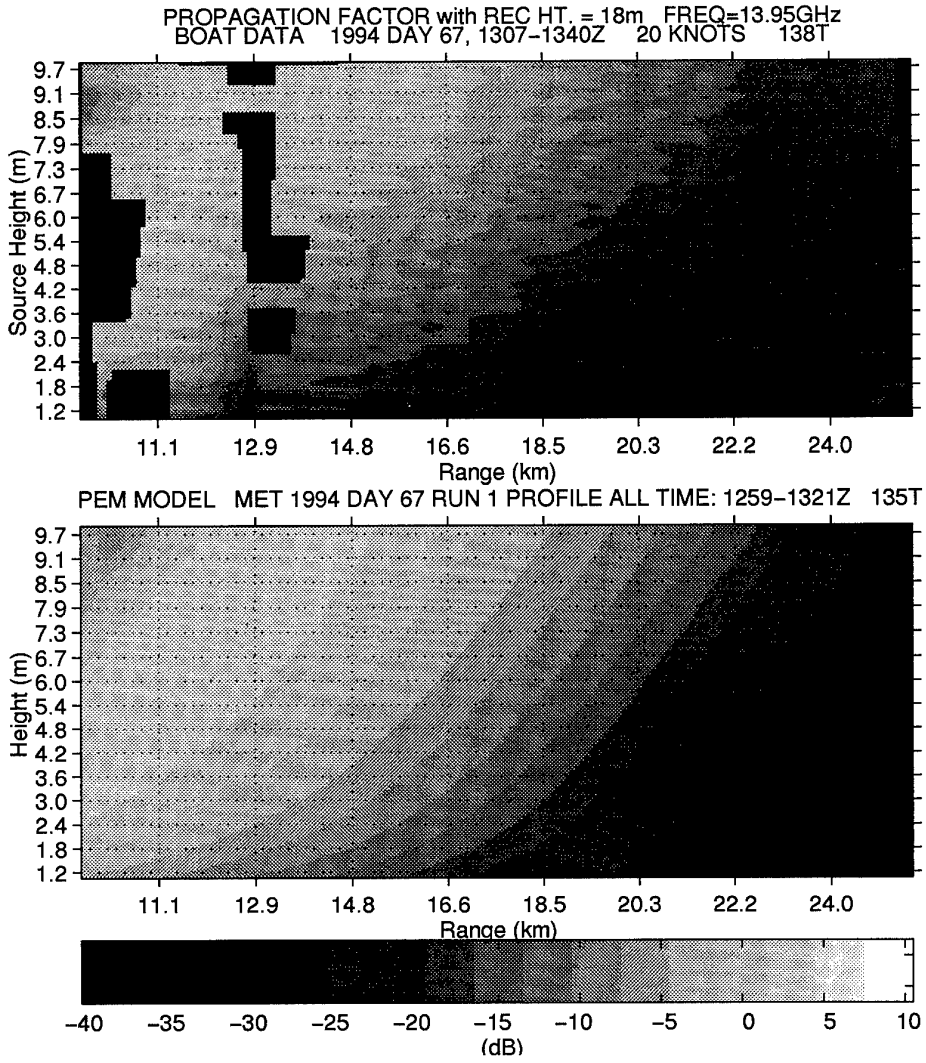


Figure 5. Subrefraction

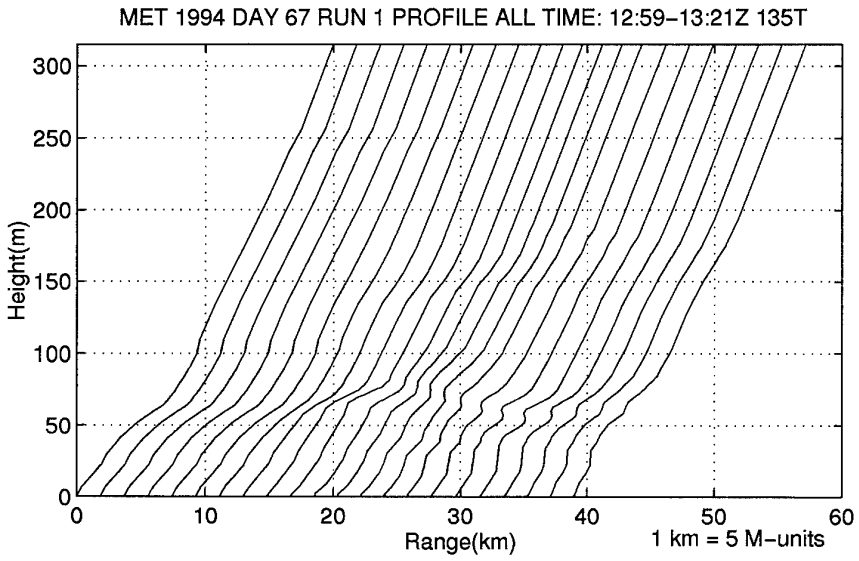


Figure 6. Measured Refractive Profiles Versus Range for a Subrefractive Environment

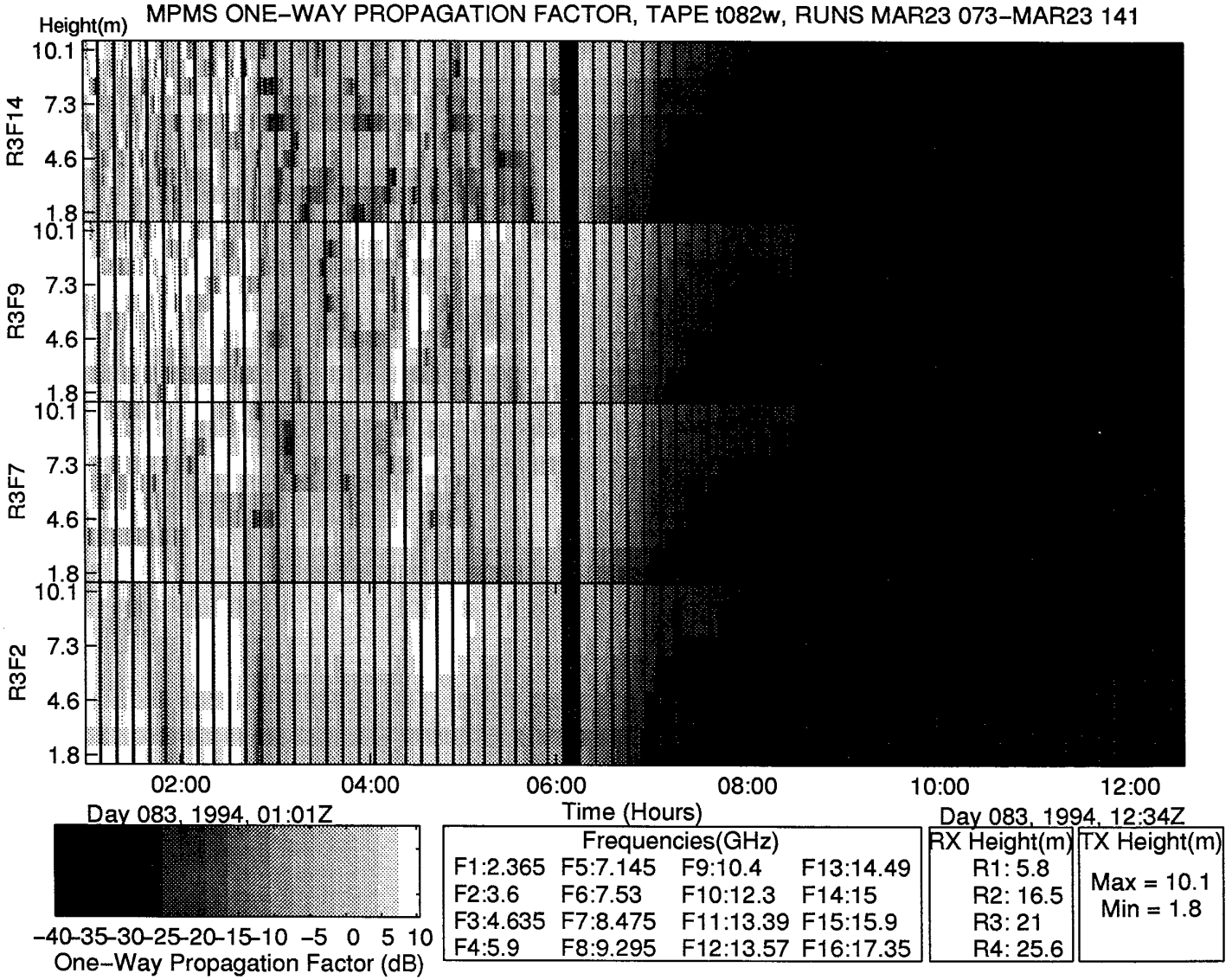


Figure 7. MPMS Measured Long Term Variability

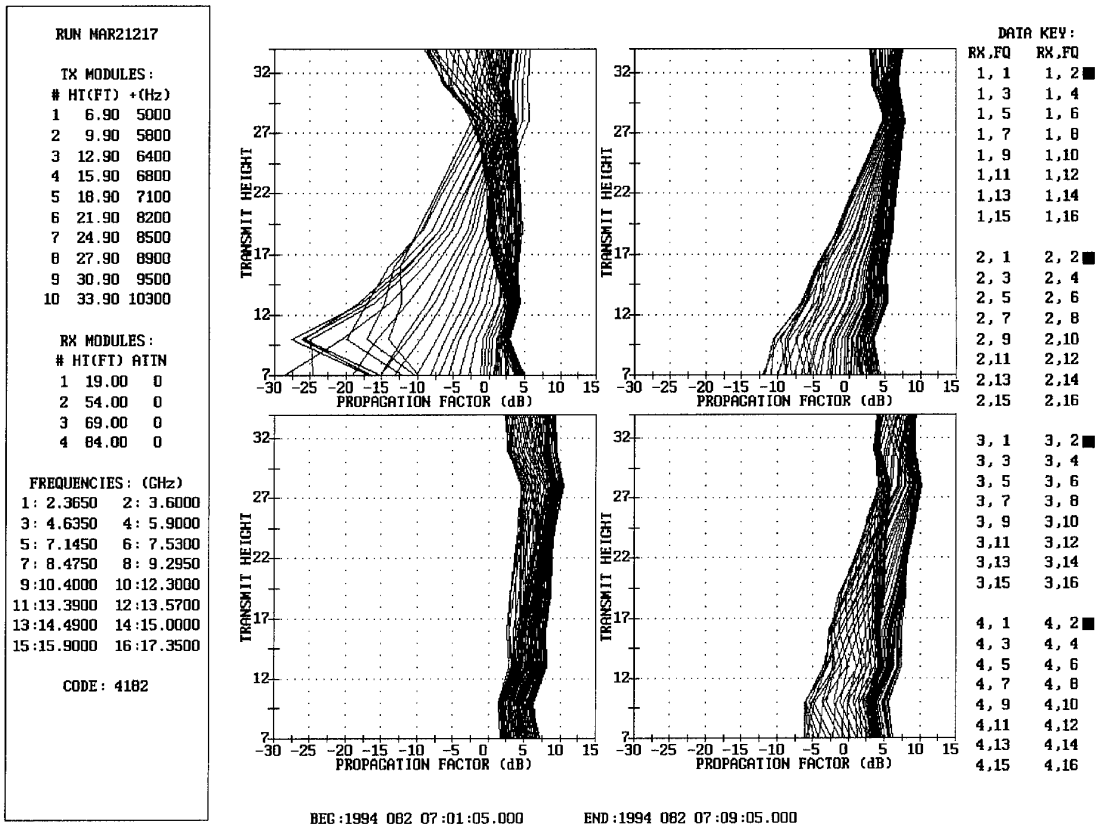


Figure 8. MPMS Measured Short Term Variability at 3.6 GHz

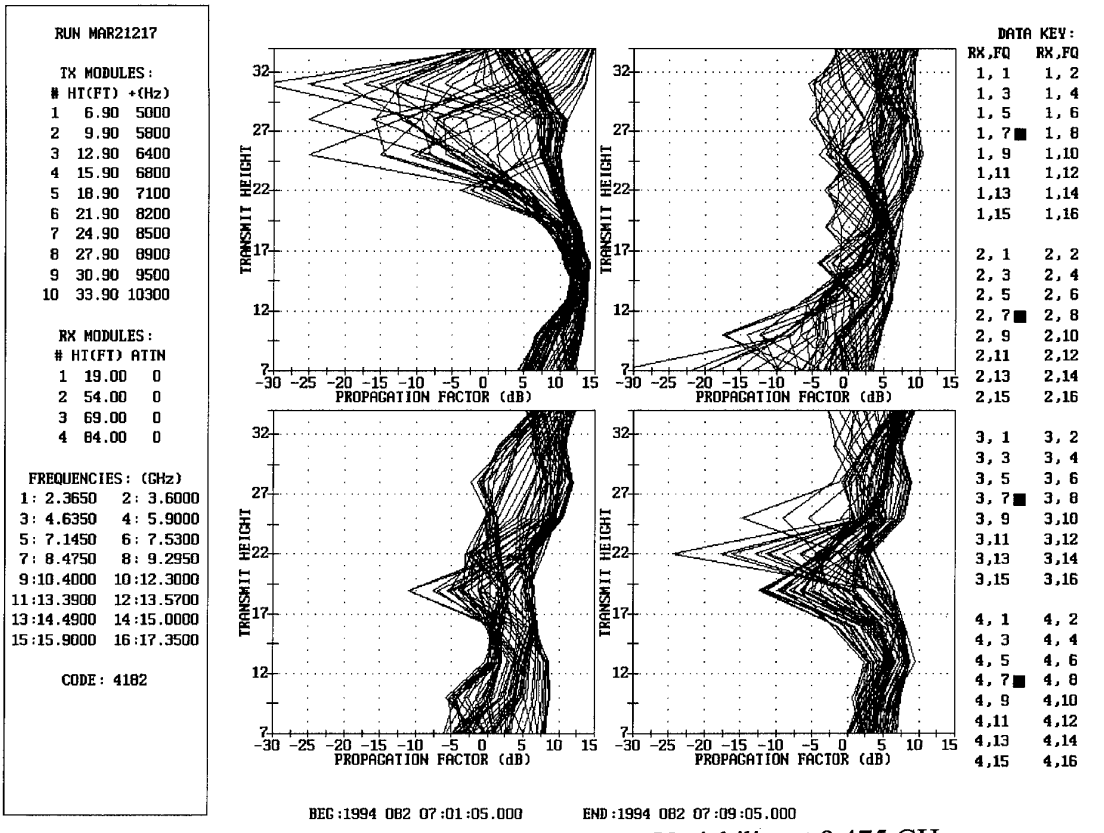


Figure 9. MPMS Measured Short Term Variability at 8.475 GHz

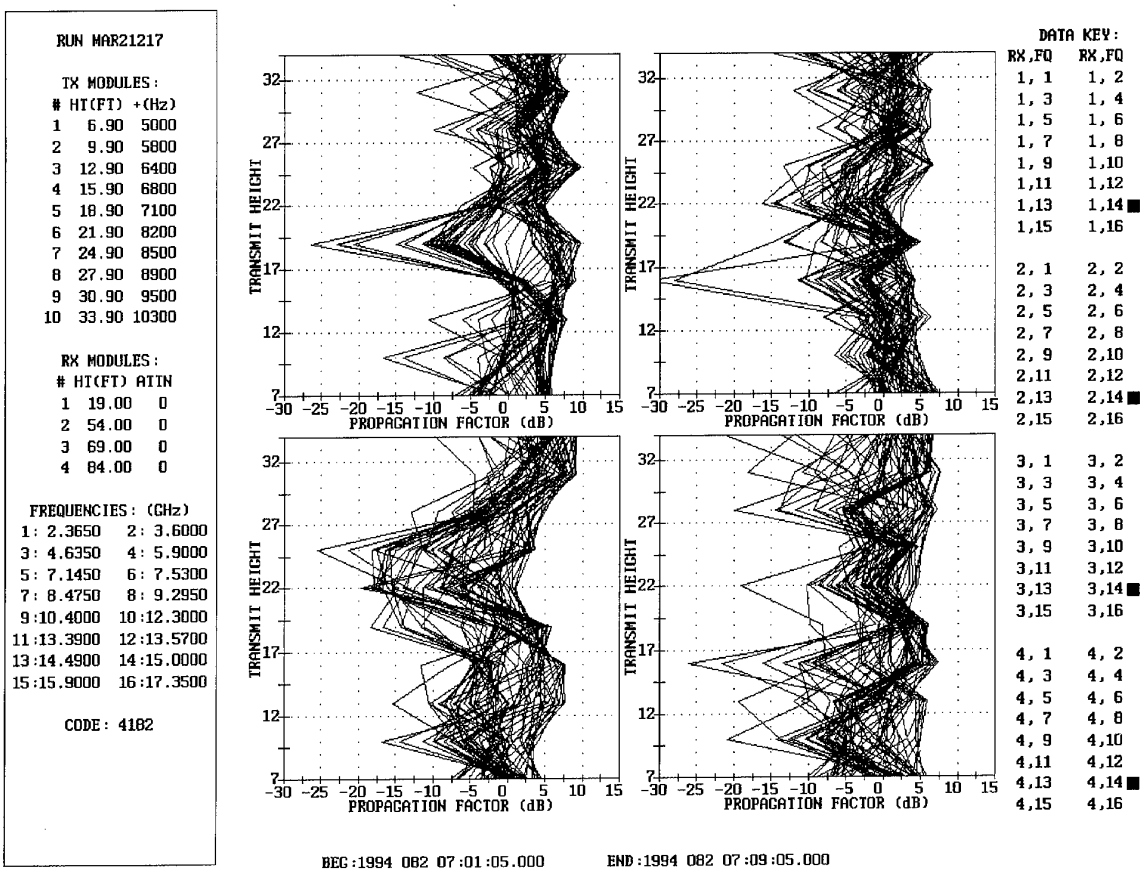


Figure 10. MPMS Measured Short Term Variability at 15.0 GHz

Paper 12

G.S. Brown (US)

What polarization did you use and do you have the capability to change during an experiment?

Author's reply:

The measurements were made using vertical polarization. The system has the capability to collect data in either horizontal or vertical polarization; however, the system rarely collected data while horizontally polarized.

U. Lammers (US)

How can you measure refractivity profiles with a helicopter without perturbing them by the rotor downwash?

Author's reply:

The range-dependent refractive profiles were measured by JHU/APL using their helicopter based meteorological sensors. JHU/APL has been performing these measurements for a number of years and have shown that as long as the helicopter's forward air speed is kept above a minimum value and as long as data is only collected during the descending legs of the helicopter's flight path, the problem of aircraft induced perturbations of the atmospheric parameters is minimal.

Modelling of rough surface effects with the matched transform PE

M.F. Levy
Rutherford Appleton Laboratory
Chilton, Didcot
Oxfordshire OX11 0QX
U.K.

SUMMARY

Generalised impedance boundary conditions allow the rigorous modelling of propagation over surfaces with angle-dependent reflection coefficients. We present a split-step PE implementation based on a matched transform approach. The method is applied to propagation over the rough sea in ducting conditions and to ground-wave calculations.

1. INTRODUCTION

Parabolic equation techniques have proved their usefulness for solving wave propagation problems which had long been thought intractable. With increasing demand for sophisticated PE methods, there has been much interest in the development of better boundary modelling techniques, in particular for Fourier/split-step implementations of the PE. The problem of finding a transform capable of modelling finite impedance boundary conditions was solved in [2,4] with the discrete mixed transform. Application of the discrete mixed transform approach to propagation over a rough surface requires estimation of local angles of incidence on the surface by other means.

Here we present an approach based on non-local (or generalized impedance) boundary conditions, which avoids the explicit calculation of angles of incidence. Basically the field incident on the horizontal boundary of interest is decomposed into plane wave components, and a rigorous boundary condition is derived in angle-space. Much work has been published on transparent boundary conditions based on the non-local approach [1,5,6,7,8,10]. More recently these ideas have been applied to acoustic wave propagation over an elastic bottom [10,11], and to propagation over a rough surface [11]. Our matched transform approach is a split-step Fourier version which is well-suited to electromagnetic wave applications. For smooth finite-impedance surfaces, the matched transform approach is very similar to the mixed transform of [2]. More generally it provides a rigorous solution for propagation over rough surfaces with angle-dependent reflection coefficients. One of the main applications is microwave propagation over the rough sea surface in the presence of ducting, a problem that is difficult to treat with conventional PE methods.

2. GENERALISED IMPEDANCE

We start with the standard parabolic equation

$$\frac{\partial^2 u}{\partial z^2} + 2ik\frac{\partial u}{\partial x} + k^2(m^2 - 1)u = 0 \tag{1}$$

where u is a reduced field component, k is the free-space wave number and m is the modified refractive index. We can assume without loss of generality that m is 1 at the surface. In what follows, we also assume that the initial field is zero at the surface. This is not a restriction for microwave applications, for which the aperture field is normally negligible near the surface. A slightly different approach is needed for ground wave applications, as outlined in the next section.

We define the one-sided Fourier transform in range x by

$$U(s,z) = \int_0^\infty u(x,z)\exp(2i\pi sx)dx \tag{2}$$

We introduce a thin homogeneous layer of modified refractive index 1 between heights 0 and ε. This layer is only needed temporarily for the derivation. Inside the layer, we assume that the initial field is zero. Then the one-sided Fourier transform of u satifies the following equation inside the layer:

$$\frac{\partial^2 U}{\partial z^2}(s,z) + 4\pi ksU(s,z) = 0 \tag{3}$$

In the decomposition of the field into plane wave components with respect to the horizontal, the spectral variable s is related to propagation angle α from the horizontal by

$$s = \frac{k}{4\pi}\sin^2\alpha \tag{4}$$

Equation (3) corresponds to the classical problem of plane wave incidence over a planar interface. The solution is given by

$$U(s,z) = C(s)\left\{\exp(-iz\sqrt{4\pi ks} + R(s)\exp(iz\sqrt{4\pi ks}\right\} \tag{5}$$

where the reflection coefficient $R(s)$corresponds to the angle of incidence defined by (4). Differentiating at the surface, we obtain

$$U(s,0) = \frac{i}{\sqrt{4\pi ks}}\frac{1+R(s)}{1-R(s)}\frac{\partial U}{\partial z}(s,0) \tag{6}$$

Inverting (6), we get the generalized impedance boundary condition (GIBC)

$$u(x,0) = \int_0^x \frac{\partial u}{\partial z}(t,0)w(x-t)dt \tag{7}$$

where the kernel w is the inverse Fourier transform of the impedance function

$$W(s) = \frac{i}{\sqrt{4\pi ks}}\frac{1+R(s)}{1-R(s)} \tag{8}$$

When the reflection coefficient is zero for all angles, this boils down to the well-known transparent non local boundary condition for a homogeneous medium [1,5,6,7,8,10]. More generally, the impedance function can model any surface with angle-dependent reflection properties.

Paper presented at the AGARD SPP Symposium on "Remote Sensing: A Valuable Source of Information", held in Toulouse, France from 22-25 April 1996, and published in CP-582.

3. MATCHED TRANSFORM APPROACH

It is straightforward to include the GIBC (7) in finite-difference implementations of the PE. However a split-step version is highly desirable since split-step algorithms are more efficient for many EM applications. The idea is to replace the field u with a new function v which matches the boundary condition. We put

$$v(x.z) = u(x,z) - \int_0^x \frac{\partial u}{\partial z}(t,z)w(x-t)dt \tag{9}$$

Then if we assume a constant refractive index between one range step and the next (split-step approximation), v satisfies

$$\frac{\partial^2 v}{\partial z^2} + 2ik\frac{\partial v}{\partial x} + 2ik\frac{\partial u}{\partial z}(0,z)w(x) = 0 \tag{10}$$

Note that v is equal to u on the initial vertical. Since v is zero on the surface, the solution can be marched in range with a sine transform. One possibility is to use the following sequence of operations: first do a sine transform in height of the matched function on the current vertical.

$$\hat{v}(x,p) = S\{v(x,z)\} = \int_0^\infty v(x,z)\sin(2\pi pz)dz \tag{11}$$

Then propagate in angle space, including the initial field derivative term.

$$\hat{v}(x+\Delta x,p) = \exp\left(-2i\pi^2 p^2 \Delta x / k\right)\left(\hat{v}(x,p) + \varphi(x,p)\right) \tag{12}$$

The initial field derivative term is given by

$$\varphi(x,p) = -2\pi p\int_0^\infty u(0,z)\cos(2\pi pz)dz \times \int_0^{\Delta x} w(x+t)e^{2i\pi^2 p^2 t/k}dt \tag{13}$$

Then return to z-space with a sine transform, retrieve the field from the matched function and multiply by the refractive index propagator

$$u(x+\Delta x,z) = \exp\left(ik(m^2-1)\Delta x/2\right)u(x,z) \tag{14}$$

Finally, construct the matched function v at the next step using equation (9). Note that in vacuum v can be marched in the whole domain, and u retrieved at the end of the calculation.

In order to retrieve the field from the matched function, the integral equation (9) must be solved. This can be done with a finite-difference scheme. Alternatively we may consider the one-sided Fourier transform of equation (9) in range.

$$V(s,z) = U(s,z) - W(s)\frac{\partial U}{\partial z}(s,z) \tag{15}$$

This can be solved for U with either a simple finite-difference scheme or with an implicit method similar to that of [2]. The results are transformed back to obtain u.

Figure 1 shows an application of the matched transform technique to the modelling of transparent boundaries: a 3 GHz source radiates in vacuum. An absorbing layer has been added at the top of the domain, while the bottom is modelled with a split-step GIBC. It is seen that the GIBC works perfectly, absorbing all the downwards propagating energy.

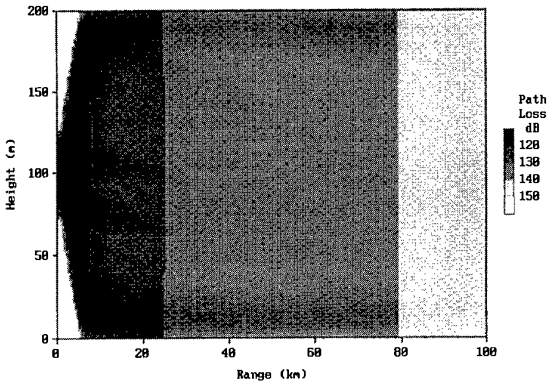


Figure 1. S-band source in vacuum with GIBC on bottom boundary

We now apply the technique to X-band propagation over the rough sea. The example given here has been treated in [2] with a discrete mixed transform method, estimating local angles of incidence with a MUSIC algorithm. The 10 GHz source is vertically polarised and located 25 m above mean sea level. The refractive index structure consists of a strong surface duct (12 M-units lapse in the first 50 m, standard slope of 118 M-units/km above 50 m). The electrical constants of the sea are taken to be $\varepsilon_r = 80$, $\sigma = 4\,\mathrm{S/m}$.

Figure 2 shows the coverage diagram for a smooth sea: the duct couples with the sea surface to produce multiple bounces.

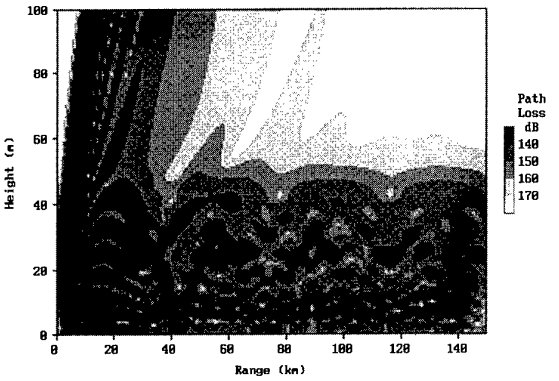


Figure 2. Coverage diagram of X-band antenna in surface duct with smooth sea

Figure 3 shows propagation factor against range for a receiver at 25 m height for the smooth sea case, and for a wind speed of 14 m/s which corresponds to a r.m.s. wave height of 1 m. The function W is calculated from the Miller-Brown roughness model [9]. Its inverse one-sided Fourier transform w is calculated with an Fast Fourier Transform.

Agreement with [2] is excellent, showing that in this case the boundary condition based on the local incidence angles obtained with MUSIC works very well. Interaction of the strong ducting environment with the rough surface produces a large loss of energy compared with the smooth sea case, demonstrating the importance of accurate roughness modelling.

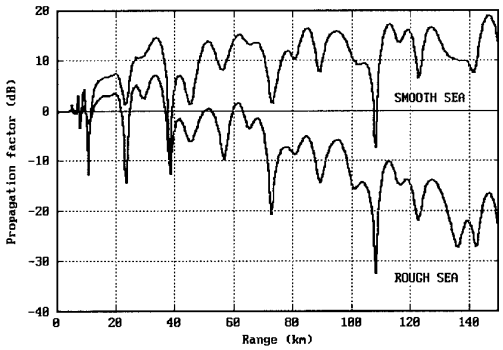


Figure 3. Propagation factor for receiver at 25 m height for smooth sea (rms wave height 0 m) and rough sea (rms wave height 1 m)

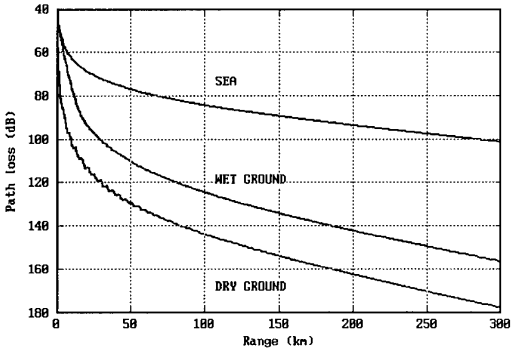


Figure 4. Ground wave curves at 3 MHz

4. SURFACE WAVE

The approach described above has to be modified for surface wave applications. We limit the discussion to grazing angle incidence, an approximation that is justified at the frequencies of interest for surface wave propagation. The boundary condition at the ground is then

$$\frac{\partial u}{\partial z}(x,0) = ik\delta u(x,0) \tag{16}$$

where δ is the surface impedance for grazing incidence. The matched function v is now simply

$$v(x,z) = u(x,z) - \frac{1}{ik\delta}\frac{\partial u}{\partial z}(x,z) \tag{17}$$

The matched function v is marched in range with a sine transform, and u is recovered by solving the ordinary differential equation (17) at each range with a finite-difference scheme. The method is very similar to the discrete mixed transform algorithm described in [2], but here we solve for u directly instead of using the inverse discrete mixed transform of [2]. Compared to the perfectly conducting surface case there is only a slight overhead in computation time in solving for u at each range.

Figure 4 shows path loss results for a 3 MHz vertically polarised source at ground level and a receiver at ground level. Three curves are shown: propagation over the sea ($\varepsilon_r = 70$, $\sigma = 5\,\mathrm{S/m}$), over wet ground ($\varepsilon_r = 30$, $\sigma = 10^{-2}\,\mathrm{S/m}$) and over dry ground ($\varepsilon_r = 7$, $\sigma = 3\times10^{-4}\,\mathrm{S/m}$). Agreement with the ITU ground-wave curves [3] is excellent.

5. CONCLUSIONS

Generalised impedance boundary conditions provide a rigorous solution for propagation over a rough surface which can be described in terms of reflection coefficients for plane wave incidence. The main advantage of the GIBC formulation is that because of its non-local nature, it does not require estimation of incidence angles on the surface. The matched transform approach gives a robust split-step PE implementation of GIBCs. We have demonstrated its usefulness for treating microwave ducting over the rough sea surface and for surface wave applications.

6. ACKNOWLEDGMENT

This work has been funded as part of the National Radio Propagation Programme by the Radiocommunications Agency of the Department of Trade and Industry, U.K.

7. REFERENCES

[1] R.A. Dalrymple and P.A. Martin, Perfect boundary conditions for parabolic water-wave models, Proc. R. Soc. London A, vol. 437, pp. 41-54, 1992.

[2] G.D. Dockery and J.R. Kuttler, An improved impedance algorithm for Fourier split-step solutions of the parabolic wave equation, to appear in IEEE Trans. AP.

[3] ITU-R Recommendations, Rec. 368, in Propagation in ionized media. 1994 PN Series Volume.

[4] J.R. Kuttler and G.D. Dockery, Theoretical description of the parabolic equation/Fourier split-step method of representing electromagnetic propagation in the troposphere, Radio Sci, vol. 26, pp. 381-393, 1991.

[5] M.F. Levy, Fast PE models for mixed environments, AGARD CP No 567, paper 18, 1995.

[6] M.F. Levy, Transparent boundary conditions for the parabolic wave equation, submitted for publication.

[7] S.H. Marcus, A generalized impedance method for application of the parabolic approximation to underwater acoustics, J. Acoust. Soc. Am., vol. 90, pp. 391-398, 1991.

[8] S.H. Marcus, A hybrid (finite difference--surface Green's function) method for computing transmission losses in an inhomogeneous atmosphere over irregular terrain, IEEE Trans. AP, vol. 40, pp. 1451-1458, 1992.

[9] A.R. Miller, R.M. Brown and E. Vegh, New derivation for the rough-surface reflection coefficient and for the distribution of sea-wave elevations, IEE Proc., Vol. 131, Pt H, pp. 114-116, 1984.

[10] J.S. Papadakis, Exact, nonreflecting boundary conditions for parabolic-type approximations in underwater acoustics, J. Comput. Acoust., vol. 2, pp. 83-98, 1994.

[11] D.J. Thomson and M.E. Mayfield, Using an acoustic PE with a non-local boundary condition to model propagation over an elastic bottom, Second International Conference on Theoretical and Computational Acoustics, 1995.

Paper 13

L. Felsen (US)

The local impedance concept is an idealization for plane layered penetrable media, and even more so when penetrable interfaces are irregular. Formally, by nonlocal impedance representations, the concept can be made rigorous. In the paper, the author treats the exact nonlocal problem within the confines of the PE forward scattering assumption.

Question: To what extent is it consistent with the "real" problem to solve an approximate propagation algorithm (PE) exactly with respect to boundary conditions? What are the validations where the assumptions work? Can one take the next important step to apply the model when a target is present? Target ID is very sensitive to phase, backscatter cannot be ignored, and deterministic - stochastic target-in-clutter issues become relevant.

Author's reply:

The nonlocal representation used in the paper is derived within the PE framework and hence can only be used for forward scatter modelling. Further work is required for backscatter applications. The main problem in my view is not the PE framework, but the derivation of the modified impedance for rough surfaces, and its relation to physical parameters for the case of the sea surface.

K. Anderson (US)

Have you tried using the Transparent Boundary Condition to replace the upper absorbing layer?

Author's reply:

Yes, this has been implemented with a finite difference code, with very good results. Unfortunately, the FFT approach does not lend itself to the implementation of boundary conditions on the top and bottom of the domain simultaneously. Hence using the matched transform to replace the upper absorbing layer would impose a lower absorbing layer and would not improve computation times.

G. Brown (US)

Our results with "exact" numerical studies of the GIBC indicate that as long as one is willing to stay in the near-specular scattering direction, the GIBC is a valid predictor of the rough surface effects. Conversely, it is not something that one should use in the backscatter direction, especially near low grazing angle incidence.

Author's reply:

Comparison of GIBC results with more rigorous representations of rough surfaces are essential. It is very encouraging that your results show the GIBC approach is valid in the forward scatter direction, which is the one of interest in this paper.

Paper 13

N. Douchin (FR)

Did you make comparisons between both finite differences and split-step Fourier algorithms in the particular case you have shown, using a rough surface in case of finite differences? (Shadow effects).

Author's reply:

Comparisons of the GIBC finite difference and split-step Fourier implementations show identical performance. The more challenging problem of comparing the GIBC with the explicit rough surface representation with a finite difference code has not yet been addressed.

L. T. Rogers (US)

Please give some examples of run times for your modified P.E. code.

Author's reply:

For the X-band example presented here, the run times for a smooth surface are of the order of a few seconds on a Pentium PC, and about a minute for a rough surface.

ABSTRACT

of

Paper 14

"TOPEX POSEIDON FOLLOW-ON RADAR ALTIMETER DESIGN AND DEVELOPMENT STATUS

Philippe Escudier*,Philippe Raizonville*,Guy Carayon*,Pierre de Chateau Thierry**

* Centre National d'Etudes Spatiales, 18 avenue Edouard Belin, 31055 Toulouse Cedex, France.
** Alcatel Espace, 26 bis avenue Jean François Champollion, 31037 Toulouse Cedex, France.

The TOPEX POSEIDON FOLLOW-ON (TPFO) mission will be the continuation of the present TOPEX POSEIDON mission, whose major aim is to observe the ocean circulation on a large scale basis in terms of space and time.
The data provided by the TOPEX mission have a level of accuracy never reached before and allow a continuous survey of the oceans particularly worthy for the scientists involved in oceanography and in climate related studies.
This quality of the data has been made possible by the high performances of the instruments, which have all reached or exceeded the specifications.

Among these, POSEIDON altimeter is a CNES provided radar altimeter, on board TOPEX as an experimental instrument, with a limited (10%) time of operation and with novel design in order to allow a low mass, low power, low telemetry rate; first, the solid-state concept was selected, consisting in a transistor technology for the amplifier instead of a travelling wave tube as for previous generation altimeters; second, industrial technology of the altimeter contractor, ALCATEL ESPACE Toulouse, was intensively used to optimize the design and reduce size, mass and power demand. Third, data on board processing was made as efficient as possible to reduce data transmission needs.

All these efforts lead to a very lightweight, low power instrument, with a 25 kg mass and 50 Watts power demand, not considering the antenna nor redundancy.
This altimeter concept is now flight proven; it has been operating successfully since mid-1992, and produced data with a high level of accuracy.
For these reasons, POSEIDON altimeter has been selected as the baseline unique altimeter for the TPFO mission.
In this framework, its design has yet evolved, in order to cope with upgraded specifications, namely the need for two bands of operation, one at C band and one at Ku band.
The TPFO altimeter architecture is very similar to the POSEIDON one, at both frequencies; technological improvements are also introduced, to keep the mass and power consumption as low as possible while C band is added: the transmitted signal is a 320 MHz chirp signal generated by a digital generator, the pulses are amplified in C band and Ku bands by solid state amplifiers with a new technology allowing better power efficiency, signal real time processing is done partly by ASICs dedicated to prefilitering and Fourier transform.
Signal processing is also modified in order to manage both frequencies, and to further improve peformances.

Paper presented at the AGARD SPP Symposium on "Remote Sensing: A Valuable Source of Information", held in Toulouse, France from 22-25 April 1996, and published in CP-582.

The TPFO altimeter is presently under design at ALCATEL ESPACE; phase B started in September 1994, and is planned to end in April 1996; it includes a lot of hardware breadboarding: the digital chirp generator, the power amplifiers, the C band diplexer, the ASICs..... Real time software studies are also conducted.

The Phase C (realization) of the altimeter is planned to start in May 1996, the delivery for satellite integration is scheduled for March 1998, and the launch in March 1999. Thanks to the efforts made as concerns instruments mass limitation, it will be possible to achieve the TPFO mission using a small satellite platform, presently under definition at CNES and in the French space industry.

Using Remote Refractivity Sensing to Predict Tropospheric Refractivity from Measurements of Microwave Propagation

Donald Boyer, Greg Gentry, and Janet Stapleton
Naval Surface Warfare Center Dahlgren Division
Dahlgren, VA 22448-5100 USA
(540)653-1908
dboyer@nswc.navy.mil

Stephen Burk and John Cook
Naval Research Laboratory
Marine Meteorology Division
Monterey, CA 93943-5502 USA

SUMMARY

Direct measurements of vertical profiles of humidity and temperature at various ranges, from which range dependent refractivity structures are derived, are very difficult to obtain onboard a ship, and these profiles are needed in order to predict the performance of a radar system in a coastal marine environment. Recently remote sensing techniques have been proposed which attempt to deduce the refractivity structure over a path by measuring relative signal levels from known emitters. Since in a naval battle group there are known transmitters at different frequencies, whose signals can be monitored by any friendly ship, these radio propagation techniques offer high potential for implementation as part of a shipboard sensor assessment system.

In order to derive refractivity fields in the lower atmosphere above the sea surface from measurements of propagation loss, the refractivity profiles must be represented in terms of a small number of parameters. This paper discusses using analytic representations of mean refractivity profiles in the constant-flux surface layer of the marine atmospheric boundary layer. The numerical inversion techniques used for obtaining refractivity fields from measurements of propagation loss are also included. A separate paper [2] describes some of the important effects of these profiles along with the transmitter/receiver geometry and frequency band of the propagation measurements.

1. INTRODUCTION

Radar coverage in a coastal marine environment are very sensitive to vertical and horizontal variations in the tropospheric refractivity conditions. These systems experience extended or reduced detection ranges, inaccurate altitude estimates, and increased surface clutter due to the presence of nonstandard tropospheric refractivity conditions (e.g., trapping layers or ducts, and subrefractive layers). If synoptic range dependent refractivity fields were available in a timely manner aboard a ship, an automated decision aid could be designed to adjust the sensor output to account for these propagation effects. In fact, decision aids have been developed for use on U.S. Navy ships, but their use is limited by the inadequacy of available synoptic refractivity measurements.

This paper explores some of the major issues involved in using far-field measurements of propagation to estimate the refractivity fields through which the electromagnetic waves were transmitted. If the propagation measurements consist of phase as well as amplitude information, then matched field processing techniques, which have been used successfully in acoustics, can be employed [1] These techniques would employ a modeled signal waveform, or a replica, which would be cross correlated with the measured complex-valued sensor data, to yield a tomographic reconstruction of the refractivity medium. Though there has been interest recently in electromagnetic matched field processing, the techniques discussed in this paper only employ measurements of signal amplitude at the receiver, in order to characterize the refractivity medium through which the electromagnetic waves were transmitted.

In order to estimate tropospheric refractivity fields from measurements of propagation over a path, each of the following computational tools must be employed:

1. meteorologically meaningful models, in terms of a small number of parameters, which adequately characterize the range and height variations in the refractivity environment

Paper presented at the AGARD SPP Symposium on "Remote Sensing: A Valuable Source of Information", held in Toulouse, France from 22-25 April 1996, and published in CP-582.

2. an accurate, but computationally efficient simulation of the propagation of electromagnetic waves with range and height varying refractivity fields

3. an accurate and computationally efficient numerical inversion technique, which selects numerical values of the refractivity parameters, based on matching the outputs of the propagation model to measured propagation data.

To determine the effectiveness of such an estimation procedure, two data sets of the following types are needed:

4. a collection of propagation measurements with height and frequency diversity

5. meaningful meteorological data, corresponding in time and in the vicinity of the location at which at least some of the propagation data in item 4 were collected.

The measured propagation data referred to in this paper as the Stapleton/Kang data is described in more detail by another paper in this symposium [2]. These measurements were collected at Wallops Island on the Atlantic Ocean in Virginia, during the months of February-April, 1994. The geometry consisted of a path of 29 kilometers over the water, with 10 transmitters within 12 meters of the surface, and 4 receivers within 28 meters of the surface. The signals were transmitted in the band 2–18 GHz [3]. The meteorological data and propagation data will be discussed in Section 4 as will be the meteorological models of item 1. It is important to note the geometry of the propagation path necessitates parameterization of the meteorological environment only in the lower two hundred or so meters of the atmospheric boundary layer. The propagation model employed in this study is Radio Physical Optics (RPO) model [4]. RPO is a parabolic equation (PE) model which uses the Fourier split-step algorithm to solve the parabolic approximation to variable coefficient wave equation. It simulates the effects of range and height variation in refractivity, as well as, the effects of transmitter/receiver geometry, frequency, polarization, antenna pattern, the earth's curvature, electrical properties of the sea surface and atmospheric absorption. Though RPO does not include the surface roughness effects in the propagation computation in the diffraction region, it is believed that these effects will be small over a 29 kilometer path. A more detailed discussion of PE models can found in [5] and [7]. The inversion algorithm of item 3 is based on nonlinear least squares, and uses a modified finite difference Gauss-Newton technique. The inversion algorithm with examples will be discussed in Section 3.

2. REFRACTIVITY AND PROPAGATION NEAR THE SEA SURFACE

Propagation of electromagnetic waves near the surface of the sea is greatly influenced by the vertical variations of the refractive index of the adjacent atmosphere. The refractive index, n, of air near the earth's surface, at microwave frequencies, varies between 1.00024 and 1.00040. It is traditional to introduce a quantity called radio refractivity, N, which is related to the refractive index by

$$N = (n-1) \times 10^6.$$

Clearly, radio refractivity varies between 240 and 400 N-units. The height variation of N can expressed in terms of meteorological profiles of air temperature (T), atmospheric pressure (P) and specific humidity (q) as

$$N(z) = 77.6\frac{P(z)}{T(z)}\left(1 + 4810\frac{q(z)}{\epsilon T(z)}\right) \qquad (1)$$

where T is in degrees Kelvin, P is in millibars, q is in kg/kg, and $\epsilon = 0.622$. The critical value of the slope of the refractivity curve is the curvature of the earth

$$\left(\frac{dN}{dz}\right)_c = \frac{-10^6}{a} = -0.157\frac{N}{\text{m}},$$

where $a = 6.371 \times 10^6$ meters is the mean earth's radius. Any slope of the refractivity curve more negative than the critical value will cause the waves to curve around the earth, or be trapped or "ducted"; thus a layer of air whose refractivity slope is more negative than $-0.157N/\text{m}$ is called a trapping layer or a tropospheric duct. A slightly more convenient manner in which to represent tropospheric refractivity is modified refractivity

$$M(z) = N(z) + 0.157z.$$

Clearly the trapping value for the slope dM/dz is 0, and for the study of electromagnetic wave propagation, refractivity is usually expressed in M-units. There are three refractivity states for layers of air, in addition to ducting that are usually identified relative to electromagnetic wave propagation. These are subrefraction, normal or standard, and superrefraction; the four states are defined by Table 1.

Since a trapping layer serves as a leaky waveguide with respect to radio wave propagation, in general terms, a thicker trapping layer is associate with the stronger the effect. Figures 1 and 2 illustrate the effects of a surface based trapping layer, 4/3-earth or standard refraction, and a subrefractive layer on near-surface propagation, using synthetic profiles in the RPO simulation with the geometry of the Stapleton/Kang data. Transitions from strong ducting conditions to subrefraction have been observed in the Stapleton/Kang propagation data off the east coast

$\frac{dM}{dz} > .157$	$\frac{dN}{dz} > 0$	sub
$.81 < \frac{dM}{dz} < .157$	$-.76 < \frac{dN}{dz} < 0$	normal
$0 < \frac{dM}{dz} < .81$	$-.157 < \frac{dN}{dz} < -.76$	super
$\frac{dM}{dz} < 0$	$\frac{dN}{dz} < -.157$	ducting

Table 1: Gradients of Refractivity.

of the United States. Transitions of as much as 60 dB or more of propagation factor have been observed in a time span of less than two hours. The propagation factor (F) is related to the one-way path or propagation loss by

$$L = L_{fs} - 20 \log F,$$

where the free-space path loss is given by

$$L_{fs} = 20 \log \left[\frac{4\pi R}{\lambda} \right],$$

R is the range, and λ is the wavelength. The paper by Babin [13] discusses the meteorological situations which induce a refractivity change from ducting to subrefraction off the coast of Wallops Island. Since in the troposphere the temperature generally decreases with height, the situation in which air temperature increases with height is called a temperature inversion. In the presence of a temperature inversion, if the humidity decreases with height, then a trapping layer results. While if the humidity increases with height in a temperature inversion, a subrefractive layer is generally formed. So, a movement of fog from the ocean toward the land can induce a transition from ducting to subrefraction off the coast, with the resulting effect on microwave propagation as indicated in Figures 1 and 2.

3. REFRACTIVITY ESTIMATION FROM PATH LOSS MEASUREMENTS

It is well-known in the radar community that if synoptic refractivity data were readily available, tactical decision aids could be developed to compensate for the effects of anomalous propagation on the sensors and weapon systems. However, direct measurements of atmospheric refractivity with sufficiently high resolution to describe near-surface profiles are not routinely available from radiosonde data. Point assessments of the evaporation duct height can be inferred from routine shipboard meteorological measurements using techniques generally referred to as flux-profile relationships. As this paper shows later, these techniques provide a very convenient parameterization of

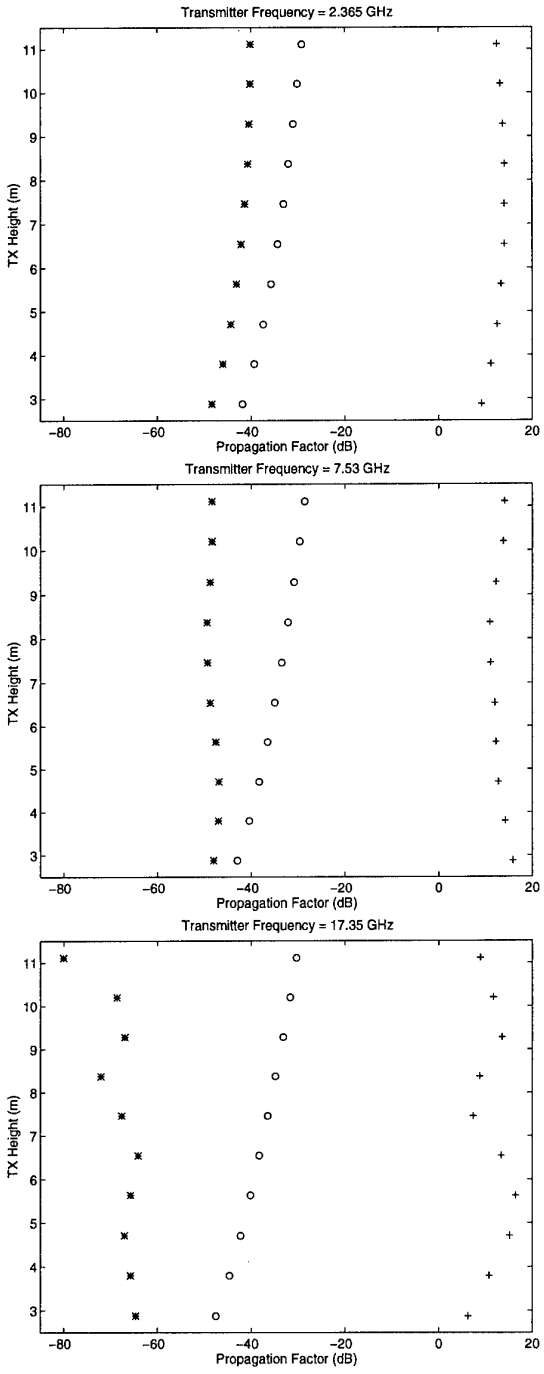


Figure 1: Receiver Height at 5.79 meters. The plus denotes a 75 meter surface-based duct, the circle denotes the standard refraction, and the asterick denotes subrefraction.

refractivity in the lower 50 meters or so of the marine atmospheric boundary layer. However, in practice, onboard a ship, these techniques are also laden with problems [6, 8]. A refractivity remote sensing technique proposed in this paper uses measured path loss of electromagnetic waves, and a numerical inversion technique to estimate the refractivity fields along the

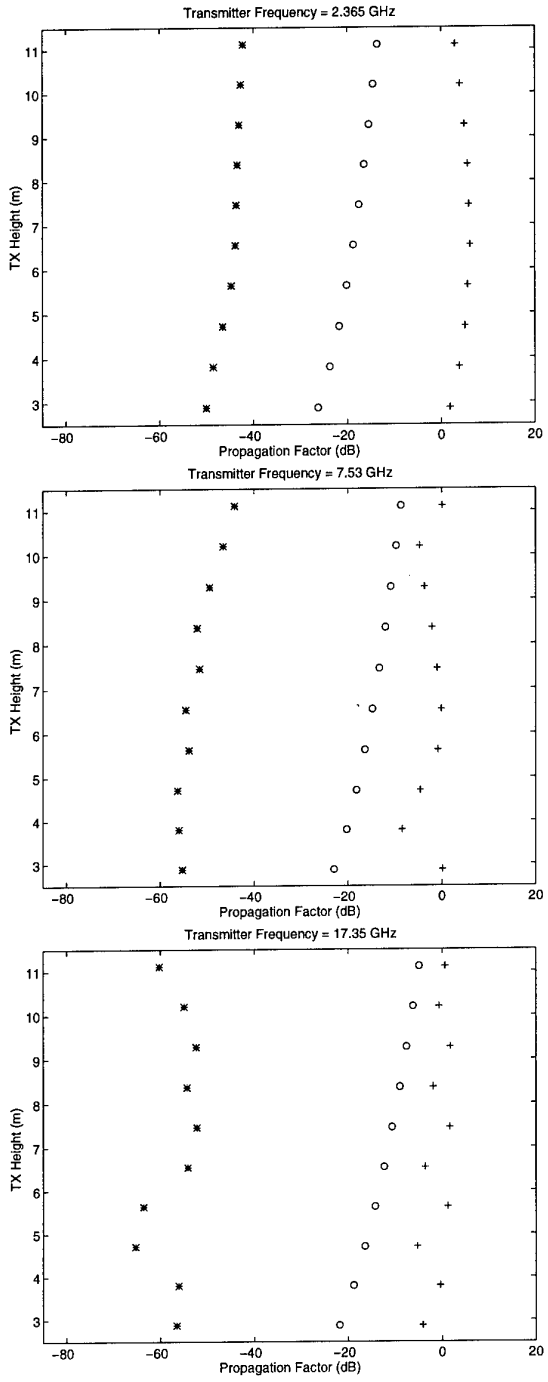


Figure 2: Receiver Height at 25.6 meters. The plus denotes a 75 meter surface-based duct, the circle denotes the standard refraction, and the asterick denotes subrefraction.

path. Such a technique might be used to supplement direct refractivity measurements as part of a sensor assessment system, and will now be described in detail.

In the last few years, Colton and Kress [9] have extended the theory of inverse scattering of acoustic and electromagnetic waves to include the inverse medium problem. The concepts have been successfully applied to a medical imaging problem [10]. Since the present refractivity estimation problem involves much more general initial and boundary conditions than Colton and Kress consider, as well as a nonhomogeneous scattering medium, it is not possible to obtain classical solutions in terms of integral equations, even for the direct problem. Since the parabolic approximation to the two-dimensional, variable coefficient Helmholtz equation, or the PE model, accounts for the initial and boundary conditions, it is used as the model to match the measured propagation data. A formal statement of the problem will now be given and the solution procedure illustrated with a very simple synthetic example.

Assume that there are l_1 transmitters' locations and l_2 receivers' locations. Though one can assume range diversity, to simplify the notation, assume a fixed range between the antennas along with k distinct frequencies. The tropospheric refractivity environment is parameterized by a set of l profiles $\{P_1, P_2, \cdots, P_l\}$, with the i-th profile P_i defined by n_i parameters. Thus, a parameter vector would be $\mathbf{x} = [p_1^1, p_2^1, \cdots, p_{n_1}^1, \cdots, p_1^l, p_2^l, \cdots, p_{n_l}^l]^T$. So when one says $\mathbf{x} \in \mathbf{R}^n$ one means $n = n_1 + n_2 + \cdots + n_l$. One can either use an analytical representation for each profile, in terms of two or three parameters, or a piecewise linear or spline characterization. Advantages of the former approach are the greatly reduced size of the parameter space that must be searched, and the more meteorologically meaningful solutions which result. Now, since a PE model outputs a path loss at a discrete set of receiver heights, for a given transmitter height and frequency, the PE model must be executed $k \cdot l_1$ times to obtain the $m = k \cdot l_1 \cdot l_2$ modeled path losses. Thus the direct medium scattering equation

$$\mathbf{y} = \mathbf{F}(\mathbf{x}) \tag{2}$$

denotes a nonlinear vector valued function with the PE model mapping the refractivity parameter space in R^n into the modeled path loss space in R^m. Now, one is not actually inverting Equation (2), since the measured path loss vector $\hat{\mathbf{y}}$ is not in the range of PE model, so one must find a $\hat{\mathbf{x}}$ in the refractivity parameter space which minimizes the Euclidean norm of the residual between modeled and measured path loss:

$$R(\mathbf{x}) = \frac{1}{2}\Big(\mathbf{F}(\mathbf{x}) - \hat{\mathbf{y}}\Big)^T \Big(\mathbf{F}(\mathbf{x}) - \hat{\mathbf{y}}\Big). \tag{3}$$

The Levenberg-Marquardt iteration technique can be written as: find a $\mu_k \geq 0$ and a $\mathbf{x}$ such that

$$\Big(\mathbf{J}_k^T \mathbf{J}_k + \mu_k \mathbf{I}\Big)\, \mathbf{x} = \mathbf{J}_k^T \left(\hat{\mathbf{y}} - \mathbf{F}_k\right) \tag{4}$$

where $\mathbf{J}_k = \mathbf{J}(\mathbf{x}_k)$ is the Jacobian matrix of partial derivatives with respect to $\mathbf{x}_k$ approximated by finite

differences using the PE model for the function evaluations. A singular value decomposition is performed on the Jacobian matrix to remove the redundant parameters. Then, μ_k can usually be set equal to 0 in Equation (4), and the much simpler Gauss-Newton method results, involving only positive definite Hessian approximations $\mathbf{H}_k = \mathbf{J}_k^T \mathbf{J}_k$. In this case, a QR decomposition of the Hessians results in a very efficient, robust implementation.

Consider the special case of the Levenberg-Marquardt algorithm when it is possible to let $\mu_k = 0$ in Equation (4), and the Gauss-Newton algorithm is then applicable. This powerful approach to solving the nonlinear least squares problem in Equation (3) can be obtained from a first-order approximation to Equation (2)

$$\mathbf{F}(\mathbf{x}) \simeq \mathbf{F}(\mathbf{x}_0) + \mathbf{F}'(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0), \qquad (5)$$

where $\mathbf{x}_0$ is an initial approximate solution. Now, letting $\mathbf{J}_0 = \mathbf{F}'(\mathbf{x}_0)$ be the Jacobian matrix of partial derivatives of output propagation factors with respect to each of the refractivity parameters, Equation (5) can be rewritten as

$$\mathbf{y} = \mathbf{y}_0 + \mathbf{J}_0(\mathbf{x} - \mathbf{x}_0). \qquad (6)$$

The propagation measurement vector $\hat{\mathbf{y}}$, can be set equal to the right hand side of the equation, and the equation

$$\hat{\mathbf{y}} = \mathbf{y}_0 + \mathbf{J}_0(\mathbf{x} - \mathbf{x}_0) \qquad (7)$$

is solved for $\mathbf{x}$, writing $\mathbf{J}_0^+$ for the pseudo-inverse of $\mathbf{J}_0$ to obtain

$$\mathbf{x} = \mathbf{x}_0 - \mathbf{J}_0^+(\mathbf{y}_0 - \hat{\mathbf{y}}).$$

An iteration procedure based on the above equation is referred to, as a finite difference Gauss-Newton type procedure for inverting the nonlinear Equation (2), or equivalently solving the nonlinear least squares problem (3). The k-th step in the iteration procedure is

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{J}_k^+(\mathbf{y}_k - \hat{\mathbf{y}}), \qquad (8)$$

and when $\mathbf{J}_k$ has full column rank, its pseudo-inverse is conventionally expressed as

$$\mathbf{J}_k^+ = \left(\mathbf{J}_k^T \mathbf{J}_k\right)^{-1} \mathbf{J}_k^T,$$

and the familiar Gauss-Newton formula for nonlinear least squares appears. Dennis and Schnabel's book [11] has a careful treatment of the numerical issues involved in nonlinear least squares. Now, a simple example should illustrate the application of these concepts to the refractivity inversion problem.

The following example uses synthetically generated propagation measurements to determine a refractivity field using the numerical inversion technique previously outlined. The geometry employed in the example corresponds to the Stapleton/Kang data, with the role of transmitter and receiver reversed. The problem parameters are:

$$\begin{aligned} \text{Frequency} &= 7.530\,\text{GHz}, \\ \text{Transmitter Height} &= 16.46\,\text{meters}, \\ \text{Receiver Heights} &= \begin{bmatrix} 2.88 \\ 5.63 \\ 8.37 \\ 11.11 \end{bmatrix} \text{meters}, \\ \text{Range} &= 29.45\,\text{kilometers}. \end{aligned}$$

The synthetic measurement vector consists of propagation factors is in dB, and the components are indexed by the corresponding receiver heights:

$$\text{propagation measurement} = \begin{bmatrix} 5.09 \\ 6.86 \\ 5.87 \\ 4.04 \end{bmatrix} \text{dB}. \qquad (9)$$

The radio horizon limit vector for this transmitter receiver geometry is

$$\text{horizon limit} = \begin{bmatrix} 23.7 \\ 26.5 \\ 28.6 \\ 30.4 \end{bmatrix} \text{km}.$$

Since the range between the two antennas is 29.45 kilometers, it is clear from sign of the elements of the propagation measurement vector that the refractivity field must contain a moderately strong surface-based trapping layer. In fact, the standard propagation vector for this geometry is easily seen to be

$$\text{standard propagation} = \begin{bmatrix} -30.07 \\ -23.48 \\ -19.12 \\ -15.60 \end{bmatrix} \text{dB}.$$

In order to invoke the numerical inversion routine, an initial approximate solution is needed, but first a parameterization of the refractivity environment must be proposed. An investigation of measured profiles and experience with examining the output from propagation models for such a case indicate that the refractivity field might be adequately approximated by two, or possibly three evaporation duct profiles. So, employing the previous notation, let $\mathbf{x} = [p_1^1, p_1^2]^T$, where $p_1^1 = d_1$ and $p_1^2 = d_2$ are duct heights in a simple log-linear evaporation duct parameterization, which is derived on the assumption of neutral thermal stability [16],

$$M(z) = M(0) + \alpha(z - d)\log\left(\frac{z + z_0}{z_0}\right), \qquad (10)$$

where z_0 is the surface roughness parameter with value taken to be 1.5×10^{-4} meters (from EREPS [16]) and α is the asymptotic value of the slope of

the profile with value taken to be 0.125 M-units per meter. So, the problem statement of Equation (2) becomes for this example

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \mathbf{F}\left(\begin{bmatrix} d_1 \\ d_2 \end{bmatrix}\right),$$

where the function symbol "**F**" represents an execution of the numerical wave equation solver RPO, with an evaporation duct of height "d_1" located at the transmitter, and an evaporation duct of height "d_2" located at the receiver. The refractivity inversion problem for this example can be stated as: *given the propagation measurement vector in Equation (9), find two duct heights d_1 and d_2 such that*

$$\mathbf{F}\left(\begin{bmatrix} d_1 \\ d_2 \end{bmatrix}\right) = \begin{bmatrix} 5.09 \\ 6.86 \\ 5.87 \\ 4.04 \end{bmatrix}.$$

The statement of the problem is very dependent on the assumed refractivity parameterization, and a very simple representation was introduced in this problem. Much more meteorologically sophisticated refractivity representations will be described later, but the simplicity of this example aids in understanding the numerical procedure.

The first step in the iteration procedure requires an initial approximate solution. Assume an initial solution to be two evaporation ducts of height 5 meters, so

$$\mathbf{x}_0 = \begin{bmatrix} 5 \\ 5 \end{bmatrix}. \tag{11}$$

Now, the initial function value is

$$\mathbf{y}_0 = \mathbf{F}(\mathbf{x}_0) = \begin{bmatrix} -17.88 \\ -12.66 \\ -9.81 \\ -7.71 \end{bmatrix}, \tag{12}$$

which is not particularly close to the measurement, but now the iteration process is applied. From Equation (8), the first step in the iteration should be

$$\mathbf{x}_1 = \mathbf{x}_0 - \mathbf{J}_0^{+}(\mathbf{y}_0 - \hat{\mathbf{y}}). \tag{13}$$

The only term in the equation which is not already computed is the pseudo-inverse of the Jacobian matrix

$$\mathbf{J}_0 = \begin{bmatrix} 1.88 & 3.12 \\ 1.88 & 2.88 \\ 1.75 & 2.50 \\ 1.61 & 1.89 \end{bmatrix}. \tag{14}$$

The elements in this matrix are computed using numerical derivatives. Each individual element, such as

$$\mathbf{J}_0(1,2) = \frac{\partial \mathbf{F}_1}{\partial x_2} = 3.12,$$

Iteration	0	2	6	8
x_1	5.0	-3.1	8.8	10.0
x_2	5.0	25.6	20.4	20.0
y_1	-17.9	3.8	5.18	5.09
y_2	-12.7	4.3	6.85	6.86
y_3	-9.8	2.7	5.77	5.87
y_4	-7.7	1.6	4.04	4.04
Residual	646	12.0	0.1	0

Table 2: Summary table for iterations.

can be interpreted physically as the rate of change of propagation factor at the lowest receiver with respect to the duct height of the second profile and in this particular case is 3.12 dB per meter of change in duct height. So, the units of the elements of $\mathbf{J}_0$ are "dB" per "meter" of duct height. Thus, the units of the elements of the pseudo-inverse must be "meter" of duct height per "dB" of propagation factor, and

$$\mathbf{J}_0^{+} = \begin{bmatrix} -1.24 & -0.32 & 0.42 & 1.99 \\ 0.95 & 0.32 & -0.19 & -1.27 \end{bmatrix}.$$

So, the first iteration yields

$$\mathbf{x}_1 = \begin{bmatrix} 0.13 \\ 15.1 \end{bmatrix}.$$

Note that the first iteration took a rather large step and the corresponding residual as computed by Equation (3) is 646 dB for $\mathbf{y} = \mathbf{y}_0$. For $\mathbf{y} = \mathbf{y}_1$, the residual is 139 dB. Since the residual was reduced significantly over this iteration, the step was in the correct direction, even though the initial solution was poorly selected. Table 2 illustrates how the refractivity inversion proceeds numerically. After eight iterations, the residuals are reduced to zero.

Even though the duct height in profile 1 in the 6^{th} iteration is more than a meter less than the correct one, the effect on the propagation factors is small. The propagation model, RPO, transitions the refractivity field smoothly between the two evaporation duct profiles, so that the effect of the higher duct height is much greater than the lower one relative to this freqency and transmitter/receiver geometry. This example illustrates that the technique performs quite well on this geometry and frequency, even with a poor initial approximation. This fact would indicate that it might perform well on measured propagation data, especially since much more sophisticated meteorological representations of the refractivity will be employed than the simplistic one-parameter evaporation duct heights employed in this simple synthetic example. Computational experience has established that Gauss-Newton type methods tend to perform well on large residual nonlinear least squares problems when the largest singular value of the Jacobian is greater than unity and the first-order approximation as in Equation (6) is valid over a modest portion of the domain [11]. These two conditions

have been satisfied in most of the synthetic examples to which this procedure has been applied. However, this does not mean that even a very robust numerical procedure is assured of performing in the presence of both modeling and mesurement errors. An attempt in the next section will be made to find more adequate parametric representations of refractivity than just log-linear evaporation duct profiles.

4. METEOROLOGICAL REPRESENTATIONS OF REFRACTIVITY PROFILES IN THE SURFACE LAYER

A refractivity height profile is determined by the humidity, temperature and pressure profiles as defined in Equation (1). A portion of the atmosphere throughout which such a meteorological profile only varies with height, is usually called an air mass. An air mass is a portion of the earth's atmosphere throughout which the meteorological properties are approximately horizontally homogenous. Air mass formation is primarily a result of the transfer of heat, moisture, and momentum that takes place at the air-sea or air-earth boundary, as well as, at air mass boundaries. Air masses experience dynamic modifications primarily as a result of changes in the underlying boundary, and internal dynamic processes which alter the vertical structure. When an air mass whose properties were determined over land, passes out over a sea of different surface temperature and roughness conditions, the refractive structure in the surface layer may undergo significant vertical, as well as, horizontal modifications. Two recent papers discuss coastal meteorology from different points of view, but both contain information of great relevance to an understanding of refractivity in a coastal marine environment [12,13].

The type of near-surface refractivity structure which exists frequently over the ocean is the evaporation duct, and it is formed by a very sharp decrease in atmospheric water vapor content in the air slightly above the sea surface. The evaporation duct height is defined as that point in the height profile at which $dM/dz = 0$, and the duct height seems to be an important quantity relative to the trapping at microwave frequencies. As noted in Section 3, the evaporation duct height can be inferred from a sea surface temperature measurement, and measurements of air temperature, relative humidity, and atmospheric pressure at a reference height a few meters above the surface, using techniques developed in the meteorological community. These techniques are generally referred to as flux-profile relationships and are based on similarity theory in fluid mechanics [14, 15]. The near-surface fluxes of momentum, heat and moisture are derived from bulk measurements of these parameters. These fluxes are assumed to be constant over the lower 10% of the atmospheric boundary layer and determine the vertical gradients of wind speed, temperature, and humidity. From Equation (1), refractivity in the surface layer may be obtained.

The simplest representation of an evaporation duct has been given in Equation (10). If the air temperature T_a and the sea temperature T_s are equal, then such a simple representation in terms of just one parameter may be adequate to match the refractivity to propagation measurements, since this profile is derived under the assumption of neutral thermal atmospheric stability conditions. Liu, Katsaris and Businger wrote a paper in 1979 [17] which is the basis of a derivation of stability dependent mean temperature and humidity profiles in the marine surface layer from the bulk parameter measurements. The technique uses models of the interfacial sublayers on both sides of the air-sea interface in order to numerically determine the air-sea exchanges of momentum, heat, and humidity, and employs the Monin-Obukhov similarity theory to derive representations of the mean profiles. The vertical profiles of potential temperature and specific humidity are

$$\theta(z) = \theta_s + \frac{\theta_*}{k\alpha}\left[\log\left(\frac{z}{z_\theta}\right) - \psi\left(\frac{z}{L}\right)\right], \quad (15)$$

$$q(z) = q_s + \frac{q_*}{k\alpha}\left[\log\left(\frac{z}{z_q}\right) - \psi\left(\frac{z}{L}\right)\right], \quad (16)$$

where

$$\psi(\xi) = \begin{cases} 2\log\left[\frac{1}{2}\left(1+\sqrt{1-16\xi}\right)\right], & \text{for } \xi < 0, \\ -7\xi, & \text{for } \xi > 0, \\ 0, & \text{for } \xi = 0. \end{cases}$$

Each of the quantities $\theta_*, q_*, \theta_s, q_s, z_\theta, z_q, L$ is computed by the LKB method using the bulk parameters: sea surface temperature (T_s), air temperature (T_a), relative humidity (R), atmospheric pressure (P), and wind speed (u), each specified at a reference height, a few meters above the sea surface. The quantities k and α are atmospheric constants. The stability conditions may be summarized in terms of the Monin-Obukhov length, L, or the air-sea temperature difference, ASTD $= T_a - T_s$ by

$\frac{1}{L} < 0$ or ASTD < 0 : for unstable or convective,

$\frac{1}{L} > 0$ or ASTD > 0 : for stable,

$\frac{1}{L} = 0$ or ASTD $= 0$: for neutral,

The refractivity inversion procedure needs parametric representations of refractivity which are general enough to encompass most mean surface layer refractivity profiles. The LKB representation, in terms of the four bulk parameters, will generate mean potential temperature and specific humidity surface layer profiles, according to Equations (15) and (16), and the temperature profile is obtained from

$$T(z) = \alpha\theta(z) - \Gamma z,$$

where $\alpha = (p_0/1000)^\kappa$, with $\kappa = 0.286$, and $\Gamma = 0.098$. The mean refractivity profile is then generated using Equation (1) with the atmospheric pressure expressed in terms of virtual temperature $T_v(z)$, by

$$p(z) = p_0 \exp\left(-\frac{z}{29.3 T_v(z)}\right),$$

where p_0 is the atmospheric pressure at the surface.

The LKB representation allows for a very general parameterization of the mean meteorological profiles in the constant-flux surface layer, in terms of just five parameters. The assumptions under which these surface layer profiles are derived, require that the atmospheric dynamics remain both steady state and horizontally homogeneous for some period of time. Considering the presence of diurnal changes, moving weather systems and the inhomogeneities due to the land-sea boundary, such an idealized representation can only hope to model the real world at some locations and some of the time. However, in spite of the very restrictive meteorological assumptions required to derive log-linear evaporation duct representations such as Equation (10), these profiles have been used successfully by radio meteorologists over the last forty years to predict and analyse radar performance in the presence of ducting over the ocean.

Consider the following synthetic meteorological example:

$$\begin{aligned} \text{water temperature} &= 18^\circ \text{ C} \\ \text{air temperature} &= 20^\circ \text{ C at 2 m} \\ \text{relative humidity} &= 80\% \text{ at 2 m} \\ \text{atmospheric pressure} &= 1013.25 \text{ mbars at 2 m} \\ \text{wind speed} &= 3.6 \text{ m/s at 2 m} \end{aligned}$$

so that the air-sea temperature difference is ASTD $= +2^\circ$. Because of the temperature inversion, the decrease in humidity, and lack of a strong wind, a moderately deep surface-based duct is created as displayed in Figure 7. Suppose that a cool air mass moves out to sea from the land and as a result the air temperature decreases by 2 degrees, and that the other conditions do not change appreciably over a time period. The cooling of the air results in ASTD $= 0^\circ$, and the corresponding log-linear evaporation duct profile with a duct height of 8.7 meters as shown in Figure 7. Figures 3 and 4 show the effects on the propagation of lowering the air temperature from 20°, 18°, and 16°, while holding the other parameters constant. Figures 5 and 6 show the same effects of air-sea temperature difference on propagation, but with a much stronger mean wind speed of 12 m/s. The stronger mean wind causes mechanical turbulence due to friction, which constrains the mean refractivity gradients to near-neutral conditions as

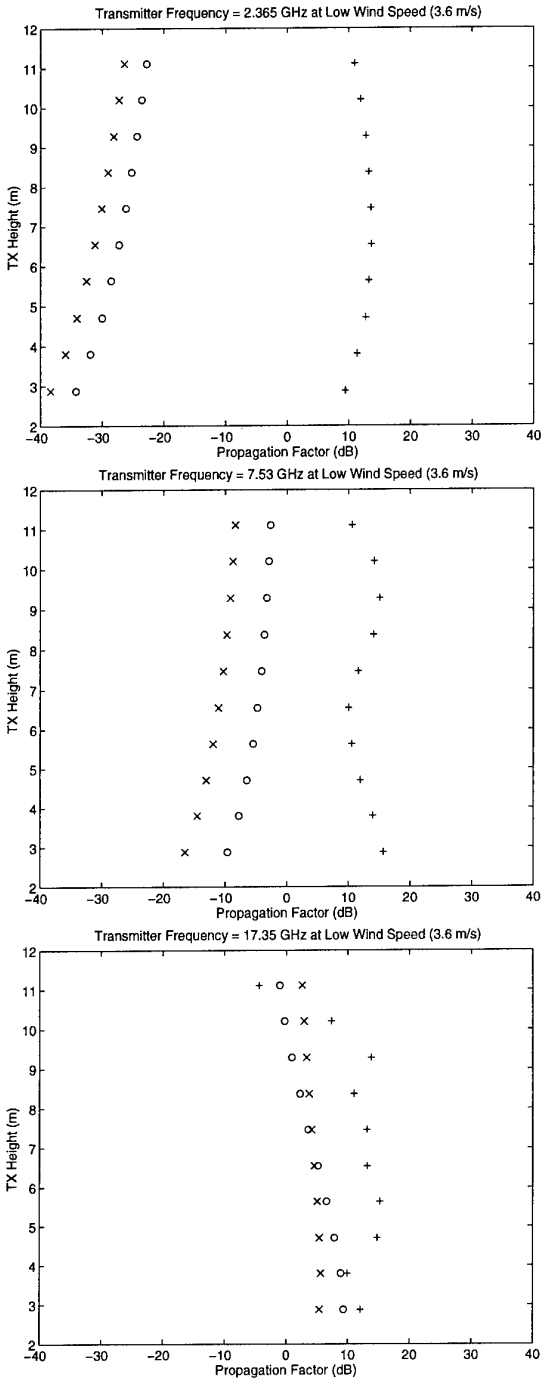


Figure 3: Receiver Height at 5.79 meters at low wind speed. The plus denotes ASTD $= +2^\circ$, the circle ASTD $= 0^\circ$, and the cross denotes ASTD $= -2^\circ$.

displayed in Figure 8. As a result, the propagation factors tend to vary much less with stability than is apparent in the absence of a strong mean wind as seen by comparing Figures 3 and 4 with Figures 5 and 6.

The parameters in the LKB representation of refractivity, which the numerical inversion procedure will

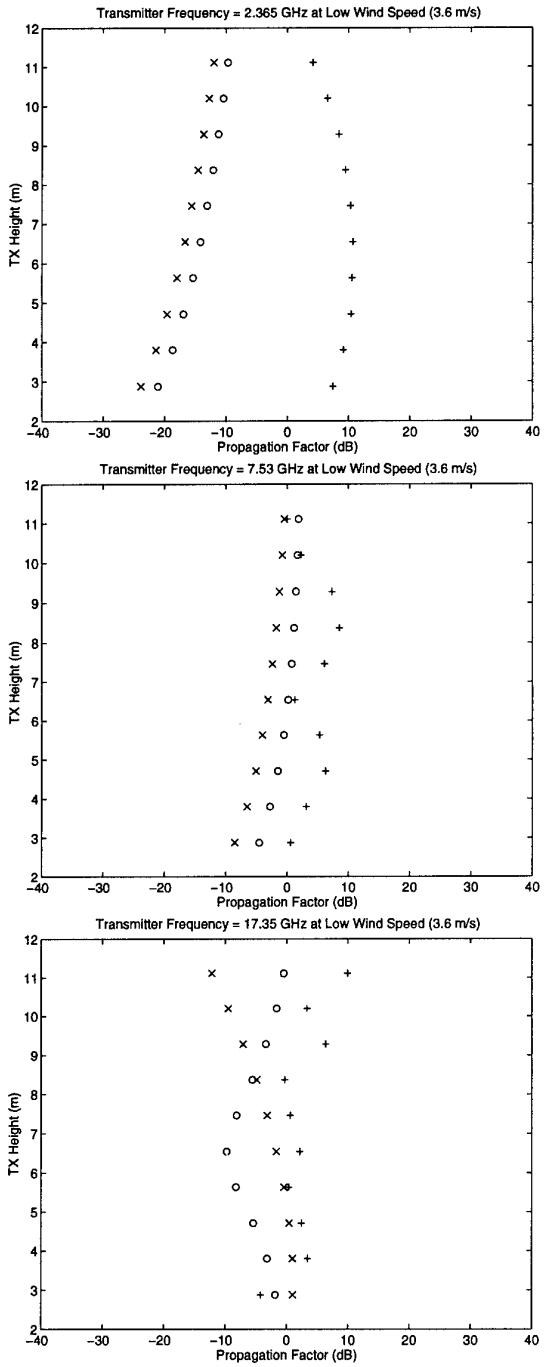


Figure 4: Receiver Height at 25.6 meters at low wind speed. The plus denotes ASTD = $+2°$, the circle ASTD = $0°$, and the cross denotes ASTD = $-2°$.

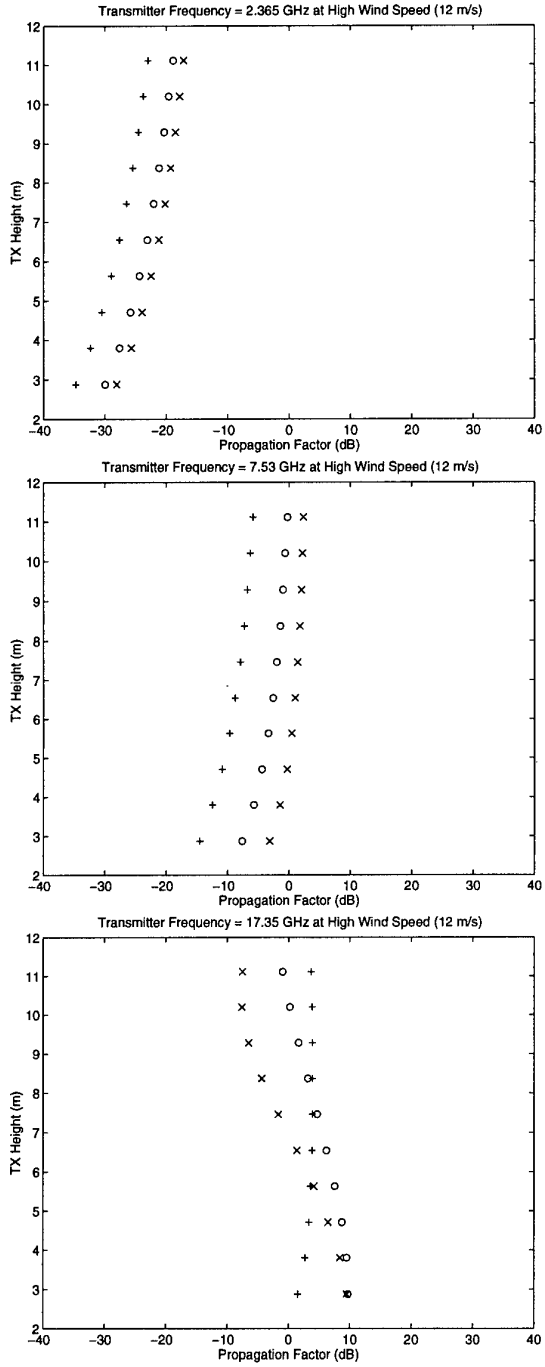


Figure 5: Receiver Height at 5.79 meters at high wind speed. The plus denotes ASTD = $+2°$, the circle ASTD = $0°$, and the cross denotes ASTD = $-2°$.

attempt to estimate from propagation measurements are wind speed, air-sea temperature difference, and relative humidity. A slightly simpler stability dependent parameterization for an evaporation duct is based on a paper by Rotheram [18]

$$M(z) = M_0 + \alpha \left\{ z - \frac{(d+z_0)}{\phi(d)} \int_0^z \frac{\phi(s[z+z_0])}{z+z_0} dz \right\},$$

where d is the duct height, and $s = 1/L$ is the Monin-Obukhov stability parameter; these two parameters will be estimated from the propagation measurements. Same standard values for z_0 and α are used as in Equation (10). The function $\phi(\xi)$ is obtained,

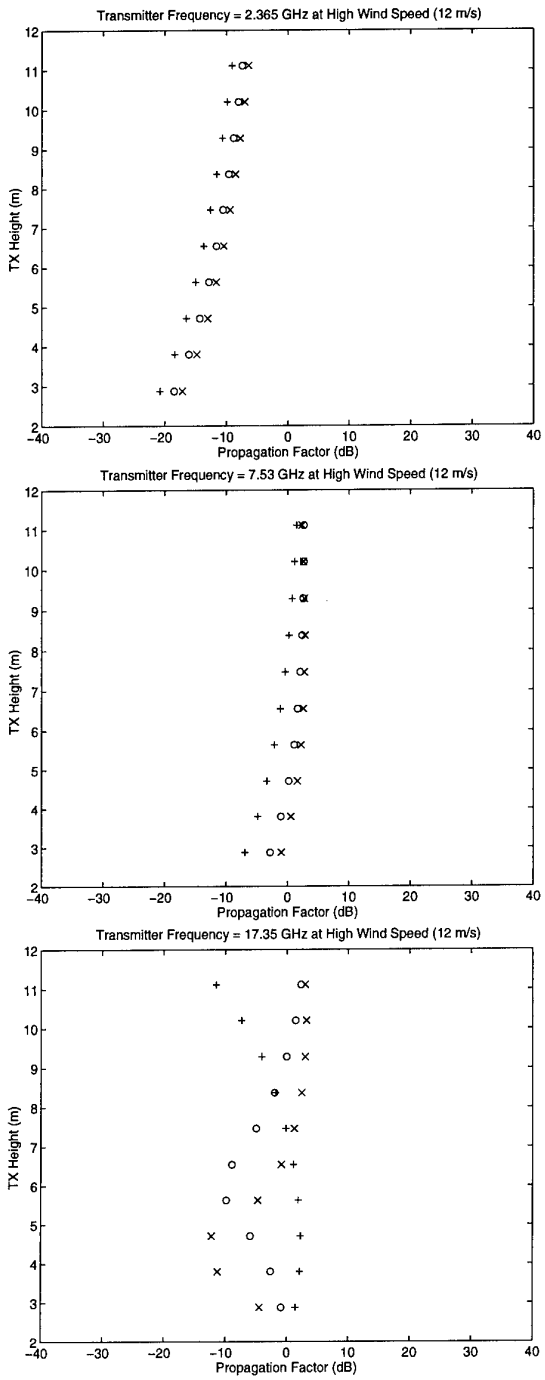


Figure 6: Receiver Height at 25.6 meters at high wind speed. The plus denotes ASTD = +2°, the circle ASTD = 0°, and the cross denotes ASTD = −2°.

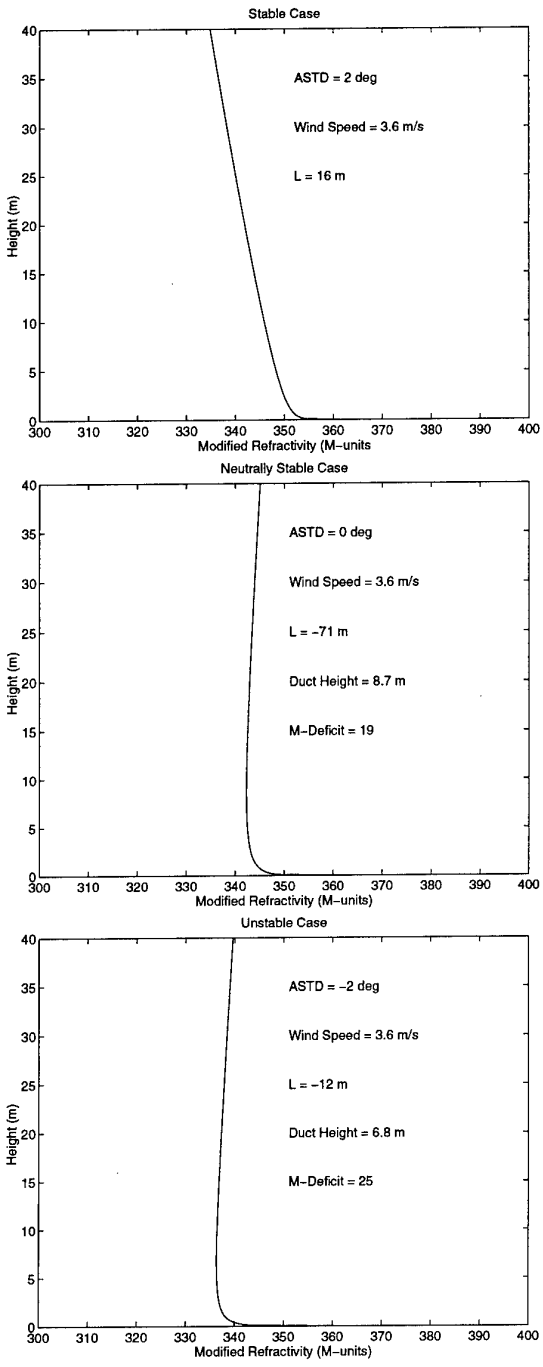


Figure 7: Modified Refractivity Profiles at low wind speed.

assuming that refractivity is a similarity variable, by

$$\phi(\xi) = \begin{cases} 1 + 8.6\xi, & \text{for } \xi > 0, \\ (1 - 12.2\xi)^{-1/2}, & \text{for } \xi < 0, \\ 1, & \text{for } \xi = 0. \end{cases}$$

This representation has some advantages over the LKB since it involves only two parameters, one of which, the duct height, is the key parameter that effects electromagnetic wave propagation in an evaporation duct. The LKB representation has the advantage that if it produces correct results, then it provides much more information about the structure of the surface layer, but it will surely be more difficult to estimate all of that meterological information uniquely from measured propagation data.

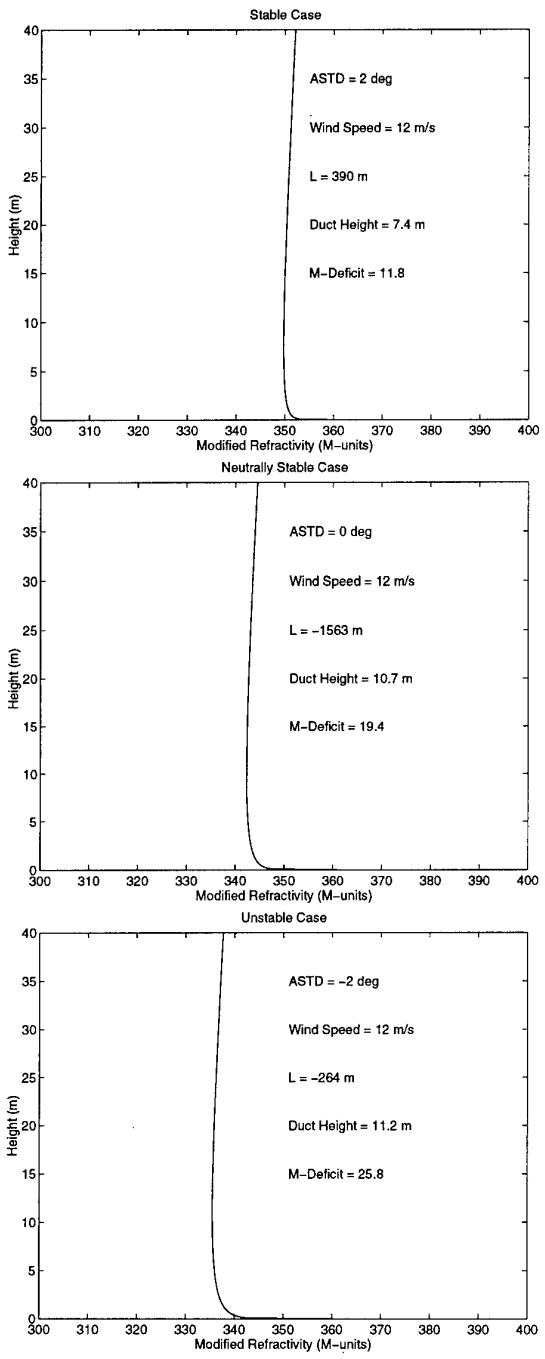


Figure 8: Modified Refractivity Profiles at high wind speed.

5. CRITERIA FOR A SUCCESSFUL REFRACTIVITY INVERSION

In the introduction it was observed that to determine the effectiveness of an estimation procedure it was necessary to have an accurate collection of propagation measurements. In addition, it would be desirable to possess some reasonable meteorological measurements collected over some portion of the timespan during which the propagation was measured. The propagation measurement system was in operation on a twenty four hour basis for three months, while the helicopter which measured the refractivity, and the boat that collected the surface meteorological measurements only operated during daylight hours on some days. Thus it is necessary to select a subset of the propagation data to which there corresponds a reasonable set of meteorological profiles and surface measurements. Once this propagation data subset is selected, then:

1. Input the refractivity profiles, for a selected time to the propagation model RPO, and obtain a vector of modeled propagation factors, $\mathbf{y}_M$, and using the corresponding measured propagation factors, $\hat{\mathbf{y}}$, compute the residual, $R_M = \frac{1}{2}(\mathbf{y}_M - \hat{\mathbf{y}})^T(\mathbf{y}_M - \hat{\mathbf{y}})$.
2. Select a reasonable refractivity parameterization, and execute the inversion procedure with a reasonable initial horizontally homogeneous approximate solution, to obtain a solution $\hat{\mathbf{x}}$.
3. Compute the residual between the estimated and the measured propagation $\hat{R} = R(\hat{\mathbf{x}})$ from Equation (3).
4. Compare $\hat{R}$ with R_M, and repeat steps 2, 3 and 4 with a different refractivity parameterization until all of the parameterization schemes have been tried.

If one of the residuals, $\hat{R}$, is less than R_M, then the inversion procedure was very successful. If one of the residuals is very close to the measured one, then the technique was successful. If each of the estimated residuals is significantly larger than R_M, then the inversion procedure did not perform as expected. One reason why the procedure could not reduce the residual significantly is most likely that the refractivity parameterization in terms of mean profiles is not adequate to capture the vertical variation in the atmosphere.

The adequacy of using mean profiles to determine refractivity in the surface layer has been questioned for years now, and more sophisticated parameterizations which incorporate turbulent fluctuations into the mean profiles are under investigation. It is hoped that these results can be incorporated into an improved inversion scheme.

6. CONCLUSIONS

A remote sensing technique for determining height and range varying refractivity fields from propagation loss measurements, using a numerical inversion procedure which incorporates a PE model, was discussed, and a synthetic example was used to illustrate the computational algorithms. The most important and difficult aspect of the problem is incorporating an adequate parameterization of the refractivity for the atmospheric surface layer into the procedure. The examples illustrate many of the types of

propagation conditions normally encountered. Additionally, these conditions can be modeled with a fairly simple refractivity profile represented in terms of a small number of parameters. The propagation data to which the inversion technique will be applied, as well as, the criteria for evaluating the success of such an application are discussed in some detail. The application of the inversion technique to Stapleton/Kang data described in this paper is underway and will soon be presented.

In order to investigate the dynamic character of the refractivity fields, the inversion technique can be applied repeatedly to generate a time series. In addition, one-dimensional and two-dimensional atmospheric boundary layer models will be executed to attempt to simulate the mean refractivity dynamics, and a large eddy simulation employed to model the turbulent fluctuations. The results of these efforts will be documented at a later time.

ACKNOWLEDGEMENT

The work is supported by the Office of Naval Research.

REFERENCES

1. Gingras, D.F., and P. Gerstoff, "Inversion for Geometric and Geoacoustic parameters in Shallow Water," *J. Acoust. Soc. Amer.*, 97 (6), June, 1995.
2. Stapleton, J., and S. Kang, "Direct Measurement of Microwave Propagation Effects," AGARD CP, 22-26 April 1996, Paper 12.
3. Queen, J.L., J. Stapleton, and S. Kang, "Wideband Low Elevation Microwave Propagation Measurements," NSWCDD/TR-95/18, February 1995.
4. Patterson, W.L., and H.V. Hitney, *Radio Physical Optics CSCI Software Documents*, NOSC Tech. Doc. 2403, December 1992.
5. Barrios, A.E., "Parabolic Equation Modeling in Horizontally Inhomogeneous Environments," *IEEE Transactions on Antennas and Propagation*, 40 (7), pp. 791-797, July 1992.
6. Cook, J., "A Sensitivity Study of Weather Data Accuracies on Evaporation Duct Height Algorithms," *Radio Science*, Vol. 26, No. 3, pp. 731-746, May-June, 1991.
7. Kuttler, J.R., and G.D. Dockery, "Theoretical Description of the Parabolic Approximation/Fourier Split-Step Method of Representing Electromagnetic Propagation in the Troposphere," *Radio Science*, pp. 381-393, March-April 1991.
8. Blanc, T.V., "Variation of Bulk-Derived Surface Flux, Stability, and Roughness Results Due to the Use of Different Transfer Coefficient Schemes," *Journal of Physical Oceanography*, Vol. 15, pp. 650-659, June, 1985.
9. Colton, D. and R. Kress, *Inverse Acoustic and Electromagnetic Wave Scattering Theory*, Springer Verlag, 1992.
10. Colton, D. and D. Monk, *The Detection and Monitoring of Leukemia Using Electromagnetic Waves*, in press.
11. Dennis, J.E. and R. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice Hall, 1983.
12. Rogers, D.P., D.W. Johnson and C.A. Friehe, "The Stable Internal Boundary Layer over a Coastal Sea. Part I: Airborne Measurements of the Mean and Turbulence Structure," *Journal of the Atmospheric Sciences*, Vol. 52, No. 6, 15 March 1995, pp. 667-683.
13. Babin, S.M. "A Case Study of Subrefractive Conditions at Wallops Island, Virginia," *J. Applied Meteorology*, Vol. 34, pp. 1028-1038, May, 1995.
14. Garratt, J.R., *The Atmospheric Boundary Layer*, Cambridge University Press, 1992.
15. Wyngaard, J.C., "On Surface Layer Turbulence,", *Workshop on Micrometeorology*, edited by D.A. Haugen, American Meteorological Society, 1973.
16. Patterson, W.L., et al, *Engineer's Refractive Effects Prediction System (EREPS) Revision 3.0*, NOSC Tech. Doc. 2648, Revision 3.0, May 1994, page 113.
17. Lui, T.L., K.B. Katsaros and J.A. Businger "Bulk Parameterization of Air-Sea Exchanges of Heat and Water Vapor Including the Molecular Constraints at the Interface," *J. Atmospheric Sciences*, Vol. 36, pp. 1722-1735, September, 1979.
18. Rotheram, S.A., "Radiowave Propagation in the Evaporative Duct," *Marconi Review*, Vol. 37, No. 192, pp. 18-40, 1974.

Paper 16

T. Rogers (US)

The inversion at evaporation duct meteorological parameter, as opposed to the inversion of just the evaporation duct height may make your inversion problem much more difficult. The Lui, Katseros and Bussinger (LKB) duct height is not a one-to-one function of the air-sea-temperature-difference (ASTD), so you are forcing yourself to solve two inverse problems instead of just one. This may transform the problem one where global optimization is required; you are currently using a local method.

Author's reply:

The issues you raise are very important. It may be much simpler to represent the refractivity in terms of a family of one parameter evaportation ducts, with the duct height as the parameter. The propagation factor may not be sensitive enough to ASTD or to wind speed, in order to force the gradient to zero fast enough for the descent method to not become confused. However, I think that it is worth attempting to obtain as much meteorological information from the propagation data as is possible. Multiple parameterizations will be attempted including surface based ducts and subrefractive layers, and the residuals will be compared.

ON THE IMPORTANCE OF UTILIZING POLARIMETRIC INFORMATION IN RADAR IMAGING AND CLASSIFICATION

E. Krogager[1], W.-M. Boerner[2]

[1]Danish Defence Research Establishment
Ryvangs Allé 1, DK-2100 Copenhagen, Denmark
T: +45 39272233 F: +45 31203315 E-mail: ek@ddre.dk

[2]University of Illinois at Chicago, UIC-EECS/CSL, M/C 154
840 W. Taylor St., SEL-4210, Chicago, Illinois 60607-7018, USA
T: +1 312.996.5480 F: +1 312.996.5480 E-mail: wolfgang.m.boerner@uic.edu

1. SUMMARY

Some basic polarization phenomenology and terminology, including the radar target scattering matrix, will be introduced. Significant advances of polarimetric radar over single polarization radar are pointed out by considering the response, i.e., the scattering matrix, for a set of significantly different elementary scatterers. The handling of polarimetric radar data requires special techniques. Examples of such techniques are explained, in particular the decomposition into physically relevant target constituents, and examples of actual applications to measured polarimetric SAR data are demonstrated.

2. INTRODUCTION

This paper is dealing with one of the special topics which have become important in modern radar, namely, polarimetric imaging techniques [1]. Although the fundamental capabilities of polarimetric radar, that is, radar utilizing the polarization of the transmitted and received waves, have been known since the early days of radar back in the fifties, it is only in recent years that such techniques have actually been implemented for practical utilization. There are a variety of reasons for that, some major reasons being the cost and complexity of polarimetric radars, but also an apparent lack of appreciation of what polarization can actually be used for. Not least within the NATO community, as regards operational systems, the utilization of information carried with the polarization of electromagnetic waves, is virtually nonexistent.

The importance of polarization must be emphasized for any radar system. In low resolution systems the extension from single polarization to full polarimetric capability virtually adds an extra dimension which of course improves the overall performance. In high resolution (imaging) systems the same holds true, possibly even to a wider extent. This is because the observable scattering characteristics become much more well-defined when each resolution cell contains only contributions from a scatterer of limited physical extent, whereby a much closer relation can be established between the gathered data and the actual nature of the target under observation.

It is our goal with this paper to try to point to the necessity of utilizing polarimetric information in order to optimize the use of radar for military as well as civilian applications. The paper is organized with several parts of theoretic background information. This, in particular the formulas, may be skipped by readers who rather are interested in the more phenomenologically descriptive parts as well as the application oriented discussions and actual examples.

3. BASIC FORMULATIONS

3.1 Plane Wave Polarization

At a specific instance of time, at a specific point in space, the direction and magnitude of an electric field is determined by the real part of a complex, time-harmonic field expression:

$$\vec{E}_{ire}(\vec{r},t) = \Re(\vec{E}_0 e^{j(\omega t - \vec{k}\cdot\vec{r})}) \qquad (1)$$

To characterize the actual behavior of the field vectors, a Cartesian coordinate system is fixed to the wave, so that the z-axis is parallel to and counted positive in the direction of propagation. The field vectors are then in the x-y plane:

$$\vec{E}_0 = E_{0x}\hat{x} + E_{0y}\hat{y} \qquad (2)$$

Depicted in the x-y plane at a fixed value of z, the tip of the electric field vector traces out an ellipse in the general case, when $\phi_x \neq \phi_y$, and the field is elliptically polarized. Special cases are *linear polarization*, when $\phi_x = \phi_y$, and *circular polarization*, when $|\phi_x - \phi_y| = 90°$ and $a_x = a_y$.

It is convenient to characterize the polarization ellipse in terms of two angles: the orientation angle around the z-axis, and the ellipticity angle (associated with the axial ratio). The overall magnitude, on the other hand, is not considered part of the polarization state. A general, normalized polarization vector $\vec{\mathbf{p}}$, characterizing a polarization ellipse with orientation angle θ ($-\pi \le \theta \le \pi$) and ellipticity angle τ ($-\pi/4 \le \tau \le \pi/4$), may be specified as,

$$\vec{p}_{\hat{x}\hat{y}}(\theta,\tau) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{pmatrix} \cos\tau \\ j\sin\tau \end{pmatrix} \qquad (3)$$

The $\hat{x}\hat{y}$ subscript is included here to emphasize that (3) is referred to an orthonormal linear basis. The sense of rotation of the vector tracing the ellipse is conveniently incorporated in the sign of τ, the ellipticity angle. By the IEEE convention, rotation is right-handed when an observer looking in the direction of propagation sees the field vector in a plane parallel to the x-y plane rotating clockwise with time. Consequently, ellipticity is negative for right-hand polarization. A possible initial phase is accounted for by a multiplicative complex scalar, $exp(j\phi_0)$, $-\pi \le \phi_0 \le \pi$. The following general representation then applies,

Paper presented at the AGARD SPP Symposium on "Remote Sensing: A Valuable Source of Information", held in Toulouse, France from 22-25 April 1996, and published in CP-582.

$$\vec{E}(z,t) = A\, e^{j\varphi_0}\hat{p}\, e^{j(\omega t - kz)} \tag{4}$$

Above, the wave polarization state was characterized by two parameters, the ellipticity angle and the orientation angle of an ellipse. The same amount of information can be represented by a single complex number. This is the idea behind the polarization ratio formalism, according to which a polarization state may be completely described in terms of the complex ratio between orthogonal components.

In the *linear basis*, the complex polarization ratio is thus defined as

$$\rho_{\hat{x}\hat{y}} = \frac{E_y}{E_x} = \frac{a_y}{a_x} e^{j(\varphi_y - \varphi_x)} = \left[\frac{\sin^2\theta\cos^2\tau + \cos^2\theta\sin^2\tau}{\cos^2\theta\cos^2\tau + \sin^2\theta\sin^2\tau}\right]^{\frac{1}{2}} \exp\left(j\arctan\left[\frac{\tan 2\tau}{\sin 2\theta}\right]\right) \tag{5}$$

while the polarization ratio in the circular basis is found to be

$$\rho_{\hat{R}\hat{L}}(\theta,\tau) = j\frac{\cos\tau + \sin\tau}{\cos\tau - \sin\tau} e^{j2\theta} = \tan(\tau + \tfrac{\pi}{4}) e^{j2(\frac{\pi}{4} - \theta)} \tag{6}$$

3.1.1 Polarization vector transformation under change of basis
Transformation of the coordinates of a polarization vector from one orthonormal elliptical basis, $\mathscr{E}_1$, to another orthonormal elliptical basis, $\mathscr{E}_2$, is accomplished by a unitary transformation matrix, whereby

$$\begin{aligned}\vec{p}_2(\mathscr{E}_2) &= \left[U(\mathscr{E}_1 \to \mathscr{E}_2)\right]\vec{p}_1(\mathscr{E}_1) \\ &= \left[\hat{u}^1_{\mathscr{E}_1}(\mathscr{E}_2)\ ,\ \hat{u}^2_{\mathscr{E}_1}(\mathscr{E}_2)\right]\vec{p}_1(\mathscr{E}_1)\end{aligned} \tag{7}$$

in which the column unity vectors of [U] are the basis vectors of $\mathscr{E}_1$, the old basis, specified in $\mathscr{E}_2$, the new basis, and $\hat{p}_1$ is the polarization vector specified in the old basis, $\mathscr{E}_1$.
[U] is a unitary Hermitian matrix, and so its inverse is given as

$$[U]^{-1} = [U]^{*T} \tag{8}$$

Consequently,

$$\begin{aligned}\vec{p}_1(\mathscr{E}_1) &= \left[U(\mathscr{E}_2 \to \mathscr{E}_1)\right]\vec{p}_2(\mathscr{E}_2) \\ &= \left[U(\mathscr{E}_1 \to \mathscr{E}_2)\right]^{*T}\vec{p}_2(\mathscr{E}_2)\end{aligned} \tag{9}$$

Transformation from a linear basis to a general elliptical basis is thus obtained by applying both a rotation and an ellipticity transformation matrix, [2],[3],

$$[U(\theta,\tau)] = \begin{bmatrix}\cos\alpha & -j\sin\alpha \\ -j\sin\alpha & \cos\alpha\end{bmatrix}\begin{bmatrix}\cos\theta & \sin\theta \\ -\sin\theta & \cos\theta\end{bmatrix} \tag{10}$$

3.2 The Scattering Matrix

To completely characterize the scattering properties of a radar target, a matrix rather than a scalar is required, due to the vector nature of the electromagnetic field. Hence, the general expression for the field received by a radar upon reflection at a given target, is given by an expression of the following form,

$$\begin{aligned}\vec{E}^r &= \begin{pmatrix}E_1^r \\ E_2^r\end{pmatrix} = \begin{bmatrix}S_{11} & S_{12} \\ S_{21} & S_{22}\end{bmatrix}\begin{pmatrix}E_1^t \\ E_2^t\end{pmatrix}\frac{1}{\sqrt{4\pi r^2}} \\ &= [S]\vec{E}^t\frac{1}{\sqrt{4\pi r^2}}\end{aligned} \tag{11}$$

where S_{ij} are complex quantities describing the reflective properties of the target for a given aspect angle and carrier frequency. [S] is known as the (Sinclair) scattering matrix, and completely describes the polarization transforming properties of the target at a single frequency in the reference direction [4].

3.2.1 Scattered field in radar fixed coordinate system
From a transmitted field specified in the coordinate system associated with the transmit antenna, we seek the scattered field at the receive antenna in the coordinate system associated with the receive antenna. However, we shall consider only the backscattering situation where both antennas are colocated, at least so that the bistatic angle between them is very small. It will therefore suffice here to consider only the situation where the coordinate system associated with the receive antenna is the same as that for the transmit antenna. For the transmitted field an expression of the following form applies,

$$\vec{E}_t(r) = A_T e^{j\varphi_T}\frac{e^{-jkr}}{\sqrt{4\pi r^2}}\vec{p}_T \tag{12}$$

At the target, at a distance of r_0, a backscattered wave is generated, determined by the complex coefficients of the scattering matrix,

$$\vec{E}_s = A_T e^{j\varphi_T}\frac{e^{-jkr_0}}{\sqrt{4\pi r_0^2}}\begin{bmatrix}S_{11} & S_{12} \\ S_{21} & S_{22}\end{bmatrix}\vec{p}_T \tag{13}$$

From there on, the field is inward in the coordinate system we have chosen. By conjugating the polarization state of the wave, we thus find for the far-field of the spherical wave with origin at the target,

$$\vec{E}_i(r) = A_T e^{j\varphi_T} A_P e^{j\varphi_P}\begin{bmatrix}S_{11} & S_{12} \\ S_{21} & S_{22}\end{bmatrix}^*\vec{p}_T^* \tag{14}$$

where the propagation factor has been represented by the amplitude and phase quantities with subscript P.

3.2.2 The voltage equation
The voltage at the terminals of the radar's receiver due to the signal that was originally transmitted by the radar's transmit antenna, and subsequently reradiated by the target characterized by a scattering matrix, [S], is given by,

$$\begin{aligned}V_r &= A_R e^{j\varphi_R}\vec{p}_R^*\cdot\vec{E}_i \\ &= A_P A_R A_T e^{j(\varphi_{0_P} + \varphi_{0_R} + \varphi_{0_T})}\langle\vec{p}_R, \vec{p}_s\rangle\end{aligned} \tag{15}$$

where

$$\vec{p}_s = [S]^*\vec{p}_T^* \tag{16}$$

and

$$\langle \vec{p}_R, \vec{p}_s \rangle \equiv \vec{p}_R^{\,*} \cdot \vec{p}_s \qquad (17)$$

defines the inner product of the polarization of the receiving antenna and the polarization of the scattered field, referenced to the radar coordinate system.

3.2.3 Transformation under a change of basis

For the polarization scattered by a target, referenced to a given basis 1, we have,

$$\vec{p}_s^{\,1} = [S^1]^* \vec{p}_T^{\,1*} \qquad (18)$$

Now, with a second basis, basis 2, we have

$$\vec{p}_T^{\,2} = [U]\vec{p}_T^{\,1} \;\Rightarrow\; \vec{p}_T^{\,1} = [U]^{-1}\vec{p}_T^{\,2} \qquad (19)$$

and

$$\vec{p}_s^{\,2} = [U]\vec{p}_s^{\,1} \;\Rightarrow\; \vec{p}_s^{\,1} = [U]^{-1}\vec{p}_s^{\,2} \qquad (20)$$

where [U] is the unitary matrix for the change of basis, cf. Section 3.1.1.

In basis 2 we seek an expression for the scattering process in terms of a scattering matrix, $[S^2]$, so that the same physical quantities are related by,

$$\vec{p}_s^{\,2} = [S^2]^* \vec{p}_T^{\,2*} \qquad (21)$$

One finds, $$[S^2] = [U]^* [S^1] [U]^{*T} \qquad (22)$$

a unitary congruent transformation yielding the desired relation between $[S^1]$ and $[S^2]$.

This relation, i.e., the possibility of transforming quantities from one basis to another, opens up for a variety of important operations and analysis tools in radar polarimetry. In particular, this means that once the complete scattering matrix has been specified (measured) in one basis, one can easily calculate (synthesize) the corresponding matrix (target response) in any other basis without actually measuring quantities in that basis.

3.2.4 The pseudo-eigenvalue problem for the scattering matrix

The eigenvalue problem associated with the scattering matrix was first considered by Kennaugh [4], and later treated in detail by notably Huynen [5] and Boerner [6]. Huynen rigorously formulated the existence of maximum and minimum (null) polarizations for the scattering matrix and pointed out how they form the so-called Huynen-fork on the Poincaré sphere.

In the present case where the operator is the complex monostatic radar target scattering matrix, the characteristic problem is to find $\vec{p}_s$ so that maximum received voltage is obtained. Thus, one can see that the condition for maximum received voltage is that $\vec{p}_s$ equals $\vec{p}_R$ within a proportionality factor. As proved by Kennaugh [7], this goal may be reached with identical receive and transmit polarization, which in practical terms means that the radar needs not have separately variable polarization on transmit and receive in order to maximize the target return. Setting therefore $\vec{p}_s = \lambda \vec{p}_R$ and $\vec{p}_R = \vec{p}_T$, we find

$$[S]^* \vec{p}_T^{\,*} = \vec{p}_s = \lambda \vec{p}_R = \lambda \vec{p}_T \qquad (23)$$

from which the characteristic pseudo-eigenvalue equation follows,

$$[S]\vec{p}_T = \lambda \vec{p}_T^{\,*} \qquad (24)$$

It should be noted that unlike traditional eigenvalue problems, a complex conjugate appears on one side of (24), hence the designation pseudo-eigenvalue problem.

It is customary to represent the diagonalized matrix, associated with the eigenvalues, in the following form,

$$[S_d] = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} = \begin{bmatrix} |\lambda_1| e^{j\varphi_1} & 0 \\ 0 & |\lambda_2| e^{j\varphi_2} \end{bmatrix} \qquad (25)$$

where λ_1 and λ_2 are the two pseudo-eigenvalues that satisfy (24), ordered without loss of generality such that $|\lambda_1| \geq |\lambda_2|$. The corresponding eigenvectors are orthogonal and span the elliptic basis in which the scattering matrix is diagonal.

Following Huynen's notation, except for the absolute phase, we may write for the eigenvalues:

$$\begin{aligned} \lambda_1 &= m\, e^{2j(\nu+\zeta)} \\ \lambda_2 &= m \tan^2\gamma\, e^{-2j(\nu-\zeta)} \end{aligned} \qquad (26)$$

whereby

$$[S_d] = m e^{j2\zeta} \begin{bmatrix} e^{j\nu} & 0 \\ 0 & e^{-j\nu} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \tan^2\gamma \end{bmatrix} \begin{bmatrix} e^{j\nu} & 0 \\ 0 & e^{-j\nu} \end{bmatrix} \qquad (27)$$

Thus, the symmetric, complex scattering matrix, containing in total 6 independent parameters in its elements, has now been represented by another 6 parameters, the Huynen-Euler parameters:

- m — ($m \geq 0$), the maximum polarization, which is the maximum attainable response from the target, obtained when the optimum polarization, associated with the largest eigenvalue, is used by the radar; this parameter represents the true magnitude (in terms of radar cross section) of the target.
- ζ — ($-180° \leq \xi \leq 180°$), the absolute phase of the scattering matrix.
- ψ — ($-90° \leq \psi \leq 90°$), the orientation angle of the target, determining the orientation angle of the optimum polarization for the target
- τ_m — ($-45° \leq \tau_m \leq 45°$), the helicity angle, i.e., the ellipticity of the optimum polarization for the target
- ν — ($-45° \leq \nu \leq 45°$), the skip angle; if the return is due to scattering mechanisms with an even number of reflections (bounces), ν will be equal to 45°; however, it should be observed that targets can have $\nu = 45°$, even when only part of the reflected signal is due to even-bounce scattering.
- γ — ($0° \leq \gamma \leq 45°$), the characteristic (or, polarizability [8]) angle; targets with $\gamma = 45°$ do not repolarize the incident wave while targets with $\gamma = 0°$ completely determine the polarization of the returned wave.

3.3 Power Reflection Matrices

3.3.1 Kennaugh matrix

The above formulas refer to measured voltages, i.e., to complex quantities. In many cases it is more relevant to deal with real quantities related to measured powers. Especially with statistically varying fields this is convenient, and the statistical independencies between different contributions have important consequences. Such a description is available with the Stoke's reflection matrix, also known as the Kennaugh matrix.

To establish the relation between the Kennaugh matrix and the Sinclair matrix, the following notation, known as the Pauli spin matrix decomposition, is introduced,

$$[\boldsymbol{S}] = \begin{bmatrix} a+b & c \\ c & a-b \end{bmatrix} = a\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + b\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} + c\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \tag{28}$$

Then, following Huynen's notation,

$$[\boldsymbol{K}] = \begin{bmatrix} A_o+B_o & F & C & H \\ F & -A_o+B_o & G & D \\ C & G & A_o+B & E \\ H & D & E & A_o-B \end{bmatrix} \tag{29}$$

whose elements are related to the elements of [S] as follows,

$$\begin{aligned} |a|^2 &= 2A_o, \quad ab^* = C-jD \\ |b|^2 &= B_o+B, \quad bc^* = E+jF \\ |c|^2 &= B_o-B, \quad ac^* = H+jG \end{aligned} \tag{30}$$

3.3.2 Covariance matrix

To characterize the polarization state of a plane wave, the complex polarization 2-vector, a real Stokes 4-vector, and a 2x2 coherency matrix are frequently encountered. In quite a similar manner, it is possible for a radar target characterized by a (symmetric) complex 2x2 Sinclair matrix or a real 4x4 Kennaugh matrix to construct a complex 3-vector and a 3x3 covariance matrix.

Two different approaches are frequently encountered, one which defines the 3-vector directly in terms of the elements of the scattering matrix, another which defines the 3-vector in terms of the Pauli spin matrix decomposition. The latter representation which, according to Huynen [9], should be preferred, is given by

$$[\boldsymbol{C}] = (a,b,c)^T(a,b,c)^* = \begin{bmatrix} |a|^2 & ab^* & ac^* \\ ba^* & |b|^2 & bc^* \\ ca^* & cb^* & |c|^2 \end{bmatrix} \tag{31}$$

Like the Kennaugh matrix, the covariance matrix is particularly suited for statistically varying fields, and it contains exactly the same information as the Kennaugh matrix which is clear from (30).

3.3.3 Other characteristic polarizations

In the above sub-sections we considered exclusively the diagonal form of the scattering matrix and the various methods for obtaining this form which provides the so-called cross-pol nulls and co-pol max polarization states. However, also other characteristic polarizations exist, in particular the copol-nulls [5],[4],[6],[10]. These may be found by solving for $S_{11}=0$ and $S_{22}=0$ instead of $S_{12}=0$ in the transformed matrix. In addition, there are the cross-pol max polarizations and cross-pol saddle-point extrema, of which the latter have only recently been discovered [11]. However, these other characteristic (or optimal) polarizations will not be further considered in the present treatment.

4. POLARIMETRIC SCATTERING BEHAVIOR OF DIFFERENT TARGETS

4.1 Canonical targets

One of the most convincing ways to consider the potential of polarimetric radar is to consider the response of some different types of elementary single scatterers. Such different scatterers with the associated different scattering mechanisms produce very different responses to a polarimetric radar while they may be impossible to distinguish by an ordinary single polarization radar.

4.1.1 Sphere

Starting with the sphere, it has the identity matrix as its characteristic scattering matrix,

$$[S]_{sphere} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{32}$$

with equal responses for HH and VV polarizations. This matrix is characteristic for any odd-bounce scatterer, like plates and tri-hedrals, i.e., three-sided corner reflectors.

4.1.2 Line target (dipole)

A dipole or thin wire target has the following characteristic scattering matrix which contains the orientation angle, θ, of the dipole around the line of sight,

$$[\boldsymbol{S}]_{wire} = \begin{bmatrix} \cos^2\theta & \frac{1}{2}\sin 2\theta \\ \frac{1}{2}\sin 2\theta & \sin^2\theta \end{bmatrix} \tag{33}$$

which means that this angle can actually be measured with a polarimetric radar. For this type of target, it makes a great deal of difference, whether HH or VV polarizations are used, while a polarimetric radar will always measure the same total power, independent of whether linear, circular, or any other reference polarizations are used.

4.1.3 Diplane

The diplane, i.e., the two-sided corner reflector, also contains the orientation angle around line of sight, and we note for this type of target a characteristic phase difference between the two co-polar responses, which is characteristic for any even-bounce reflector,

$$[S]_{diplane} = \begin{bmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{bmatrix} \quad (34)$$

We further note, that the combined response from a sphere plus a diplane yields the response from a dipole, which means that for more general, complex targets we must expect to see significant differences between the single polarization HH and VV components.

4.1.4 Helix

The helix is shown here as a more peculiar type of target with the characteristics that it transforms incident linear polarization to reflected circular polarization,

$$[S]_{helix} = \frac{1}{2}\begin{bmatrix} 1 & \pm j \\ \pm j & -1 \end{bmatrix} \quad (35)$$

Although the helix is a somehow artificial target, it should be observed that the helix scattering matrix can also be produced by two diplanes with a relative orientation of 45° and a relative displacement along line of sight of 1/8 wavelength. This means that for practical targets with two or several double-bounce scatterers in a resolution cell, helix-like scattering may also be found.

For the present purpose the above selection of different elementary targets will suffice, and it should be quite clear at this point, that a polarimetric radar is able to measure a great deal of target characteristics that would never be revealed by a single polarization radar. The single polarization radar may even happen to get zero or near zero response from a target which otherwise has a significant effective radar cross section. An important (basis invariant) parameter that is provided by a polarimetric radar is the total power, rather than the ordinary RCS (Radar Cross Section) given by just one of the elements of the scattering matrix. This is also known as the span of the scattering matrix

$$\Upsilon = Sp([\boldsymbol{S}]) = |S_{11}|^2 + |S_{22}|^2 + 2|S_{12}|^2 \quad (36)$$

which, as already mentioned, has the powerful feature of being an invariant, i.e., its value does not depend on in which basis the quantities are measured.

5. TARGET DECOMPOSITION TECHNIQUES

5.1 Background - the Huynen and other decompositions

A unique feature provided by polarimetric data is the possibility to separate scattering contributions of different nature. For this purpose several decomposition techniques have been proposed. The first such decomposition was presented by Huynen [5]. The basic philosophy was to decompose the (averaged) Kennaugh matrix into a sum of an average single target and a distributed residue component, the so-called N-target, where N refers to, but should not be too literally regarded as, noise. This decomposition is relevant to situations with fluctuating fields where statistical independence of the components implies that powers can be incoherently added. A special property is the orientation invariance which means that the forms of the component matrices are invariant under changes of orientation around the radar line of sight. More recently, various other approaches to incoherent decomposition have been presented [12],[13],[14],[15],[16],[17]. The interested reader is referred to the references. A recent review of target decomposition theorems in radar polarimetry may be found in [18].

5.2 Decomposition of the Sinclair matrix into 3 components: sphere, diplane, helix

A detailed description of the sphere, diplane, helix decomposition as applied in this paper is given elsewhere [2], [19],[20], but for easy reference the key formulas are repeated here. In the roll-invariant formulation, the sphere, diplane, helix decomposition is given by the following general representation,

$$[\boldsymbol{S}] = e^{j\varphi}\left\{e^{j\varphi_s}k_s[\boldsymbol{S}]_s + k_d[\boldsymbol{S}]_{d(\theta)} + k_h[\boldsymbol{S}]_{h(\theta)})\right\} \quad (37)$$

Due to the properties of this decomposition, appearing from Table 1 of responses of the three elementary targets to a circularly polarized system, it is convenient to deal with quantities in the circular basis.

In the circular basis, a general symmetric ($S_{RL} = S_{LR}$) scattering matrix, $[S_C]$, is given as,

$$[S_C] = \begin{bmatrix} S_{RR} & S_{RL} \\ S_{RL} & S_{LL} \end{bmatrix} = \begin{bmatrix} |S_{RR}|\, e^{j\varphi_{RR}} & |S_{RL}|\, e^{j\varphi_{RL}} \\ |S_{RL}|\, e^{j\varphi_{RL}} & |S_{LL}|\, e^{j\varphi_{LL}} \end{bmatrix} \quad (38)$$

Table 1

	LL	RR	LR
Sphere	NO	NO	YES
Diplane	YES	YES	NO
Right helix	YES	NO	NO
Left helix	NO	YES	NO

The following expressions for the parameters of the decomposition may then be found [19],[20],

$$\begin{aligned} k_s &= |S_{RL}| \\ k_d^+ &= |S_{LL}| \;;\; k_d^- = |S_{RR}| \\ k_h^+ &= |S_{RR}| - |S_{LL}| \;;\; k_h^- = |S_{LL}| - |S_{RR}| \\ \varphi &= \tfrac{1}{2}(\varphi_{RR} + \varphi_{LL} - \pi) \\ \theta &= \tfrac{1}{4}(\varphi_{RR} - \varphi_{LL} + \pi) \\ \varphi_s &= \varphi_{RL} - \tfrac{1}{2}(\varphi_{RR} + \varphi_{LL}) \end{aligned} \quad (39)$$

As appears, S_{RR} or S_{LL} represents directly the diplane component, according as $|S_{LL}| > |S_{RR}|$ or $|S_{RR}| > |S_{LL}|$, respectively. Correspondingly, the helix component is of either right or left sense, respectively.

In order to compute the decomposition in terms of quantities given in the linear basis, the following relations are useful,

$$S_{RR} = j S_{HV} + \frac{1}{2}(S_{HH} - S_{VV})$$
$$S_{LL} = j S_{HV} - \frac{1}{2}(S_{HH} - S_{VV}) \quad (40)$$
$$S_{RL} = \frac{j}{2}(S_{HH} + S_{VV})$$

As an alternative to the sphere, diplane, helix decomposition, a decomposition into a sphere and two helices (left and right, respectively) has also been considered [19],[20]. This corresponds to merely taking the magnitudes of the elements of the scattering matrix in a circular basis, say RR, RL, and LL, and thus corresponds to using directly the elements of the matrix in a linear basis, say HH, HV, and VV. However, unlike the linear basis components the circular basis components have the important property of being roll invariant, and more closely related to physical scattering mechanisms.

6. APPLICATION OF POLARIMETRIC TECHNIQUES TO SYNTHETIC APERTURE RADAR (SAR) DATA

6.1 EMISAR data

The data used for the results presented with this paper have been acquired with the Danish polarimetric SAR system which has been developed at the Electromagnetics Institute (EMI), Technical University of Denmark (TUD) [21],[22],[23]. This system, EMISAR, now makes up the key instrument for research at the Danish Center for Remote Sensing (DCRS) which was established at EMI in 1994.

The system, originally a C-band single polarization system, was upgraded to full polarimetric capability in 1993, and the performance has been demonstrated through numerous measurement campaigns. A remarkable feature of this civilian radar is its high resolution of 2 m in both range and cross-range together with the full-polarimetric, wide-swath capability, which makes it a powerful tool for many demanding applications, including classification and identification of natural and man-made targets.

A fully polarimetric L-band channel was added to the EMISAR system in 1995, and successfully tested during several missions in 1995. This has added significantly to the capabilities and the utility of the system, because the polarimetric scattering characteristics of a target are in general highly wavelength dependent [24].

Also, interferometric modes are presently implemented. A repeat-track mode in which the radar steers the aircraft enables the aircraft to fly a desired track with meter accuracy, and additional antennas are installed to provide for single pass across-track interferometry [25].

An important issue in connection with the use of fully polarimetric data is the quality of system calibration, since a correct interpretation of relative phases and amplitudes of the scattering matrix can only be made with a sufficiently accurate calibration. To ensure this, the present system accommodates unique features for internal calibration and for ensuring system stability. External calibration is carried out just before and after a mapping sequence, and calibration reflectors as well as distributed targets may be used to verify and improve the calibration [26].

6.2 Results and discussion

The images considered are generated by using various sets of parameters to modulate the rgb color-process, cf. Table 2.

Table 2

	Red	Green	Blue
LIN	HV	HH	VV
CIRC	RL	RR	LL
SDH	k_s	k_d	k_h
SIGHEL	left k_h	right k_h	-

The three sets of color-modulation parameters, denoted in Table 2 as LIN, CIRC, and SDH, all represent useful information for discrimination and contrast improvement in high resolution polarimetric SAR imaging. From the images analyzed so far, LIN and SDH appear to yield the best discrimination between different target types in the images. Both representations seem to have certain advantages for particular targets. Thus, the ability of LIN to distinguish targets with predominant HH, HV, or VV returns provides an efficient way of revealing target characteristics which favor the generation (scattering) of a specific linear polarization, related to target geometry. For example, it has been found that wind causes a significant increase in the amount of cross-polarized scattering from corn fields in the linear basis. On the other hand, the parameters of LIN are not directly related to a specific scattering mechanism, as are the parameters of SDH. For example, a horizontal or a vertical diplane (commonly found in areas with man-made targets) produces HH and VV components of equal magnitudes, which is also the case for an odd-bounce target like a sphere, a plate, or a trihedral. Hence, LIN will not discriminate between these two otherwise very different types of targets. To separate these targets, LIN must be used in conjunction with for example a plot of the phase difference between HH and VV. In the decomposition, this is already accomplished because the parameters of the decomposition incorporate the phases as well as the amplitudes of the elements of the linear basis matrix. The orientation angle, however, is hidden in the other two parameters in this approach and would have to be displayed in a different type of image.

It should be noted that only the three component magnitudes are used in these images. Additional information is actually present in the polarimetric data in the form of parameters like phases and orientation angles. For example, the phase difference between HH and VV returns may be utilized to generate images where certain types of scatterers stand out. Thus, a sphere has no phase difference between HH- and VV-components, while a diplane has a phase difference of π between these components.

The method denoted SIGHEL (signed helicity) aims at separating the helix-component of the sphere, diplane, helix decomposition into the two categories: left and right sense. Thereby it becomes possible to investigate the distribution between left and right sense of the helix-component. For distributed, natural targets with azimuthal symmetry, the correlation between co- and cross-polarized components in the

linear basis is usually assumed to be zero, and helicity might be expected to be evenly distributed between left and right sense in such areas. The more general conditions under which this correlation is zero is a subject of considerable current interest due to the relevance in connection with SAR cross-talk calibration on the basis of distributed targets for which the true co- and cross-polarized returns are assumed to be uncorrelated. However, it has been found that certain types of targets exhibit consistently one dominating sense of helicity, depending on carrier wavelength, which implies that for such areas the assumption may not apply [2],[27]. This observation has for instance been made in agricultural areas, and is likely to be related to the fact that various crop types have distinct preferences and characteristics with regards to twine and composition [28]. However, further research is required before any firm conclusions can be reached.

In order to keep this paper in grayscale print, examples of the images discussed above are included on the last pages in the form of a black and white print of an rgb image as well as the corresponding separated red, green, and blue component images. Although they do not provide the same overview of the data as a one-page color-composite image would do, some of the features provided by polarimetric processing are clearly seen. The scene is from the agricultural research center at Foulum, Denmark, with the research center located in the lower left corner of the images, and a variety of farm fields, woods, etc. are seen around the center.

6.3 Possible improvement of SAR and ISAR image generation by the use of fully polarimetric data

Several problems are still identified in connection with the SAR and ISAR (Inverse SAR) imaging technique for general non-cooperative targets and scenarios. The two key elements of the ISAR principle are a motion compensation and a rotation compensation which are required to obtain a sufficiently high cross-range resolution. Both of these pose a major problem because the phase information required for carrying out these compensations is difficult to establish with sufficient reliability.

It is of interest to consider whether the use of fully polarimetric data could be advantageous in this connection. In the light of the decomposition of the scattering matrix presented in subsection 5.2, it is apparent that the chance of localizing suited individual scatterers would increase significantly if fully polarimetric data were available for each resolution cell. Thus, instead of just searching for a resolution cell with a strong scatterer, one could then search for a resolution cell with a pure type of scattering, either of the odd-bounce or the even-bounce type. If such a resolution cell can be found, chances are much better that the corresponding back-scattering is due to a distinct single scatterer, which is required in order to obtain a well suited phase history for motion compensation. Similar considerations can be applied to the other data corrections that were mentioned above.

Although two-dimensional imaging of course provides a much better visualization of the radar reflectivity distribution of a given target, there may be situations where it is only practical to implement either the cross-range resolution or the range resolution. The first requires only a CW radar while the latter avoids the long integration times. In such cases, polarization makes up a useful option for extending the capability of a given system so as to provide as much information about the target as possible.

The utilization of polarimetric information should therefore be expected to significantly enhance the basic processing and thereby the overall image quality, in addition to providing a lot of inherent extra information about the target.

7. CONCLUSIONS

The fundamentals of polarimetric techniques in relation to high resolution radar applications were reviewed, followed by a presentation of a special three-component decomposition of the scattering matrix. By decomposing the scattering matrix into three different components, it is possible to resolve different types of scatterers even if they are within the same resolution cell of the image. This allows for a better resolution of the target scattering properties as well as a better characterization of the type of scattering for the individual contributions. The usefulness of fully polarimetric data in connection with radar target imaging was demonstrated by applying polarimetric techniques to fully polarimetric, well calibrated EMISAR data. The components of the decomposition may be imaged separately, and in combination by utilizing the three components to modulate a color-mixing process. Both representations are found to be powerful tools for handling and interpreting polarimetric imaging data. A particularly troublesome problem associated with dynamic radar imaging is the various phase compensations that must be carried out before the desired images are obtained. It was pointed out that polarimetric data should be used in such procedures and should be expected to greatly enhance this part of the processing, in addition to providing a lot of extra target related information as compared with single polarization data. Hence, fully polarimetric data should be expected to greatly improve the possibility of successful radar target discrimination in a wide variety of applications.

8. ACKNOWLEDGMENTS

During the presentation of the paper, some further examples of polarimetric SAR images will be shown, courtesy of the Danish Center for Remote Sensing (DCRS), Electromagnetics Institute, Technical University of Denmark, c/o Profs. S.N. Madsen and E. Lintz Christensen. Thanks are also due to the following personnel of DCRS, Dr. J. Dall, J. Grinder-Pedersen, S.S. Kristensen, and J. Mohr, for assistance in accessing, handling, and processing the data on DCRS facilities. DCRS is financed by the Danish National Research Foundation.

9. REFERENCES

1. Boerner, W.-M., et al. eds., Inverse Methods in Electromagnetic Imaging Polarimetry, Proc. NATO-ARW-DIMRP (W.-M. Boerner, Director), 1988 Sept. 18-24, Bad Windsheim, FRG, NATO-ASI-Series C, (Math. & Phys. Sci.), Vol. 350, Part I (pp. 1-1064) and II (pp. 1065-1938), D. Reidel Publ. Co., Dordrecht/Boston, 1992.

2. Krogager, E., *Aspects of Polarimetric Radar Imaging*, Doctoral Thesis,Technical University of Denmark, May 1993 (Danish Defence Research Establishment, P.O.Box 2715, DK-2100 Copenhagen, Denmark).

3. Xi, A.-Q., and W.-M. Boerner, "Determination of the characteristic polarization states of the radar target scattering matrix [S(AB)] for the coherent monostatic and reciprocal propagation space by using the complex polarization ratio transformation formulation", *Journal of the Optical Society of America*, A, vol. 9, no. 3, pp. 437-455, Mar. 1992.

4. Kennaugh, E.M., *Effects of Type of Polarization on Echo Characteristics*, The Ohio State University, Antenna Laboratory, Columbus, OH, Reports 381-1 to 394-24, 1949-1954.

5. Huynen, J.R., *Phenomenological Theory of Radar Targets*, Doctoral Thesis, Technical University, Delft, The Netherlands. Rotterdam: Bronder-Offset, 1970.

6. Boerner, W.-M., M.B. El-Arini, C.-Y. Chan, and P.M. Mastoris, "Polarization Dependence in Electromagnetic Inverse Problems", *IEEE Trans. Antennas Propagat.*, vol. 29, no. 2, pp. 262-271, Mar. 1981.

7. Kennaugh, E.M., *Effects of Type of Polarization on Echo Characteristics*, The Ohio State University, Antenna Laboratory, Columbus, OH, Report 389-12, p. 14, 1952.

8. Holm. W.A., "MMW Radar Signal Processing Techniques", in N.C. Currie and C.E. Brown (eds.), *Principles and Applications of Millimeter-Wave Radar*, Artech House, Norwood, MA, 1987.

9. Huynen, J.R., "Physical Reality and Mathematical Process in Radar Polarimetry", ICAP 1991, *IEE Conference on Antennas and Propagation*, York, England, 257-261, Apr. 1991.

10. Agrawal, A.P., and W.-M. Boerner, "Redevelopment of Kennaugh's Target Characteristic Polarization State Theory Using the Polarization Transformation Ratio Formalism for the Coherent Case", *IEEE Trans. Antennas Propagat.*, vol. 27, no. 1, pp. 2-14, Jan. 1989.

11. Boerner, W.-M., and A.-Q. Xi, "The Characteristic Radar Target Polarization State Theory for the Coherent Monostatic and Reciprocal Case Using the Generalized Polarization Transformation Ratio Formulation", *Archiev für Elektronik und Übertragungstechnik*, AEÜ, vol. 44, no. 4, pp. 273-281, 1990.

12. Cloude, S.R., "Target Decomposition Theorems in Radar Scattering", *Electronics Letters*, 21, 1, 22-24, Jan. 1985.

13. Holm, W.A. and Barnes, R.M., "On Radar Polarization Mixed Target State Decomposition Techniques", *IEEE National Radar Conference*, 249-254, 1988.

14. van Zyl, J.J., "Unsupervised Classification of Scattering Behavior Using Radar Polarimetry Data", *IEEE Transactions on Geoscience and Remote Sensing*, 27, 1, 36-45, 1989.

15. Cameron, W.L. and L.K. Leung, "Feature motivated polarization scattering matrix decomposition", *IEEE International Radar Conference*, 549-557, 1990.

16. Pottier, E., "On radar target decomposition theorems with application to target classification by using neural network method", ICAP 1991, *IEE Conference on Antennas and Propagation*, York, England, 265-268, 1991.

17. Freeman, T., S. Durden, and R. Zimmerman, "Mapping Sub-Tropical Vegetation using Multi-Frequency, Multi-Polarization SAR Data", *Proc. IGARSS'92*, May 1992.

18. Cloude, S.R., and E. Pottier, "A Review of Target Decomposition Theorems in Radar Polarimetry", *IEEE Transactions on Geoscience and Remote Sensing*, 34, 2, 498-518, Mar. 1996.

19. Czyż, Z.H., "Alternative Approaches to Polarimetric Signal and Image Processing", *Proc. 1994 Progress in Electromagnetics Research Symposium, PIERS'94*, European Space Agency, Noordwijk, The Netherlands, Jul. 11–15, 1994.

20. Krogager, E. and Z.H. Czyż, "Properties of the Sphere, Diplane, Helix Decomposition", *Proc. Journées Internationales de la Polarimétrie Radar*, JIPR'95, pp. 106-114, Nantes, France, 21-23 Mar., 1995.

21. Madsen, S.N., E.L. Christensen, N. Skou, and J. Dall, "The Danish SAR System: Design and Initial Tests", *IEEE Transactions on Geoscience and Remote Sensing*, 29, 3, 417-426, May 1991.

22. Christensen, E.L., S.N. Madsen, J. Dall, N. Skou, J.H. Jørgensen, K. Woelders, A. Netterstrøm, J. Granholm, and M. Dich, "The Danish Polarimetric SAR for Remote Sensing Applications", *Proc. IGARSS'94, 1994 International Geoscience and Remote Sensing Symposium*, pp. 1361-1364, Pasadena, CA, USA, 8-12 Aug., 1994.

23. Skou, N., J. Dall, S.N. Madsen, E.L. Christensen, K. Woelders, A. Netterstrøm, and J.H. Jørgensen, "EMISAR: An Airborne L & C-Band Polarimetric SAR for Future European Campaigns", *First International Airborne Remote Sensing Conference and Exhibition*, pp. III-711-722, Strasbourg, France, 11-15 Sep., 1994.

24. Freeman, A., "Classification of Multi-frequency, Multi-temporal Polarimetric SAR Images of Natural Vegetation", *Journées Internationales de la Polarimétrie Radar*, JIPR'92, Nantes, France, 8-11 Sep., 1992.

25. S.N. Madsen, N. Schou, J. Granholm, K. Woelders, and E. Lintz Christensen, "A System for Airborne SAR Interferometry", *European Conference on Synthetic Aperture Radar, EUSAR'96*, Königswinter, Germany, 171-176, Mar. 1996.

26. Skriver, H., J. Dall, and S.N. Madsen, "External Calibration of the Danish Polarimetric C-Band SAR", *Proc. IGARSS'94, 1994 International Geoscience and Remote Sensing Symposium*, Pasadena, CA, USA, 8-12 Aug., 1994.

27. Krogager, E., "Coherent integration of scattering matrices", *Proc. Journées Internationales de la Polarimétrie Radar*, JIPR'95, pp. 708-719, Nantes, France, 21-23 Mar., 1995.

28. Krogager, E. and S.N. Madsen, "Comparison of Various Decompositions for Analysis, Interpretation, and Classification of Polarimetric SAR Images", *European Conference on Synthetic Aperture Radar, EUSAR'96*, Königswinter, Germany, 105-108, Mar. 1996.

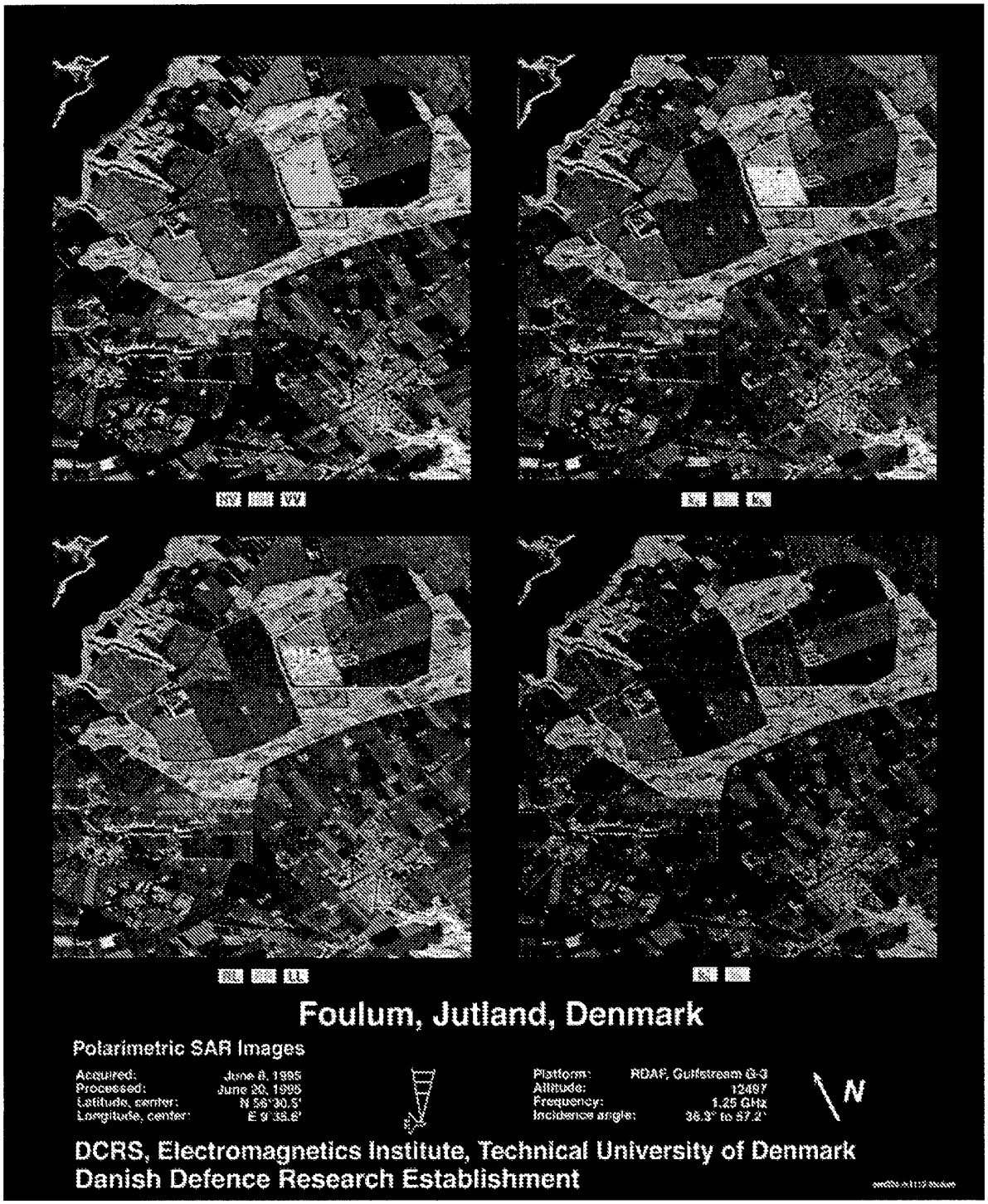


Figure 1. Grayscale print of example of color composite images according to Table 2.

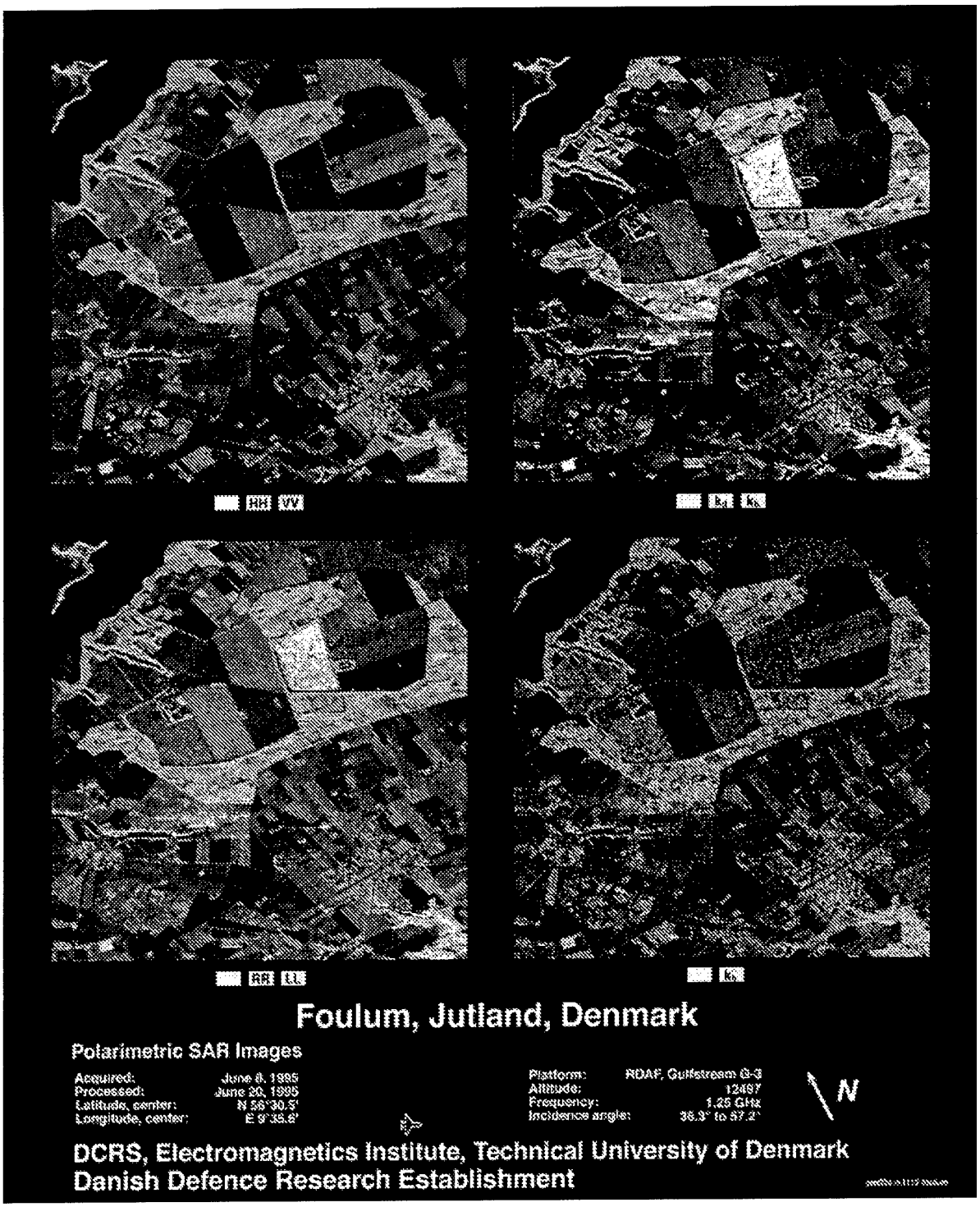


Figure 2. Red channel of Figure 1.

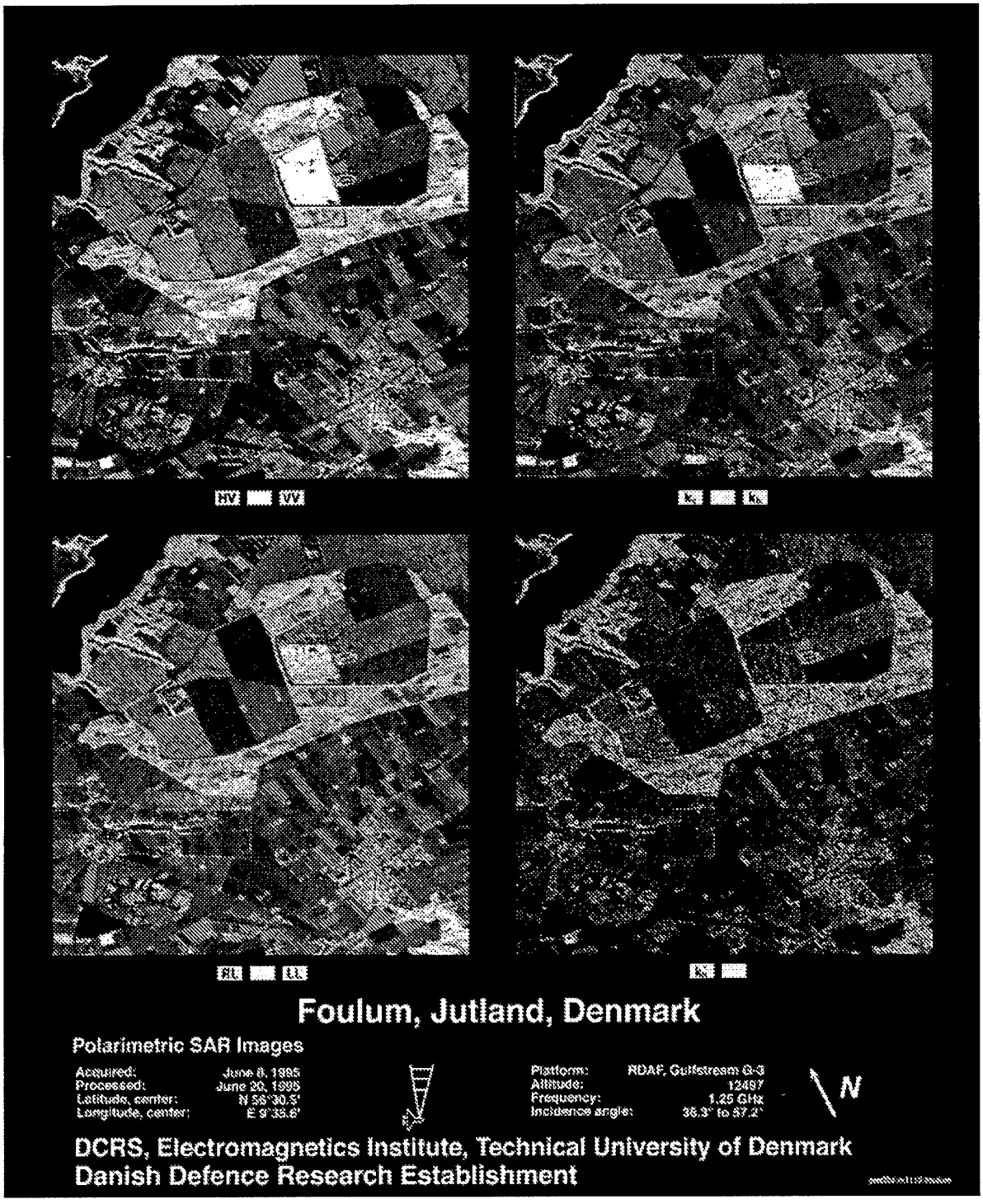


Figure 3. Green channel of Figure 1.

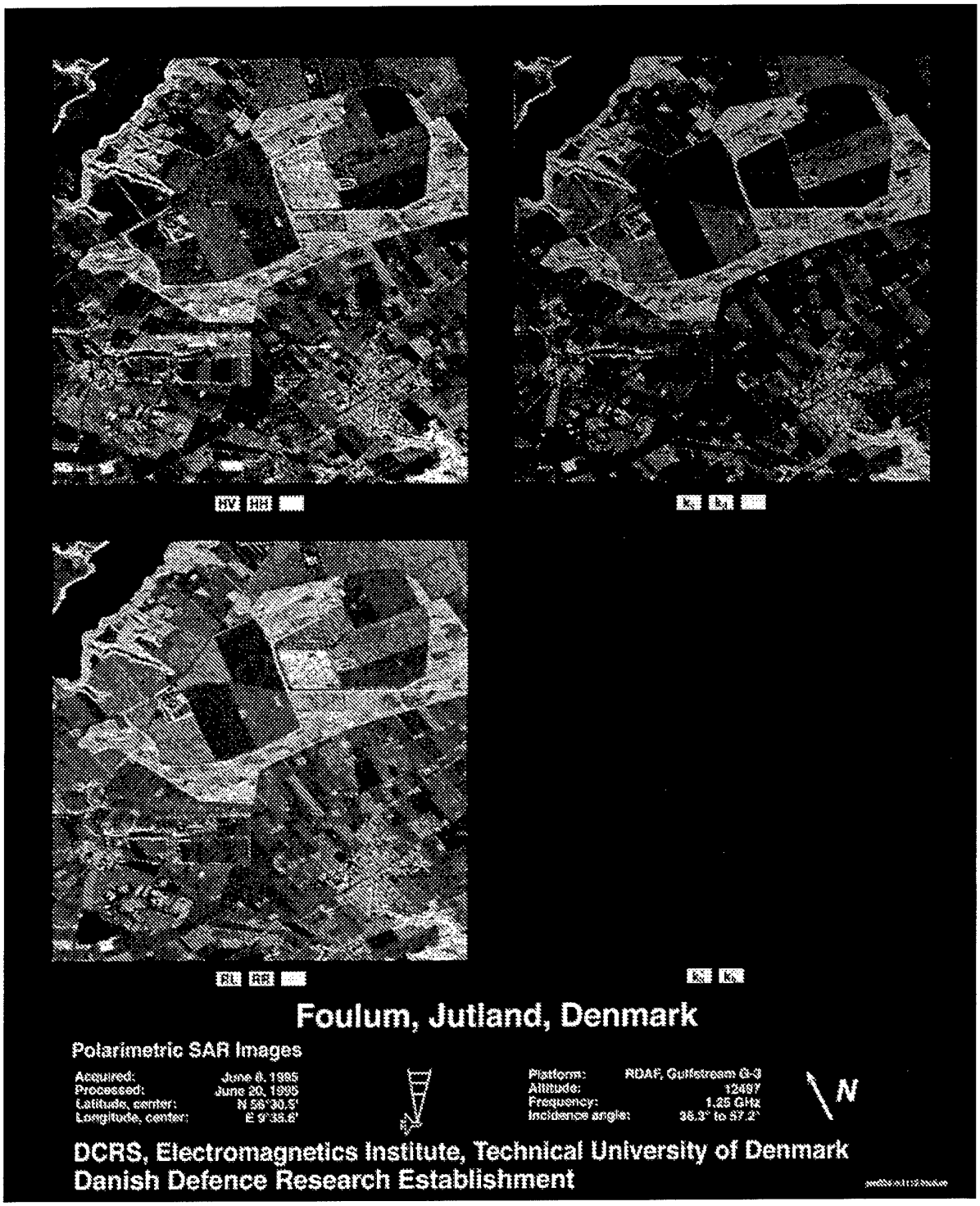


Figure 4. Blue channel of Figure 1.

Paper 17

C. Zeisse (US)

Is S_{VH} always equal to S_{HV}?

Author's reply:

For backscattering, where transmitter and receiver are co-located, the reciprocity theorem, stating that $S_{VH} = S_{HV}$, is usually assumed to apply to most practical cases. Polarimetric SAR-data are usually calibrated and processed so as to make S_{VH} equal to S_{HV}. For bistatic scattering, however, where transmitter and receiver are not co-located, S_{VH} is in general not equal to S_{HV}.

Problème de modélisation lié à la limitation temporelle de l'impulsion émise par le radar-altimètre POSEIDON

BARCKICKE Jean
Météo France/C.N.E.S., 18 avenue Edouard Belin, 31055 Toulouse Cedex, France.
Téléphone: 61 27 40 84, Télécopie: 61 28 25 95

1. SOMMAIRE

L'altimètre permet d'accéder à des paramètres caractéristiques de la surface de la mer par une comparaison du signal réfléchi avec un modèle d'écho.

Après la présentation des symboles dont le chapitre suivant fait l'objet, le fonctionnement de l'instrument est rappelé en *chapitre trois*. Le modèle d'écho classique est obtenu par une méthode originale fondée sur l'utilisation de la réponse indicielle des systèmes physiques traversés par le signal.

Le *chapitre quatre* étudie les traitements subis par le signal, en y intégrant le modèle d'écho.

La résolution classique du problème posé par la limitation temporelle de l'impulsion, qui consiste en l'ajout d'une convolution du signal reçu théoriquement avec une fonction sinus cardinal élevée au carré est exposée en *chapitre cinq*. Un examen précis de cette méthode et des propriétés du signal altimétrique permet de montrer que les hypothèses nécessaires à la résolution classique ne sont pas vérifiées. Les conséquences concrètes du traitement sont alors examinées, afin d'expliquer l'aspect bruité des échos réels.

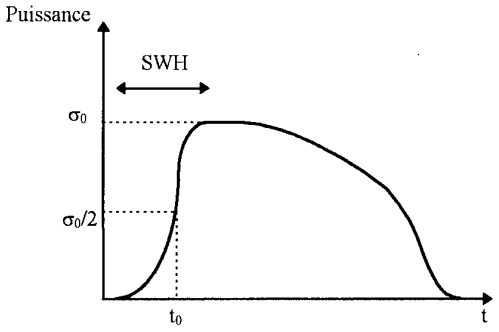


Figure 1

Le *chapitre six* tente d'apporter des éléments de réponse au problème posé. Une solution théorique, nécessitant des modifications matérielles du système récepteur est proposée. Enfin, les caractéristiques d'un simulateur numérique en cours de développement sont présentées. Destiné à modéliser l'instrument, il permettra de tester ses différentes configurations ainsi que les hypothèses des modèles d'écho.

2. LISTE DE SYMBOLES

CAG	(boucle de) Contrôle Automatique de Gain
t_0	époque = temps aller/retour entre le radar et la surface
SWH	Significant Wave Height = hauteur significative des vagues
σ_0	surface efficace de rétro-diffusion radar
ψ	incrément de phase du champ électrique à la réflexion
R(0)	coefficient de réflexion de Fresnel à zéro degrés
d	distance entre le radar et un point de la surface
t	temps (aller/retour pour parcourir cette distance)
c	vitesse de la lumière=3 10^8 m/s
H	altitude du satellite au-dessus de la surface=1336 km
z	altitude d'un point au-dessus de la surface de la mer
R	rayon de la terre
$H_{1/3}$	moyenne du tiers supérieur des vagues (≈SWH)
E[ ]	moyenne d'ensemble d'une variable aléatoire
$\overline{A}$	moyenne temporelle de A
$\otimes$	symbole de convolution
x^*	complexe conjugué de x
δ	distribution de Dirac
x(f)	signal spectral entrant
h(f)	réponse impulsionnelle spectrale du système traversé
s(f)	signal spectral sortant

3. FONCTIONNEMENT DU RADAR-ALTIMETRE

3.1 Principe général

L'altimètre est un radar à visée verticale qui envoie une impulsion proche d'un **créneau** vers la surface océane. Le temps aller-retour et la forme de l'impulsion réfléchie, appelée écho ou forme d'onde, permettent d'accéder aux paramètres recherchés: l'époque, la hauteur significative des vagues et la surface efficace de rétro diffusion radar. La Figure 1 montre l'aspect théorique du signal reçu par l'altimètre, après réflexion par la surface.

L'époque t_0 est liée au temps mis par l'impulsion pour effectuer le trajet aller-retour. Plaçons l'origine des temps sur le front de montée de l'impulsion **émise**. L'époque correspond alors à la partie du signal réfléchi provenant de la réflexion sur les points de la surface de la mer situés à hauteur nulle.

La précision requise sur la distance est telle qu'on ne peut se contenter d'une valeur approximative de t_0. Une première estimation est fournie en sortie du traqueur, qui est une boucle alpha/bêta sur les temps aller-retour des impulsions successives. De même, une estimation de la surface efficace de rétro-diffusion est obtenue en sortie de la boucle de contrôle automatique de gain (CAG).

L'estimation fine de ces paramètres, ainsi que de la hauteur significative des vagues (SWH), reliée à la pente du front de montée, à valeur crête de l'impulsion réfléchie fixée, est obtenue par comparaison de la puissance reçue avec un modèle d'écho. Cette comparaison peut être faite au sens des moindres carrés. En pratique, pour des raisons de temps de calcul, on utilise à bord un algorithme de maximum de vraisemblance [Dumont, 1]. Dans tous les cas, la précision des estimations dépend fortement de la qualité du modèle.

3.2 Modèle théorique d'écho en l'absence de troncature

L'altimètre POSEIDON est un radar à compression d'impulsion, et utilise donc des impulsions "longues". On peut cependant lui appliquer un modèle de radar impulsionnel idéal, dans le

Paper presented at the AGARD SPP Symposium on "Remote Sensing: A Valuable Source of Information", held in Toulouse, France from 22-25 April 1996, and published in CP-582.

domaine spectral, comme cela sera montré dans la quatrième partie. Le modèle de puissance reçue par un radar-altimètre impulsionnel idéal, émettant une impulsion de Dirac, et dont la réponse impulsionnelle est un Dirac (le système radar est supposé ne pas modifier pas le signal), est donné par la convolution de deux termes: la **réponse impulsionnelle de la mer sans vagues**, assimilée à un miroir plat ou à une sphère lisse, et **"la réponse impulsionnelle des vagues"**, liée à la répartition des hauteurs des points spéculaires.

Faisons pour l'instant l'hypothèse que seule la réflexion à l'interface modifie la puissance du signal, et que tous les points situés à la même distance d du radar possèdent les mêmes propriétés réflectives, caractérisées par le coefficient de Fresnel sous incidence normale:

$$R(0) = \sqrt{\sigma_0}\, e^{j\psi}$$

R(0) peut contenir une phase, mais σ_0, facteur de réflexion en puissance, est réel [Ulaby et al., 2].

Le radar émet l'impulsion à t=0. La puissance reçue à t=2d/c est proportionnelle à l'élément de surface infinitésimal correspondant à l'ensemble des points spéculaires situés à la distance d, car selon l'hypothèse choisie pour la réflexion, les ondes réfléchies par ces différents points sont en phase.

La réponse impulsionnelle puissance est donc proportionnelle à la surface des points spéculaires considérés. Mathématiquement, il est plus rigoureux de calculer la réponse indicielle puissance, qui est proportionnelle à la portion de surface comprise à l'intérieur de la sphère de rayon d = ct/2, chaque point étant affecté d'un coefficient correspondant à la contribution en puissance qu'il fournit à l'écho.

En supposant qu'il est possible de séparer les effets respectifs de la mer plate et des vagues, l'incidence peut être prise égale à zéro en première approximation. Tous les rayons réfléchis sont alors captés par l'antenne, et la réponse indicielle de la mer plate peut s'écrire:

$$\mathrm{Ind(t)=Y(t-t_0)} \int_0^{\sqrt{\left(\frac{ct}{2}\right)^2 - H^2}} \sigma_0 \frac{G^2(r)}{H^4}\, \mathrm{r\ dr} \qquad (1)$$

où Y(t) est la fonction échelon unitaire, t_0 = 2H/c, et G(r) est le gain d'antenne correspondant à un angle d'émission/réception pour un point de la surface situé à la distance r du nadir. En dérivant Ind(t), on obtient la réponse impulsionnelle:

$$\mathrm{Imp(t)=}\ Y(t-t_0)\ \frac{\sigma_0 c^2 t}{4H^4} G^2\left(\sqrt{\left(\frac{ct}{2}\right)^2 - H^2}\right) \qquad (2)$$

On utilise une simplification du gain d'antenne donné par [Brown, 3]:

$$G(r)= G_0\ e^{-\frac{2}{\gamma}\left(\frac{r}{H}\right)^2}$$

On obtient ainsi la réponse de la mer plate en u = t-t_0 , à l'ordre 2 en r/(H-z) et à l'ordre 0 en z/H, ce qui impose de considérer u<< t_0, c'est à dire de rester à l'ordre 0 en u/t_0:

$$\mathrm{Imp(u)=}\ \frac{\sigma_0 G_0^2 c}{2H^3} Y(u)\ e^{-\frac{4c}{\gamma H}u} \qquad (3)$$

Repartons de l'intégrande de (1) sous la forme $\sigma_0(r)G2(r)/d^4$. La réponse indicielle devient:

$$\mathrm{Ind(t)=Y(t-t_0)} \int_0^{r_{lim}(t)} \sigma_0 \frac{G^2(r)}{H^4}\left[1 - 2\left(1+\frac{H}{R}\right)\left(\frac{r}{R}\right)^2\right] \mathrm{r\ dr}$$

où $r_{lim}(t) = \sqrt{\dfrac{\left(\frac{ct}{2}\right)^2 - H^2}{1+\frac{H}{R}}}$ et $\left[1 - 2\left(1+\frac{H}{R}\right)\left(\frac{r}{R}\right)^2\right]$ est en ordre 3 sur la correction de H loin du nadir.

En effectuant un développement limité de G(r), comme dans le cas de la mer plate, la réponse de la surface sphérique s'écrit:

$$\mathrm{Imp(u)=}\ \frac{\sigma_0 G_0^2 c}{2H^3\left(1+\frac{H}{R}\right)} Y(u)\ e^{-\frac{4c}{\gamma H\left(1+\frac{H}{R}\right)}u} \qquad (4)$$

On utilise la formule (3), car la correction de courbure peut être intégrée en fin de calcul, en remplaçant H par H(1+H/R) dans les convolutions et en divisant le résultat final g(t) par (1+H/R).

La réponse de la mer plate est choisie sous la forme:

$$\mathrm{P_{FS}(t)=\ Y(t-t_0)}\, e^{-\frac{4c}{\gamma H}(t-t_0)}$$

et celle des vagues sous la forme $v(t)=\frac{c}{2}P\left(-\frac{ct}{2}\right)$,

où $P(x)=\frac{1}{\sigma_v\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma_v^2}}$ avec $\sigma_v=\frac{H_{1/3}}{4}$

En supposant σ_0 = cste., il suffit de calculer de manière exacte:

$$\int_{-\infty}^{+\infty} Y(u-t_0)\, e^{-\frac{4c}{\gamma H}(u-t_0)}\, \frac{c}{2} P\left[-\frac{c}{2}(t-u)\right] du$$

Nous obtenons, à une constante multiplicative près, le modèle de puissance reçue sous la forme $\sigma_0 g(t)$, où σ_0 est la surface efficace de rétro-diffusion moyenne sur la tache, et

$$\mathrm{g(t)=}\ \frac{G_0^2 c}{4H^3}\left[1 + erf\left(\frac{t - t_0 - \frac{4c}{\gamma H}\sigma_c^2}{\sigma_c\sqrt{2}}\right)\right] e^{-\frac{4c}{\gamma H}\left(t - t_0 - \frac{2c}{\gamma H}\sigma_c^2\right)} \qquad (5)$$

avec $\sigma_c=\frac{2\sigma_v}{c}$.

4. TRAITEMENT DU SIGNAL DU RADAR-ALTIMETRE

Le radar POSEIDON émet le champ électrique $E(t)=E_0 \cos \varphi(t)$, de fréquence $f(t)=f_0-Kt$ avec f_0=13.6 GHz et K=320 MHz/100µs. L'impulsion réelle est tronquée à l'intervalle [-T/2, T/2], où T=100µs.

On considère un point de mer plate à la distance d du radar, sous incidence normale. La phase du champ électrique incident à l'interface est égale à $\varphi(t-d/c)$, et le champ électrique réfléchi par chaque point de l'intersection entre la sphère de rayon d et la

surface moyenne (mer plate) est égal au champ incident complexe multiplié par [R(0)](d).
Ces signaux réfléchis (infinitésimaux) arrivent en phase au niveau de l'antenne, qui reçoit un champ électrique retardé de nouveau de d/c par le trajet surface/radar, avec une phase augmentée de ψ(d), et de module proportionnel à $\sqrt{\sigma_0(d)}$.

De plus, la puissance de ce champ, proportionnelle à l'élément de surface élémentaire associé à l'ensemble des points situés à la distance d, est égale à la réponse impulsionnelle puissance d'un radar impulsionnel théorique, prise à l'instant t = 2d/c, soit $\sigma_0(d)\ g(2d/c)$.
L'ensemble des "points spéculaires" situés à la distance d du radar produit alors un champ électrique réfléchi élémentaire au niveau de l'antenne, égal à:

$$\sqrt{\frac{2\sigma_0(d)}{c}g\left(\frac{2d}{c}\right)}E_0\cos\left[\varphi\left(t-\frac{2d}{c}\right)+\psi(d)\right] \qquad (6)$$

Dans l'hypothèse impulsion infinie dans le temps, on considère la voie I (en phase). Le signal reçu est mélangé avec la réplique $E_0 \cos\varphi(t-t_0)$. On s'intéresse aux deux phases:

$\varphi_1=\varphi(t-2d/c)$ pour le signal $\varphi_2=\varphi(t-t_0)$ pour la réplique

On obtient en sortie un signal de phase $\varphi_1+\varphi_2$ et un signal de phase $\varphi_1-\varphi_2$. Le signal en $\cos(\varphi_1+\varphi_2)$ est éliminé par le mélangeur et/ou son filtre de prédétection. Il subsiste le signal défini par:

$$\frac{\varphi_1-\varphi_2}{2\pi}=[2K\frac{d-H}{c}](t-t_0)-\left(\frac{f_0}{K}+\frac{d-H}{c}\right)[2K\frac{d-H}{c}] \qquad (7)$$

La puissance du champ électrique (6) due aux points spéculaires correspondant à un élément de distance radiale Δd est égale à $\frac{2\sigma_0(d)}{c}E_0^2 g(\frac{2d}{c})$ Δd, et la relation entre la fréquence du signal et cette distance, comme le montre (7), est $f = 2K\frac{d-H}{c}$.
La puissance élémentaire peut donc s'écrire $\frac{2\sigma_0(d)}{c}E_0^2 g(\frac{2d}{c})\frac{c}{2K}\Delta f$.

L'ensemble des points spéculaires situés à la distance d est à l'origine d'une composante du signal en sortie de mélange. Cette composante est caractérisée par:

- une fréquence: $f(d) = 2K\frac{d-H}{c}$
- une phase: $-2\pi\left(t_0+\frac{f_0}{K}+\frac{f}{2K}\right)f+\psi(d)$, (8)
- une amplitude: $E_0\sqrt{\frac{\sigma_0(d)}{K}g\left(\frac{2H}{c}+\frac{f}{K}\right)}$

Le signal infini, somme des composantes décrites ci-dessus est défini par:

$$s(t)=\int_{-\infty}^{+\infty}E_0\sqrt{\frac{\sigma_0(d)}{K}g(t_0+\frac{f}{K})}\ e^{-j2\pi\left(t_0+\frac{f_0}{K}+\frac{f}{2K}\right)f+j\psi(f)}e^{j2\pi ft}df \qquad (9)$$

Le signal réel est tronqué à un intervalle temporel de taille T centré sur l'époque. On suppose que les déphasages à la réflexion sont peu différents de π, d'ou $\psi(f)=\pi+\varphi_0$, avec $\varphi_0(f)<<\pi$. Le signal a donc pour transformée de Fourier:

$$A(f)=(E_0\ e^{\varphi_0(0)-j2\pi t_0 f}) \qquad (10)$$
$$[\sqrt{\frac{\sigma_0(f)}{K}g(t_0+\frac{f}{K})}e^{-j2\pi\left(\frac{f_0}{K}+\frac{f}{2K}\right)f+j\psi(f)}]\otimes[\frac{\sin(\pi Tf)}{\pi f}]$$

On utilise alors le module carré de A(f), disponible en sommant les carrés des voies I (en phase) et Q (en quadrature):

$$I^2+Q^2=E_0^2|s(f)|^2,$$

où $s(f)=x(f)\otimes h(f)$, avec $x(f)=\sqrt{\frac{\sigma_0}{K}g(t_0+\frac{f}{K})}e^{-j2\pi\left(\frac{f_0}{K}+\frac{f}{2K}\right)f}$
- ψ(f) est négligé - et $h(f)=\frac{\sin(\pi Tf)}{\pi f}$.

En pratique, la réflexion est partiellement cohérente, et x(f) doit être affecté d'un coefficient compris entre $\frac{\rho(f)}{\sqrt{N_s(f)}}$ (cohérence inexistante) et 1 (cohérence totale), ou ρ(f) suit une loi de Rayleigh de paramètre égal à 1, et N_s est le nombre de points spéculaires situés à la distance $d = H+\frac{cf}{2K}$ [Dorrer, 4].

5. IMPACT DE LA TRONCATURE TEMPORELLE

Le problème à résoudre est le calcul de la convolution s(f), afin d'aboutir à un modèle d'écho prenant en compte la réponse impulsionnelle h(f) du radar.

5.1 Méthode de résolution classique

x(f) étant un signal aléatoire, on utilise la relation [Dorrer, 4]:

$$s(f)=x(f)\otimes h(f)\Rightarrow E\left[|s(f)|^2\right]=E\left[|x(f)|^2\right]\otimes|h(f)|^2 \qquad (11)$$

En prenant les modules carrés du signal spectral, et en moyennant sur N impulsions élémentaires, la forme d'onde finale est égale à:

$$\overline{I^2+Q^2}=E_0^2\left\{g\left(t_0+\frac{f}{K}\right)E\left[\frac{\sigma_0(f)}{K}\frac{\rho^2(f)}{N_s}\right]\right\}\otimes\frac{\sin^2(\pi Tf)}{(\pi f)^2}$$

Notons maintenant $\Sigma_0(f)$ la moyenne d'ensemble des $\sigma_0(f)$, pondérées par les $\rho^2(f)/N_s(f)$, qui représentent la "quantité de cohérence" dans la somme des réflexions élémentaires contribuant à la puissance reçue à la fréquence f. Si nous supposons que le profil de surface efficace de rétro-diffusion moyen $\Sigma_0(f)$ est constant, après mise à niveau par la boucle de CAG, nous obtenons la formule classique du modèle d'écho:

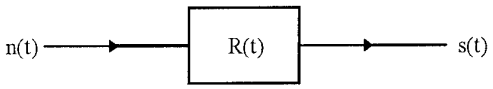
$$\frac{\Sigma_0}{K}g\left(t_0+\frac{f}{K}\right)\otimes\frac{\sin^2(\pi Tf)}{(\pi f)^2}$$

La relation (11) permet de s'affranchir des phases présentes dans les signaux x(f) et/ou h(f).

5.2 Application de la méthode au signal traité

On peut aisément montrer, à l'aide d'un contre-exemple, que (11) n'est mathématiquement pas vérifiée. Il est cependant intéressant de chercher ses limites d'application dans le cas de l'altimètre, par exemple en quantifiant l'impact sur les résultats de l'utilisation de (11) au lieu d'une relation mathématiquement exacte. Une façon de mesurer cet impact sera exposée dans la partie 6. Nous nous intéressons ici aux problème théorique posé par la relation (11).

On considère un signal aléatoire n(t) traversant un système linéaire stationnaire de réponse impulsionnelle R(t).



$$s(t)= n(t)\otimes R(t) \quad \Leftrightarrow \quad s^*(t)= n^*(t)\otimes R^*(t)$$

Posons $U(t)=E[|s(t)|^2]= E[s(t)s^*(t)]$, c'est à dire:

$$U(\tau)=E\left[\int_{t=-\infty}^{+\infty}\int_{t'=-\infty}^{+\infty} n(\tau-t)n^*(\tau-t')R(t)R^*(t')dt'dt\right]$$

En pratique, on réalise une moyenne sur N impulsions, donc

$$U(\tau)= \frac{1}{N}\sum_{i=1}^{N}\int_{t=-\infty}^{+\infty}\int_{t'=-\infty}^{+\infty} n_i(\tau-t)n_i^*(\tau-t')R(t)R^*(t')dt'dt$$

$$= \int_{t=-\infty}^{+\infty}\int_{t'=-\infty}^{+\infty}\frac{1}{N}\sum_{i=1}^{N} n_i(\tau-t)n_i^*(\tau-t')R(t)R^*(t')dt'dt$$

Dans notre cas, il est rigoureux d'écrire:

$$U(\tau)= \int_{t=-\infty}^{+\infty}\int_{t'=-\infty}^{+\infty} E\left[n(\tau-t)n^*(\tau-t')\right]R(t)R^*(t')dt'dt \qquad (12)$$

5.2.1 Démonstration classique

Si on a la propriété

$$E[n(t_1)n^*(t_2)]=\delta(t_1-t_2)\,E[|n(t_1)|^2] \qquad (13)$$

alors:

$$U(\tau)= \int_{-\infty}^{+\infty}\left\{\int_{-\infty}^{+\infty}\delta(t-t')E\left[|n(\tau-t)|^2\right]R^*(t')dt'\right\}R(t)dt \quad \Leftrightarrow$$

$$U(\tau)= \int_{-\infty}^{+\infty}E\left[|n(\tau-t)|^2\right]R^*(t)R(t)dt = E[|n(\tau)|^2]\otimes|R(\tau)|^2 \qquad (14)$$

5.2.2 Etude des possibilités d'application de (12)

Si on suppose que $f(t,u)= E[n(t)n^*(t+u)]$ a un aspect proche de la fonction de la figure 2, et que son support l(t) est de longueur

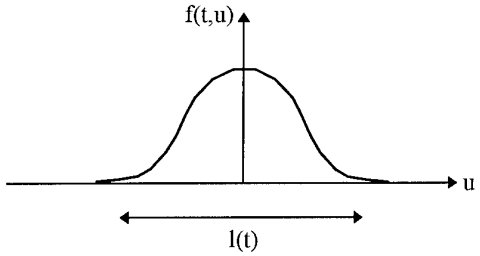


Figure 2

faible devant les échelles de variation de R(t), U(τ) devient:

$$\int_{t=-\infty}^{+\infty}\left[\int_{-\infty}^{+\infty} f(\tau-t,t-t')R^*(t')dt'\right]R(t)dt$$

$$= \int_{t=-\infty}^{+\infty} R^*(t)\left[\int_{-\infty}^{+\infty} f(\tau-t,t-t')dt'\right]R(t)dt$$

La convolution de R(u) et f(t,u) est remplacée par $R(u)\int_{-\infty}^{+\infty} f(t,u)du$, comme si f(t,u) était une impulsion de Dirac.

Nous avons en fait substitué $\delta(u)\int_{-\infty}^{+\infty} f(t,u)du$ à $E[n(t)n^*(t+u)]$, c'est à dire $\delta(t_2-t_1)\int_{-\infty}^{+\infty} E[n(t_1)n^*(t_2)]dt_2$ à $E[n(t_1)n^*(t_2)]$, membre de gauche de (13). Le résultat est l'égalité (14), dans laquelle $E[|n(t)|^2]$ doit être remplacé formellement par:

$$\int_{-\infty}^{+\infty} E[n(t)n^*(t')]dt' \approx E[|n(t)|^2]\,l(t)$$

Il est donc justifié d'appliquer (13) lorsque l'on a en même temps:

- $E[n(t)n^*(t+\tau)]$ à support de taille réduite autour de $\tau=0$ (15)
- indépendance de la longueur l(t) de ce support avec t (16)

Par exemple, pour des signaux physiques n(t) centrés, $\int_{-\infty}^{+\infty} E[n(t)n^*(t')]dt' = \int_{-\infty}^{+\infty} cov[n(t),n(t')]dt'$. Avec une longueur de corrélation courte du type de celle présentée en Figure 2, et ne dépendant pas du point de calcul, le résultat peut être appliqué.

Dans le cas des signaux de l'altimètre, examinons la condition (15). Le terme à étudier, selon les notations de (13), est égal à:

$$\sqrt{g(t_0+\frac{f_1}{K})g(t_0+\frac{f_2}{K})}\,e^{j2\pi\left[\left(\frac{f_0}{K}+\frac{f_2}{2K}\right)f_2-\left(\frac{f_0}{K}+\frac{f_1}{2K}\right)f_1\right]}$$

$$E\left[\sqrt{\frac{\sigma_0(f_1)\rho^2(f_1)}{N_s(f_1)}\frac{\sigma_0(f_2)\rho^2(f_2)}{N_s(f_2)}}\right]/K$$

La conclusion dépendra des propriétés de $E[\sqrt{\frac{\sigma_0(f_1)\rho^2(f_1)}{N_s(f_1)}\frac{\sigma_0(f_2)\rho^2(f_2)}{N_s(f_2)}}]$.

Les surfaces efficaces de rétro-diffusion $\sigma_0(f_i)$ évoluent en général de manière covariante, quel que soit $d_i = H+\frac{cf_i}{2K}$, en fonction des paramètres géophysiques, supposés varier à des échelles grandes devant la tache au sol. Ceci est d'autant plus vérifié que ces termes résultent de sommes de signaux réfléchis sur les cercles de rayon d_i, et correspondent à des moyennes spatiales. Il s'agit d'ailleurs d'une hypothèse de base pour les systèmes altimétriques, puisque σ_0 est considéré constant dans les modèles d'écho. Quelle que soit l'incertitude sur ces hypothèses, il s'agit de termes positifs, subissant des variations conjointes.

On obtient donc en sortie de l'analyseur de spectre, à la fréquence F:

$$\frac{K}{E_0^2}\overline{I^2+Q^2}(F)=$$

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left\{\sqrt{g(t_0+\frac{F-f}{K})g(t_0+\frac{F-f'}{K})}\right.$$

$$e^{j2\pi\left[\left(\frac{f_0}{K}+\frac{F-f'}{2K}\right)(F-f')-\left(\frac{f_0}{K}+\frac{F-f}{2K}\right)(F-f)\right]}$$

$$E\left[\frac{\sqrt{\sigma_0(F-f)\rho^2(F-f)}}{N_s(F-f)}\frac{\sqrt{\sigma_0(F-f')\rho^2(F-f')}}{N_s(F-f')}\right]$$

$$\left.\frac{sin(\pi Tf)}{\pi f}\frac{sin(\pi Tf')}{\pi f'}\right\}df'df$$

$E\left[\frac{\sqrt{\sigma_0(F-f)\rho^2(F-f)}}{N_s(F-f)}\frac{\sqrt{\sigma_0(F-f')\rho^2(F-f')}}{N_s(F-f')}\right]$ ne peut être assimilé à une fonction $\delta(f-f')$, pour plusieurs raisons, par exemple la positivité des σ_0. Il en résulte que le support de cette fonction n'est pas de taille inférieure à ceux des autres termes de l'intégrande, et aucune approximation similaire à celle exposée précédemment n'est applicable.

5.3 Effet de la phase dans la convolution des signaux

Nous tentons dans cette partie d'expliquer de manière "physique" l'effet des traitements subis par le signal, afin de pouvoir définir des axes de recherche d'améliorations.

Le spectre qui nous intéresse résulte de la convolution du signal x(f) complexe avec la fonction $h(f)=\frac{sin(\pi Tf)}{(\pi f)}$, x(f) étant physiquement bruité par un speckle dû au terme $\sqrt{\frac{\sigma_0(f)}{N_s(f)}}\rho(f)$.

Une variation de fréquence est liée à une variation de distance radiale selon la première équation de (8). Un développement limité, effectué au voisinage du nadir, permet de lier la variation de distance radiale à une variation de distance horizontale Δr, ce qui permet alors d'obtenir une relation entre variation de fréquence et variation de distance horizontale: $\Delta f_{nad}=\frac{(\Delta r)^2 K}{2Hc}$.

En s'éloignant du nadir, la variation de fréquence devient inférieure à celle obtenue par cette relation.

Plaçons-nous au nadir, pour observer les variations de puissance les plus lentes le long de la forme d'onde. Supposons que les fluctuations de cohérence à la réflexion deviennent sensibles sur des longueurs d'environ 50 m. L'échelle de variation correspondante sur la fréquence, en utilisant cette relation, est d'environ 10 Hz. Loin du nadir, cette échelle de variation se fait sur une plage de fréquence encore plus courte. Le signal spectral avant troncature est donc fortement bruité (on utilise en finale un point tous les 10 kHz) par un speckle pouvant produire des variations dont la "fréquence" peut dépasser 0.1s, soit 1000 fois la "fréquence" d'échantillonnage de 0.1 s (les domaines temporel et fréquentiel étant ici inversés, le terme "fréquence" a pour unité la seconde, le signal étudié étant en fait spectral).

Ce signal est multiplié par une porteuse de "fréquence" $-f_0/K$. Il est donc de "fréquence" beaucoup plus rapide que l'échantillonnage utilisé, (f_0/K=4.25 ms, pour une "fréquence" d'échantillonnage T= 0.1 ms). La convolution avec h(f), de portée 1/T=10 kHz, effectue une moyenne de x(f), alors que celui ci décrit plus de 30 ondulations par lobe de h(f), selon une "période" K/f_0=235 Hz. Le résultat est alors très faible devant les valeurs crêtes de x(f), car les aires positives et négatives s'éliminent fortement.

De plus, suivant que les composantes spectrales de la partie non complexe de x(f), c'est à dire $\sqrt{g(t_0+\frac{f}{K})}$ entaché de speckle, entrent en phase ou non avec la partie complexe, la dynamique du résultat peut changer fortement numériquement, d'où des variations de puissance le long de la forme d'onde, non liées uniquement aux variations de surface efficace de rétro diffusion.

On peut aussi aborder le problème en s'intéressant aux composantes fréquentielles de x(f), qui sont en fait dans l'espace physique. x(f) est centré sur la porteuse $-f_0/K$=-4250 µs. A mesure que l'on s'éloigne de cette fréquence, en s'approchant de la bande [-50 µs, 50 µs], on récupère de plus en plus de composantes du speckle, au détriment de celles dues au signal utile $\sqrt{g(t_0+\frac{f}{K})}$. Lors de la multiplication du spectre par la fenêtre [-50 µs, 50 µs], le résultat se trouve alors plus entaché de bruit que si le retard apparent f_0/K présent dans le signal physique n'existait pas.

6. SOLUTIONS POTENTIELLES ET INVESTIGATIONS

6.1 Diminution d'une partie du "bruit instrumental"

Le signal reçu par le radar, en provenance de la surface est défini par la phase $\varphi(t)=2\pi f_0 t-\pi Kt^2$. Examinons l'effet d'un récepteur à deux étages de démodulation.

- Supposons qu'avant mélange avec la réplique, on multiplie d'abord le signal par une porteuse f_1 proche de f_0.

La phase de l'entrée devient $\varphi_e(t)=2\pi(f_0-f_1)t-\pi Kt^2$, l'autre terme étant éliminé par le filtre de prédétection.

- De même, on utilisera une réplique centrée sur f_0-f_1, définie par sa phase $\varphi_r(t)=2\pi(f_0-f_1-\frac{K}{2}t)t$. Prise en t-2H/c, c'est à dire retardée du temps aller-retour par rapport à l'émission, nous obtenons:$\varphi_2(t)=2\pi[f_0-f_1-\frac{K}{2}$ (t-2H/c)](t-2H/c), que l'on mélange à $\varphi_1(t)=2\pi[f_0-f_1-\frac{K}{2}$ (t-2d/c)](t-2d/c).

Nous nous trouvons dans le contexte de l'étude entreprise en partie 3, avec formellement f_0-f_1 au lieu de f_0.
En particulier, revenons sur la partie 5.3. La période des oscillations du signal d'entrée devient égale à $\frac{K}{f_0-f_1}$.

Dans le cas idéal où $\frac{f}{2K} << \frac{f_0-f_1}{K} <<$T, les ondulations dues à la porteuse ne sont pas visibles dans la convolution (10), car le sinus cardinal atteint zéro avant que la phase de x(f) ne varie. Pour que ces hypothèses soient pleinement vérifiées, la différence entre f_0 et f_1 doit être d'environ 7 à 8 MHz, ce qui revient à utiliser une porteuse intermédiaire idéale f_1 =13.592 GHz. En pratique, il est possible d'utiliser une différence entre les deux porteuses atteignant 100 MHz. On est alors moins strict sur la condition:

$$\frac{K}{f_0-f_1} >> 1/T \qquad (17)$$

ces deux termes devenant respectivement égaux à 32000 Hz et 10000 Hz.

Remarque: ce type de traitement est effectué sur TOPEX, car le signal réfléchi passe à travers deux étages de démodulation. Un premier étage ramène le signal autour de 500 MHz, puis le mélange avec la réplique effectue la démodulation finale, produisant un signal de largeur de bande 640 kHz autour de zéro [Zieger et al., 5]. Cependant, avec les valeurs qui ont été choisies, la condition (17) n'est pas vérifiée. En effet, le premier terme vaut environ 6400 Hz, et le second 10000 Hz. La phase de x(f) peut varier notablement avant que le sinus cardinal soit proche de zéro. Ceci explique certainement que, malgré le choix de cette technologie, les résultats de TOPEX ne soient guère différents de ceux de POSEIDON.

La modification proposée ne prétend pas résoudre l'ensemble des problèmes soulevés ici, mais seulement réduire de façon importante le "speckle" instrumental dû à la présence d'une phase à variation très rapide. Le speckle physique subsistera et nécessitera, pour être mieux pris en compte, des améliorations du traitement et surtout du modèle d'écho.

6.2 Simulation d'écho

Les principes de base du simulateur de forme d'onde qui va être présenté doivent nous permettre d'atteindre deux buts:

- Comparer les deux membres de (11), et tenter de quantifier l'erreur commise en utilisant cette égalité dans le cas de l'altimètre,
- Etudier la fonction de transfert entre les estimations des paramètres géophysiques et leurs valeurs réelles

L'idée générale consiste à simuler numériquement des échos, qui seront ensuite envoyés en entrée de l'estimateur, chargé de retrouver les paramètres géophysiques initiaux.
Le signal spectral entrant x(f) est la somme de N_s signaux élémentaires provenant des points spéculaires situés à la distance d telle que $f = 2K\frac{d-H}{c}$. Considéré comme le nombre d'éléments de surface élémentaires à la distance d, N_s est proportionnel à l'élément de surface infinitésimal situé à cette distance, ce qui permet de le remplacer par $H+\frac{cf}{2K}$. Le modèle du signal entrant est alors:

$$x(f)=\sqrt{\frac{g(t_0+\frac{f}{K})}{H+\frac{cf}{2K}}}e^{-j2\pi(\frac{f_0}{K}+\frac{f}{2K})f}\sqrt{\sigma_0(f)}\rho(f)e^{j\Psi(f)}$$

Les différentes composantes à prendre en compte sont:

- $g\left(t_0+\frac{f}{K}\right)$, donné par le modèle d'écho (5)
- t_0 = 2H/c avec H = 1336 km et c = 3 10^8 m/s
- K = 320 Mhz/100µs et f_0 = 13,6 Ghz
- $\sigma_0(f)$, le profil de surface efficace de rétro diffusion le long de la forme d'onde, que l'on peut écrire $\sigma_0(d)$ en utilisant la relation $f=2K\frac{d-H}{c}$, ou encore $\sigma_0(r)$, r étant la distance au nadir. L'approximation de tir vertical avec séparation des effets mer (plate ou sphérique) et vagues est cohérente avec l'hypothèse $\sigma_0(r)$= cste =σ_0. Pour affiner le modèle, on prendra σ_0 variable (aléatoire gaussienne). On calculera alors pour chaque réalisation du signal aléatoire x(f) une valeur issue d'un algorithme de loi normale centrée réduite.
- $\rho(f)$ variable aléatoire suivant une loi de Rayleigh de paramètre égal à un
- $\psi(f)$ aléatoire sur [0,2π], uniforme ou gaussienne.

On génère N_f (=86 pour POSEIDON) impulsions numériques x(f), avec un pas d'échantillonnage fréquentiel suffisamment fin pour que les fréquences présentes dans les différents termes soient conservées (en considérant la cohérence à la réflexion constante sur une distance de 50 m, on obtient un pas de 1Hz), et on compare les deux branches présentées en Figure 3:

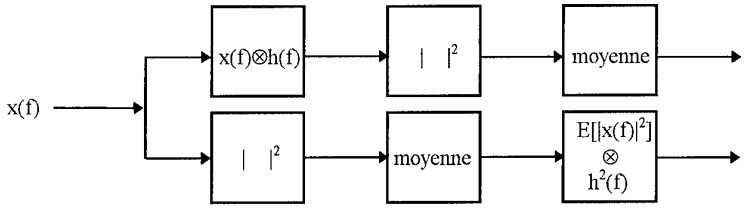


Figure 3

La première branche simule le traitement réel, et la seconde simule le résultat de l'application du second membre de (11). La configuration du code est aisément modifiable.

- Il est possible d'intégrer une autre branche en aval du calcul de x(f), afin de tester une relation différente de (11), qui permette d'aboutir à un modèle d'écho utilisable.
- Le simulateur est largement paramétré. On pourra ainsi tester (principalement) l'impact
  - du modèle d'écho g(t), en utilisant des modèles plus sophistiqués (prise en compte de la dissymétrie des vagues),
  - de la répartition des σ_0 le long de la forme d'onde,
  - de la variation de la cohérence à fréquence donnée, en jouant sur la loi suivie par $\rho(f)$ et $\psi(f)$,
  - de la valeur de la porteuse f_0, et par conséquent d'un étage supplémentaire de démodulation.

D'autres tests sont envisageables, moyennant quelques développements peu coûteux en temps.

7. CONCLUSION ET PERSPECTIVES

Cette étude à permis de soulever de façon théorique le problème inhérent à la théorie classique de l'estimation par un modèle d'écho, pour un radar à compression d'impulsion. Elle sera prochainement validée par une simulation numérique, afin de tester la part de bruit contenue dans les paramètres estimés, due à l'utilisation d'un modèle fondé sur une relation de convolution inexacte. Ce simulateur permettra également de tester les hypothèses physiques utilisées dans le modèle d'écho.

Une amélioration du système, proche du principe de fonctionnement de l'altimètre TOPEX est envisageable. Elle devrait permettre de réduire le bruit instrumental contenu dans les formes d'onde. Le simulateur destiné à valider la partie théorique de l'étude sera utilisé afin d'estimer l'impact de la modification matérielle sur les résultats physiques.
Enfin, le modèle, qui n'est théoriquement adapté au signal d'entrée que lorsque ce dernier est proche d'un bruit blanc devra être revu, afin de prendre en compte au mieux cette amélioration matérielle.

8. REFERENCES

Thèse de doctorat

1. Dumont , J. P. , " Estimation des paramètres altimétriques des signaux radar POSEIDON", Thèse de doctorat, Toulouse, septembre 1985

Livre

2. Ulaby, F.T. et al. , " Microwave Remote Sensing " , Addison-Wesley Publ. Company

Article périodique

3. Brown, G.S., " The Average Impulse Response of a rough Surface and its Applications, IEEE Transactions on Antennas and Propagation , Vol. AP-25, No. 1, pp 67-74, January 1977

4. *Notre interne CNES*
   Dorrer, M., "Traitement du signal d'un altimètre", Note interne CNES/CT/OT/CH/RE-18, février 1983

Compte-rendus de conférence

5 Zieger, A.R. et al. , " Nasa Radar Altimeter for the TOPEX / POSEIDON Project " , Proceedings of the IEEE, Vol. 79, No. 6, june 1991

Characterization of Natural Targets Using Multiparametric SIR-C/X-SAR Imagery

L. Castellano, P. Murino, A. Siciliano
Dipartimento di Scienza ed Ingegneria dello Spazio "L.G.Napolitano"
University of Naples
P.le Tecchio, 80 - 80125 - Napoli - Italy

1. ABSTRACT

The paper deals with the characterization of natural targets in terms of backscattering properties using multifrequency, multipolarization SAR data acquired over the Campanian area (Southern Italy) during the SIR-C/X-SAR shuttle missions of April 1994.
The knowledge of the complete scattering matrix, usually measured in a linear basis (HV), allows to compute a covariance matrix once that a polarimetric characteristic vector as been defined. The scattering properties of natural targets, imaged by airborne and spaceborne SAR polarimeters, can be statistically described by such a matrix whose decomposition provides useful information about the nature of scattering mechanisms acting inside the resolution cell.
If the polarimetric characteristic vector is defined using a set of simple matrices having a physical significance and if no correlation exists between the co-polarized and the cross-polarized channels, the covariance matrix assumes a very simple form and its decomposition provides a simple tool to get insight into the target scattering behavior.

2. INTRODUCTION

At the time of writing a number of satellites for Earth observation are regularly acquiring single frequency, fixed polarization SAR imagery which is employed in a variety of applications. These radars operate by transmitting and receiving the same (linear) polarization so they are able to measure, for a given frequency, only one co-polarized (HH or VV) complex element of the scattering matrix, also called Sinclair matrix.
In the microwave region, where usually radars operate, the module of this element, which is proportional to the radar cross section (RCS) of the target for the given polarization state, is strongly affected by the physical/electrical properties of the surface under observation so it can be widely employed in a variety of disciplines concerned with remote sensing. On the other hand its (absolute) phase, if properly processed, may be used to derive the topography of the imaged scene (SAR interferometry) as well as to detect small vertical variations in the terrain height (SAR differential interferometry).
Recently, during the SIR-C/X-SAR shuttle missions of 1994, a great amount of multifrequency full-polarimetric SAR data have been collected from space by a dual-channel, dual-frequency SAR developed by the NASA/JPL [1]. This radar was designed to measure simultaneously amplitude and phase of the signal on the co-polarized and cross-polarized channels: first a pulse horizontally (H) polarized is transmitted and the echo is received with both H and V polarizations, then a pulse vertically (V) polarized is transmitted and the echo is again received with both H and V polarizations; in such a way the four receive/transmit combinations of two orthogonal and linear polarization states are available and a full complex scattering matrix can be measured [2].
The information content of this vectorial measurement greatly extends the capabilities offered by a scalar measurement to such an extent that it is really difficult, if not impossible, to define an "optimum" strategy to handle polarimetric SAR data. Nevertheless, one of the most interesting possibilities offered by polarimetric data is that, according to a given model, the contributions due to different scattering mechanisms acting inside the resolution cell can be separated.
Target decomposition (TD) theory is concerned with this approach to polarimetric data analysis and provides the necessary formalism to the techniques for achieving a decomposition.
In the recent literature three formulations of TD analysis have been proposed based on three different target matrix representations: the *scattering*, the *Mueller* and the *covariance* matrices [3]. Even though it is not the purpose of this work to investigate or to compare these techniques, a subject treated extensively by several authors [4, 5, 6], it should be noted that a concept basically drives the development of each TD technique: the matrix is decomposed in a sum of matrices and it is expected that these matrices will contain information about significant geometrical properties of the target.
Some of these TD methods proved to be useful in the detection of deterministic targets in a stochastic background (or clutter) but there is little matter about their relative performances in the characterization of natural targets. It may be noted that, if the morphology of the scene imaged by the radar does not affect the distribution of local incidence angles dramatically, natural surfaces are, in general, easily recognizable with respect to built-up areas, due to their very different scattering behavior; the discrimination among natural surfaces is a quite more difficult task, particularly at higher frequencies [7].
In the following an attempt will be made to characterize such a surfaces with a TD approach: in particular we shall solve the characteristic polynomial associated to a simplified form of covariance matrix. To this purpose the Sele river plain (southern Italy), imaged during the first SIR-C/X-SAR mission, has been selected as study area and a number of sample targets have been individuated. Each target like water, vegetation, bare soil and trees, has been described as a sum of the contributions due to three simple scattering mechanisms, defined by the eigenvectors of the covariance matrix and weighted by its eigenvalues. The applied TD technique draws its inspiration from the results achieved by Cloude [8] and van Zyl [9] and it provides an efficient tool to investigate the scattering properties of natural targets.

3. THE DECOMPOSITION OF COVARIANCE MATRIX

It is known that in a generic orthogonal basis AB the scattering matrix **S** relates the transmitted (t) to the received (r) electric fields and if the path losses are neglected the following expression holds:

$$\begin{bmatrix} E_A^r \\ E_B^r \end{bmatrix} = \begin{bmatrix} S_{AA} & S_{AB} \\ S_{BA} & S_{BB} \end{bmatrix} \begin{bmatrix} E_A^t \\ E_B^t \end{bmatrix} \quad (1)$$

which constitutes the fundamental equation of radar polarimetry. The most natural choice of the basis AB is of course a linear HV basis since linear polarizations are the simplest to be obtained by the antenna engineer. Equation (1) is then assumed in the form:

$$\begin{bmatrix} E_H^r \\ E_V^r \end{bmatrix} = \begin{bmatrix} S_{HH} & S_{HV} \\ S_{VH} & S_{VV} \end{bmatrix} \begin{bmatrix} E_H^t \\ E_V^t \end{bmatrix} \quad (2)$$

Paper presented at the AGARD SPP Symposium on "Remote Sensing: A Valuable Source of Information", held in Toulouse, France from 22-25 April 1996, and published in CP-582.

since the radar measures the complex elements of **S** in such a basis. Some authors carry out their own experiments in a circular (RL) basis; in this case an unitary transformation must be applied to matrix **S** in (2).
It is interesting to note that in backscattering measurements the scattering matrix should be symmetrical since reciprocity dictates that if the receive and transmit antenna polarization states are interchanged no variations in the received echo must be observed ($S_{HV} = S_{VH}$); however, due to a number of factors such as the delay existing between the HV and VH acquisitions, actual measurements of the cross-polarized terms show that in general $S_{HV} \neq S_{VH}$. The JPL, for example, forced this symmetry [10] prior to the delivery the polarimetric images acquired during the AIRSAR campaigns.
The covariance matrix [11,12] describes through variance and correlation the statistical properties of the scattering matrix elements and may be defined in terms of a set of simple 2x2 matrices α_i through the polarimetric vector Ω whose components are:

$$\Omega_i = \mathrm{Tr}(\mathbf{S}\cdot\alpha_i) \qquad i = 1, 2, 3 \tag{3}$$

The covariance matrix **T** is then defined according to the expression:

$$\mathbf{T} = <\Omega\,\Omega^{+}> \tag{4}$$

where the symbol <...> denotes a time/spatial average and the symbol + the transpose complex conjugate. Equations (3) and (4) provide the formal link between scattering and covariance matrices.
In principle, as pointed out by Cloude [8], any set $\alpha^{(i)}$ satisfying the requisites of completeness and normalization can be used to form a basis for the expansion of **S** but usually the choice is restricted to a set having a physical significance such as:

$$\alpha_1^{(1)} = A_1^{(1)}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix};\ \alpha_2^{(1)} = A_2^{(1)}\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix};\ \alpha_3^{(1)} = A_3^{(1)}\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix} \tag{5.1}$$

$$\alpha_1^{(2)} = A_1^{(2)}\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix};\ \alpha_3^{(2)} = A_3^{(2)}\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix};\ \alpha_2^{(2)} = A_2^{(2)}\begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix} \tag{5.2}$$

where symmetry ($S_{HV} = S_{VH}$) is taken into account.
The constants A_i must preserve the total power received by the radar which is invariant with respect to any change of basis (*Span*(**S**)=*Trace*(**T**)=*const*).
It is evident that each matrix α_i in the sets (5.1) and (5.2) describes an elementary reflection property of the target. The set (5.1) represents the Pauli basis where the α_i can be viewed as the scattering matrices of a simple targets such as a sphere or a trihedral corner reflector (first matrix in (5.1)), a dihedral corner reflector (second matrix in (5.1)) and a target which transforms H and V polarizations in their orthogonal states (both third matrices in (5.1) and (5.2)).
On the other hand the set (5.2) has a different interpretation since the first two matrices are representative of targets like cylinders with horizontal and vertical orientations respectively. The latter set derives from the so-called lexicographic expansion of **S** and, if no correlation exists between co-polarized and cross-polarized returns, may be used to build a very simple form of covariance matrix:

$$\mathbf{T} = C\begin{pmatrix} \dfrac{<|S_{HH}|^2>}{C} & 0 & \dfrac{<S_{HH}S_{VV}^{*}>}{C} \\ 0 & \dfrac{2<|S_{HV}|^2>}{C} & 0 \\ \dfrac{<S_{VV}S_{HH}^{*}>}{C} & 0 & \dfrac{<|S_{VV}|^2>}{C} \end{pmatrix} =$$

$$= C\begin{pmatrix} \dfrac{1}{\zeta} & 0 & \rho \\ 0 & \eta & 0 \\ \rho^{*} & 0 & \zeta \end{pmatrix} \tag{6}$$

where:

$$C = \sqrt{<|S_{HH}|^2><|S_{VV}|^2>}$$

$$\zeta = \frac{<|S_{VV}|^2>}{C}$$

$$\eta = \frac{2<|S_{HV}|^2>}{C}$$

$$\rho = \frac{<S_{HH}S_{VV}^{*}>}{C}.$$

In (6) the use of a constant C allows to introduce in the matrix **T** the coefficient of correlation ρ between the co-polarized channels. If set (5.1) was used instead of set (5.2) a more complicated form of (6) would be found.
The (unique) decomposition of covariance matrix may be written as

$$\mathbf{T} = \lambda_1\mathbf{k}_1\mathbf{k}_1^{+} + \lambda_2\mathbf{k}_2\mathbf{k}_2^{+} + \lambda_3\mathbf{k}_3\mathbf{k}_3^{+} \tag{7}$$

where λ_i and $\mathbf{k}_i$ are respectively the eigenvalues and the eigenvectors of **T** given by

$$\lambda_1 = \frac{C}{2}\left(\xi + \sqrt{\Delta}\right);\quad \lambda_2 = \frac{C}{2}\left(\xi - \sqrt{\Delta}\right);\quad \lambda_3 = C\eta \tag{8}$$

$$\mathbf{k}_1 = \sqrt{\frac{\left(\tau+\sqrt{\Delta}\right)^2}{\left(\tau+\sqrt{\Delta}\right)^2 + 4|\rho|^2}}\begin{pmatrix}\dfrac{2\rho}{\tau+\sqrt{\Delta}}\\ 0\\ 1\end{pmatrix};$$

$$\mathbf{k}_2 = \sqrt{\frac{\left(\tau-\sqrt{\Delta}\right)^2}{\left(\tau-\sqrt{\Delta}\right)^2 + 4|\rho|^2}}\begin{pmatrix}\dfrac{2\rho}{\tau-\sqrt{\Delta}}\\ 0\\ 1\end{pmatrix}; \tag{9}$$

$$\mathbf{k}_3 = \begin{pmatrix}0\\ 1\\ 0\end{pmatrix}$$

with $\xi = \dfrac{\zeta^2+1}{\zeta};\quad \tau = \dfrac{\zeta^2-1}{\zeta};\quad \Delta = \tau^2 + 4|\rho|^2.$

Since the covariance matrix is hermitian and semi-definite positive, its eigenvalues are real and its eigenvectors are orthogonal. Note that Eq. (8) holds also if set (5.1) is used for the definition of covariance matrix **T** since the eigenvalue spectrum does not depend on the choosen basis.
The eigenvectors (9) define a target space where only one eigenvector (the third) is constant; the other eigenvectors have two non-zero components and their amplitudes are essentially a function of ζ and |ρ| while the phase of the first (complex) component is the phase of ρ. In the target space $\mathbf{k}_1$ e $\mathbf{k}_2$ lie in a plane orthogonal to $\mathbf{k}_3$ and they rotate on this plane as the actual scattering mechanims varies.
As mentioned, some natural targets act like deterministic targets if they are sufficiently smooth at the wavelength scale or, more easily, if a condition of normal incidence occurs. Under these hypotheses ($|\rho| \cong 1$; $\zeta \cong 1$; $\eta \cong 0$) one mechanism is sufficient to describe the target so only one eigenvalue (λ_1) of the covariance matrix is not zero and the corresponding eigenvector ($\mathbf{k}_1$) is directed towards a principal direction, individuated by the phase of ρ: trihedral reflector (or sphere) if the phase is zero and dihedral reflector if the phase

approaches 180°.
If multiple reflections and/or volumetric scattering occur inside the reolution cell, λ_2 and λ_3 may increase significantly and $\mathbf{k}_1$, $\mathbf{k}_2$ begin to deviate from principal directions in the target space.
In general, if the sum $\lambda_{tot} = \lambda_1 + \lambda_2 + \lambda_3$ si defined the following condition holds for deterministic targets:

$$\frac{\lambda_1}{C} = \frac{\lambda_{tot}}{C} = \left(\frac{\lambda_{tot}}{C}\right)_{min} = 2 \qquad (10)$$

The opposite condition $\lambda_1 = \lambda_2 = \lambda_3$ often reported in the literature, identifies a completely stochastic target but has never been observed even though, as we shall see, some natural targets in L-band exhibit an eigenvalue spectrum which approximates the stochastic condition.
The eigenvalues of covariance matrix can be also expressed as ratio to λ_{tot} by:

$$\gamma_i = \frac{\lambda_i}{\lambda_{tot}} \qquad i=1, 2, 3 \qquad (11)$$

The first side of (9) can be viewed as the "dispersion" associated with the corresponding eigenvalue; note that for simple targets we have $\gamma_2 = \gamma_3 = 0$ while for stochastic targets $\gamma_1 = \gamma_2 = \gamma_3 = 1/3$. Eq. (11) allows to analyze the structure of covariance matrix with no regard to power information

4. SAR DATA AND STUDY AREA

The selected test area is the plain of Sele river (South Italy) where no effects due to morphology are expected; the area was imaged during the first SIR-C/X-SAR shuttle mission (April 1994). All the image/sensor parameters are shown in Tab.1.

Data take ID	098-20
Scene ID	Plain of Sele river
Frequencies	1.254 GHz and 5.287 GHz
Number of Looks	1
Nominal resolutions	21.3m (rng) x 6.1m (az)
Incidence angle (image center)	48.5°
Calibration	Absolute
Geometric representation	Slant range
Radiometric representation	Complex scattering matrix

Tab.1 - Sensor and image parameters.

The original single-look image has been multilooked in the azimuth direction to get a 4-looks image with a pixel approximatively square and nominal resolution of about 21m x 24m (rng/az). The incidence angle varies less than 1° in the azimuth direction, so it can be considered constant along both directions.
The test area is shown in Fig.1 where both C-band and L-band total power images are reported. The Sele river runs across a flat region and, after some bends, flows into the Tyrrhenian sea (upper side in Fig.1); here a pine-wood, looking as a bright stripe on the L-band image, separates the coast from the hinterland. Wild vegetation and soft trees cover both the banks and all the upper course of the river (lower side).
Fig.1b also shows a number of fields covered with vegetation (right side), uncovered (left side) and partially flooded as well as an urban agglomerate and two areas near the Sele river covered with a dense vegetation.
A look to Figg. 1a and 1b gives us a right feeling about a better overall performance of L-band in the characterization of natural targets. It is a simple matter to note that, at these incidence angles, the total power image appears more homogeneous at C-band than at L-band; at a wavelength of 6 cm, for example, the cross-polarized power weighs more in the sum of the total received power.

5. EXPERIMENTAL RESULTS

Turning now to the target analysis, each target shown in Fig.1 has been represented by a covariance matrix computed over a homogeneous window 11x11 samples wide; the elements of these matrices have been used, according to the expressions (6), (8) and (11), to generate the results reported in Tab.2 and Tab. 3 for L- and C-band respectively.
If we look at a general behavior of natural targets as a function of frequency we may note what follows:

- the ratio ζ increases with frequency. The average values are $\zeta_L = 0.944$ and $\zeta_C = 1.055$ but the power unbalance between the co-polarized channels may vary more significantly with frequency over particular targets;
- all the other parameters increase with frequency with the exception of λ_2 and γ_2. We have observed the following average values: $|\rho|_L = 0.228$, $|\rho|_C = 0.480$; $\lambda_{1L} = 1.372$, $\lambda_{1C} = 1.547$ ($\gamma_{1L} = 53.6\%$, $\gamma_{1C} = 57.9\%$); $\lambda_{3L} = 0.50$, $\lambda_{3C} = 0.71$ ($\gamma_{3L} = 18.7\%$, $\gamma_{3C} = 22.9\%$) while $\lambda_{2L} = 0.72$, $\lambda_{2C} = 0.50$ ($\gamma_{2L} = 27.6\%$, $\gamma_{2C} = 19.1\%$);
- the average λ_{tot} at C-band (2.76) is about the 7% greater then that at L-band (2.58).

As it can be noted none of the considered targets satisfies the condition of simple target which requires the symmetry of the target with respect the horizontal and vertical wave polarization states, as provided by a flat conducting plate, by a trihedral corner reflector or by a cloud of cilinders whose orientation angle is equally distributed in $[0,2\pi]$.
For what concerns natural target a "deterministic" condition can be easily observed over hilly terrains where local slopes may reduce significantly the incidence angle.

5.1 THE EFFECTS OF INCIDENCE ANGLE

To illustrate the role played by the incidence angle it may be interesting to consider the plots of Fig.2 where, for the case of sea scattering, the dispersions computed according to expression (11) are reported as a function of incidence angle. Here the considered data take was acquired during the AIRSAR '91 airborne campaign (June 28, 1991) by an early version of the radar that flew on board the shuttle. The image (ID CM3389; P-, L- and C-band full-polarimetric; 4-looks; pixel spacing rng 6.66m, az 12.5m) is relative to the island of Ischia and has been considered since the incidence angle varies significantly ($23° < \theta < 62°$) across the swath.
The covariance matrices have been computed over 121 azimuth samples in order to keep the incidence angle costant among different samples. The results for $\theta = 25°$ are reported in Tab.4 where it may be observed that a smooth sea surface is a quite deterministic target.
At the near range where the incidence angle does not exceed 32°, γ_1 is very close to unity ($\gamma_1 > 0.9$) while γ_2 and γ_3 are very low, especially in L-band. If the incidence angle is increased up to about 50° we measure at C-band $\gamma_1 = 0.7$, $\gamma_2 = 0.2$, $\gamma_3 = 0.1$; even though the sea surface is quite smooth the C-band eigenvalues indicate that the sea is no longer a "simple" target if $\theta = 50°$.
If the surface is rougher or if the frequency is increased the target deviates from a "deterministic" condition at lower incidence angles. For example, the SIR-C/X-SAR sea scattering results of Tabb. 2 and 3 are very different to those of Fig.2 relative to $\theta = 50°$, especially at L-band, due to different wave spectra of sea surface. Fig. 2 must be regarded as a qualitative description of the effects of the incidence angle.
In general the maximum incidence angle at which a target may be considered "deterministic" depends on the particular target but also on the degree of approximation which can be tolerated.

5.2 DISCUSSION

Turning now to the analysis of natural targets and with reference to Tab.2 and Tab. 3, it can be noted that samples labeled "Field 3" at L-band and "Vegetation 2" at C-band exhibit the lower values of λ_{tot} (2.324 and 2.224 respectively); they correspond to a flat, unvegetated and semi-flooded field (Field 3) and to a thick vegetated area (Vegetation 2).
Since the shuttle flew over the test site during a rainy week, we expect over bare soils no significant effects due to volumetric

scattering, also in L-band. This means that the surface roughness of "Field 3", as viewed at a given wavelength, increases with frequency. This is clearly shown by the fact that, as the frequency increases, λ_3, γ_3 and λ_{tot} increase while λ_1 and γ_1 decrease.
If the target is constituted by thick vegetation its scattering properties change dramatically with frequency. In C-band the dimensions of leaves and of small branches are usually comparable with the wavelength and the upper layer of the vegetation acts like an irregular surface with a large "facet" component. this is confirmed by observing that:

- the target is symmetric ($\zeta_C \cong 1$);
- the HH and VV mechanisms are correlated to some extent ($|\rho|_C > 0.7$);
  γ_{3C} is only about 10% of $\lambda_{tot} = 2.224$;
- γ_{1C} is greater than 77%.

In L-band, due to the greater wavelength (23cm), a significant contribution from the lower layer of the vegetation appears and also the soil may contribute to the backscattered energy. In this case we have:

- the target is asymmetric ($\zeta_L \cong 0.5$);
- the correlation between HH and VV mechanisms is low ($|\rho|_L \cong 0.2$);
- $\gamma_{3L} \cong 2\gamma_{3C}$; $\gamma_{2L} \cong 2\gamma_{2C}$;
- γ_{1L} is decreased to about the 50%.

Finally we must consider two targets that exhibit some peculiarities such as the pine-wood and the urban, even if the latter connot be classified as a natural target.
The pine-wood, with its cloud of needles, approaches the condition of completely stochastic target ($\lambda_1 \cong \lambda_2 \cong \lambda_3$) but only in L-band. It may observed that, at least at these incidence angles, the scattering from very rough surfaces may lead to very high values of λ_3 (as for C-band sea scattering) but an additional large volumetric effect is necessary to produce a covariance matrix with degenarate eigenvalues. For pine-wood the following applies:

- the target is symmetric ($\zeta_L \cong 1$);
- the HH and VV mechanisms are not correlated ($|\rho|_L \cong 0$);
- γ_{1L} reaches the minimum value (39%).

The urban area shown in Fig.1, as it can be expected, exhibits a number of peculiarities and can be easily discriminated from natural targets. A typical result describes built-up areas as:

- the target is asymmetric ($\zeta_C \cong \zeta_L \cong 0.5$)
- the correlation between HH and VV mechanisms is greater than 0.5
- $\lambda_{3L} \cong \lambda_{3C} \cong 0$;
- λ_{1L} and $\lambda_{1C} > 2$.

6. CONCLUSIONS

The decomposition of the covariance matrix in terms of its eigenvalues and eigenvectors has been applied to a number of targets selected over a test area in Campania (Italy) and imaged during the first SIR-C/X-SAR shuttle mission. Such a decomposition allows to interpret the target as a coherent and weighted sum of three simple scattering mechanisms. One of these mechanisms is costant among different targets and has a physical significance since it corresponds to a target that trasforms horizontal in vertical polarization and vice versa. The other two mechanisms are coupled because they are a function of the same variables: the power unbalance and the coefficient of correlation between the co-polarized channels.
A target is termed "deterministic" if only one eigenvalue of its covariance matrix is not zero. In the case of natural targets this usually occurs if a condition of normal incidence is verified, regardless of SAR operating frequency. In these circumstances neither unbalance nor decorrelation exist between the co-pol channels and the targets acts as a flat plate or, equivalently, as a thriedral reflector. As the incidence angle increases, also smooth targets deviate from the deterministic condition even though for such a targets the first eigenvalue still dominates the other.
The comparison of L- and C-band results shows that, over the test area, the contributions of surface and volume scattering may evaluated to some extent and that only the presence of both mechanisms allows to reach the condition of stochastic target.
The proposed approach has proven to be a very interesting tool for the interpretation of target properties

7. AKNOWLEDGEMENTS

The authors thank JPL/NASA for providing the AIRSAR and SIR-C/X-SAR data, and the Italian Space Agency (ASI) for the financial support (contract ASI RS94).

8. REFERENCES

[1] "SIR-C Science Plan" - Jet Propulsion Laboratory, 1986
[2] M. W. Boerner, W. L. Yan, A. Q. Xi, Y. Yamaguchi - "Basic Concepts of Radar Polarimetry" - *Direct and Inverse Methods in Radar Polarimetry*, Boerner&Kluwer Academic Publishers, pp.155-245, 1992.
[3] S. R. Cloude - "Measurements and Analysis in Radar Polarimetry" - *Direct and Inverse Methods in Radar Polarimetry*, Boerner&Kluwer Academic Publishers,pp.773-791, 1992.
[4] W. M. Boerner, C. L. Liu, X.Zhang - "A Rigorous Optimization Procedure for Alternate Covariance Matrix Formulations in Radar Polarimetry" - *European Telecommunications Journal*, Special Issue on Radar Polarimetry, 1992.
[5] J. R. Huynen - "Comments on Target Decomposition Theorems" - *Direct and Inverse Methods in Radar Polarimetry*, Boerner&Kluwer Academic Publishers, pp.387-399, 1992.
[6] E. Krogager - "Decomposition of the Sinclair Matrxi into Fundamental Components with Application to High Resolution Radar Target Imaging" - *Direct and Inverse Methods in Radar Polarimetry*, Boerner&Kluwer Academic Publishers, pp.1459-1477, 1992.
[7] W.A. Holm, R.M. Barnes - "On Radar Polarization Mixed Target State Decomposition Techniques" - IEEE 1988 Radar Conference.
[8] L. Castellano - "Nuovi metodi di analisi di dati SAR polarimetrici ripresi da piattaforme aerospaziali" - *Ph.D. Dissertation* - National Libraries of Rome and Florence - 1994.
[9] Cloude S.R. - "Uniqueness of Target Decomposition Theorems in Radar Polarimtery" - W.M.Boerner ed. - Kluver publ. - pp.267-296 - 1992.
[10] J. J. van Zyl - "Application of Cloude's Target Decomposition Theorem to Polarimetric Imaging Radar Data" - *Proceedings of SPIE, Vol.1748, Radar Polarimetry*, San Diego CA, pp. 208-216, July 1992.
[11] Elachi C. et al. - "Radar Polarimetry for Geoscience Applications" - Artech House, 1990.
[12] E. Rignot, P. Dubois - "Unsupervised Segmentation of Polarimetric SAR Data Using the Covariance Matrix"- *IEEE Transactions on Geoscience and Remote Sensing* Vol. 30, N.4, 1992
[13] M. W. Boerner, C. L. Xin Zhang - "Comparison of Optimization Procedures for 2x2 Sinclair, 2x2 Graves, 3x3 Covariance and 4x4 Mueller Matrices in Coherent Radar Polarimetry and its Application to Target Versus Background Discrimination in Microwave Remote Sensing and Imaging", ESA/EARSEL *Advances in Remote Sensing*, Vol.1, pp. 55-82, 1/1993.

Sample	ζ	$\lvert\rho\rvert$	λ_1 (γ_1)	λ_2 (γ_2)	λ_3 (γ_3)	λ_{tot}
Sea	0.905	0.085	1.144 (43.2%)	0.869 (32.7%)	0.639 (24.1%)	2.650
Pine-wood	0.956	0.081	1.093 (39.2%)	0.909 (32.8%)	0.779 (28.0%)	2.781
Field 1	0.995	0.265	1.265 (50.7%)	0.735 (29.4%)	0.496 (19.9%)	2.496
Field 2	0.987	0.047	1.049 (42.0%)	0.951 (38.0%)	0.500 (20.0%)	2.500
Field 3	1.598	0.568	1.859 (79.4%)	0.364 (15.6%)	0.118 (05.0%)	2.342
Veg. 1	0.922	0.215	1.233 (45.7%)	0.773 (28.7%)	0.690 (25.6%)	2.697
Veg. 2	0.723	0.189	1.433 (53.9%)	0.673 (25.3%)	0.552 (20.8%)	2.658
Urban	0.509	0.522	2.131 (85.1%)	0.341 (13.6%)	0.032 (01.3%)	2.505

Tab. 2 - Decomposition of covariance matrix: L-band results (1994 SIR-C/X-SAR mission NASA/JPL/DLR/ASI).

Sample	ζ	$\lvert\rho\rvert$	λ_1 (γ_1)	λ_2 (γ_2)	λ_3 (γ_3)	λ_{tot}
Sea	1.218	0.611	1.662 (43.4%)	0.378 (10.1%)	1.788 (46.5%)	3.828
Pine-wood	1.098	0.238	1.261 (45.3%)	0.748 (26.9%)	0.774 (27.8%)	2.783
Field 1	1.176	0.459	1.501 (58.4%)	0.526 (20.5%)	0.542 (21.1%)	2.568
Field 2	1.060	0.440	1.446 (58.5%)	0.558 (22.6%)	0.466 (18.9%)	2.470
Field 3	1.102	0.202	1.229 (50.6%)	0.781 (32.1%)	0.420 (17.3%)	2.429
Veg. 1	1.055	0.373	1.378 (57.6%)	0.624 (26.1%)	0.389 (16.3%)	2.391
Veg. 2	1.002	0.718	1.718 (77.2%)	0.282 (12.7%)	0.224 (10.1%)	2.224
Urban	0.567	0.672	2.065 (86.8%)	0.266 (11.2%)	0.047 (02.0%)	2.378

Tab. 3 - Decomposition of covariance matrix: C-band results (1994 SIR-C/X-SAR mission NASA/JPL/DLR/ASI)

Band	ζ	$\lvert\rho\rvert$	λ_1 (γ_1)	λ_2 (γ_2)	λ_3 (γ_3)	λ_{tot}
L	1.044	0.962	1.964 (97.1%)	0.039 (1.9%)	0.019 (1.0%)	2.021
C	1.085	0.897	1.904 (94.5%)	0.103 (5.1%)	0.007 (0.4%)	2.015

Tab.4 - Sea scattering results at 25° incidence angle (1991 AIRSAR Campaign - NASA/JPL/ASI)

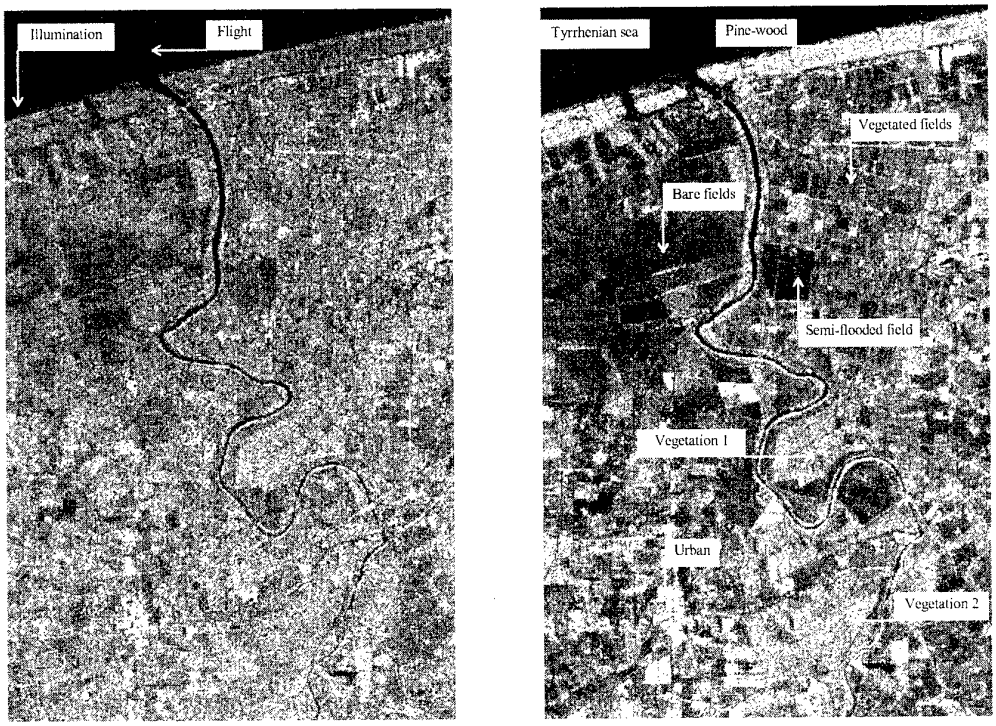


Fig.1 - SIR-C total power images of the Sele river plain (Italy). The data set was acquired during the shuttle mission in April 1994 by the NASA JPL polarimetric SAR operating at C-band (a) and L-band (b).

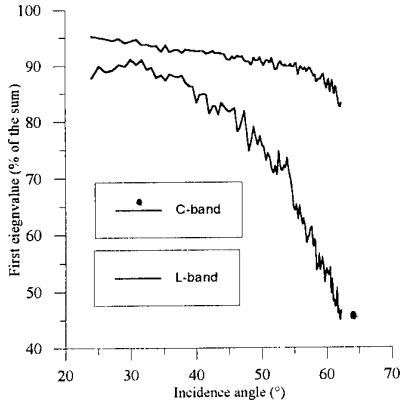


Fig. 2a - Sea scattering. The dispersion γ_1 as a function of incidence angle. 121 samples average.

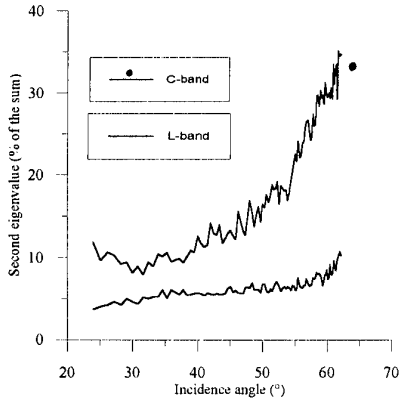


Fig. 2b - Sea scattering. The dispersion γ_2 as a function of incidence angle. 121 samples average.

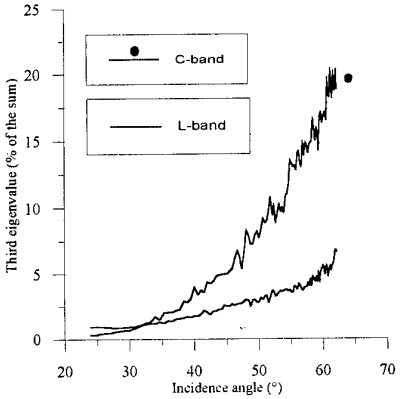


Fig. 2c - Sea scattering. The dispersion γ_3 as a function of incidence angle. 121 samples average.



Fig. 3 Brigthness (red), greenness (green) and wetness (blue) trasformation of a LANDSAT/TM (June 1994) image of the Sele river area.

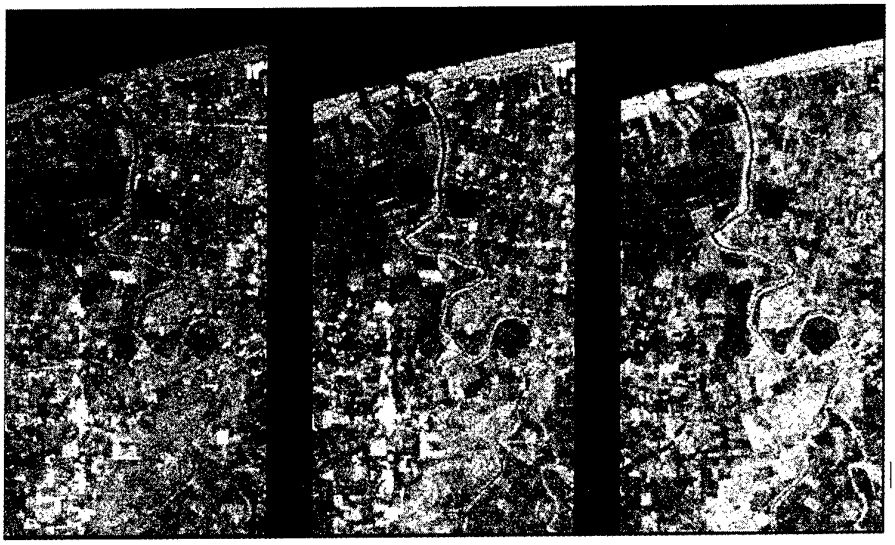


Fig. 4 - First, second and third eigenvalue images of the Sele river plain - L-band eigenvalues are represented in red and C-band eigenvalues are represented in green+blue.

Paper 19

E. Krogager (DE)

Comment

Thank you for your interesting presentation which showed another important way of utilizing polarimetric data in terms of power descriptors as compared to the complex voltage descriptors that I used in my previous presentation. It is interesting to try to compare these different approaches, and one may find that one approach is most suited for some applications, and others for other applications.

Microwave Remote Sensing of Natural Targets : Theory and Experiment

Jean-Claude SOUYRIS - Thuy LE TOAN - Nicolas FLOURY
CESBIO, 18 Av. Edouard Belin, BPI 2801, 31055 Toulouse Cedex, France
Ph : +33.61.55.85.84 E-mail : souyris@cesbio.cnes.fr

Chih-Chien HSU - Li-Fang WANG - Jin Au KONG
Department of Electrical Engineering and Computer Science,
Massachusetts Institute of Technology, Cambridge, MA 02139, USA
Ph : +1.617.253.8209 E-mail : chih@ewt.mit.edu

1. SUMMARY

The overall objective of this work is the study of radar backscattering from forest canopies and soil surfaces, in order to assess the use of multi-frequency, multi-incidence, and multi-polarisation radar data for the understanding of natural media.

To this aim, polarimetric data acquired over the Landes forest test site by the Spaceborne Imaging Radar-C/XSAR (polarimetric L- and C- bands , X band VV) are analysed. The results are interpreted by a model based on radiative transfer theory coupled with a branching model taking into account coherent scattering effects due to the clustered vegetation structures. The modeling is further used to determine the major contributions in the tree backscattering mechanism and the wave attenuation through the canopy, for a given configuration of the radar signal (frequency, incidence, polarization).

Further to the use of intensity information (i.e. backscattering coefficients σ^ohh, σ^ovv, σ^ohv) for clutter characterization (forest, non forest, water, urban areas), the information content of polarimetric measurements is assessed. Using polarisation synthesis techniques, a collection of polarimetric discriminators is calculated. In several cases, the polarimetric information is shown to improve clutter segmentation. The following results are derived : *1)* The degree of coherence at C band between HH and VV signals provides the best results for forest/non forest segmentation. *2)* L band HV is the most appropriate radar configuration for biomass estimation. *3)* The use of circular polarised intensity signals enhances the roughness effect over bare soils surfaces.

2. INTRODUCTION

For several years, radar data have been proved to be of great interest for remote sensing applications in which estimation of physical parameters is addressed [1]-[5]. Compared to data derived from optical sensors *(SPOT, LANDSAT TM)* which are usually more suited to classification purposes, radar backscattering holds several advantages : *1)* The observation is independent of weather and sun illumination conditions. *2)* The radar wave has penetration capabilities. The penetration depth, which gets larger as the wavelength increases (i.e. frequency decreases) potentially permits detection of deeper volume within the media as well as detection of buried targets *3)* Through the dielectric constant, the electromagnetic backscattering is sensitive to the medium water content. *4)* For remote sensing applications, the radar wave is traditionnally included in the centimetric wave spectrum. At this scale, the corresponding media geometrical features (surface roughness, tree structure, ...) affect the radar backscattering.

Based on these potentialities, the overall objective of this work has been to study the backscattering from forest canopies and soil surfaces, and to assess the use of multi-frequency, multi-incidence, and multi-polarisation radar data for the understanding of natural media. For this, polarimetric properties of the Landes forest test site in South-Western France have been investigated through the extensive analysis of a data set provided by the *Spaceborne Imaging Radar-C/XSAR (SIR-C/XSAR)*, which has flown aboard the space shuttle Endeavour in April and October 1994. *SIR-C/XSAR* is the latest of a series of experimental spaceborne imaging radars, beginning with *SIR-A* (1981) and *SIR-B* (1984). *SIR-C/XSAR* moreover marks a significant step as it is the first spaceborne radar capable of acquiring simultaneous multifrequency and mutipolarisation data. It consists of two radars operating at *L* band *(1.28 GHz)* and *C* band *(5.17 Ghz)* with four polarisations. The *XSAR* system is made of a radar operating at *X* band *(9 Ghz)* and *VV* polarisation.

The Landes experiment data sets have been analysed. Results have been interpreted by a theoretical modeling [6]-[7]-[8]. The model is based on the radiative transfer theory coupled with a branching model. Coherent scattering effects due to the clustered vegetation structures are taken into account. The radiative transfer equations are solved iteratively up to the second order, in order to provide estimations of backscattering coefficients for co- and crosspolarisation, as well as polarimetric channels correlation coefficients. The tree structure is modeled by multiscale dielectric cylinder clusters. Coherent scattering effect due to clustered vegetation structure is shown to be important for co-polarised returns. Also, to estimate the cross-polarisation scattering return accurately, as well as polarimetric correlation coefficients, the multiple scattering effects generated by second order contributions need to be taken into account. The theoretical modeling is moreover helpful to determine the major contributions in

Paper presented at the AGARD SPP Symposium on "Remote Sensing: A Valuable Source of Information", held in Toulouse, France from 22-25 April 1996, and published in CP-582.

the tree backscattering mechanism, for a given configuration of the radar signal (frequency, incidence, polarisation), but also to define the optimal combinations accordingly to specific objectives (biomass inversion, clutter segmentation, ...) as well as the wave attenuation through the canopy.

In most cases, the polarimetric information in remote sensing applications is only partly analysed. The coherence information between polarimetric channels (i.e. the phase difference information) is not fully taken into consideration. This is for example the case in recent works devoted to classification algorithms based on *SIR-C/XSAR* data. One of our objectives has been to focus on the polarimetric aspect of the measurement. To this aim, besides the use of intensity information (i.e. backscattering coefficients $\sigma^o hh$, $\sigma^o vv$, $\sigma^o hv$), the full polarimetric information is considered, in order to investigate the polarisation dependence of the backscattering from different natural media. It consists in deriving the target response for any set of transmitting and receiving polarisation, from the knowledge of the polarimetric response derived in a linear polarisation basis $(\hat{h}, \hat{v})$. For any polarisation basis, several polarimetric discriminators are calculated. Among others are the depolarisation ratio (i.e. the ratio between cross- and copolarised returns), as well as the degree of coherence between copolarised returns.

3. TEST SITE AND EXPERIMENTAL DATA

3.1 Study area and ground data

The site under test is the Landes forest in South-Western France. The forest covers nearly one million hectares on a flat topography, is almost totally formed of maritime pine (pinus pinaster), and is managed in a consistent fashion, in homogeneous forest stands.

The study area (Nezer test site) located within the Landes forest includes a collection of stands of various ages. The ground data available over the Nezer site consist in an updated reference biomass map, providing information about the location and age of more that *50* stands of maritime pines. The ages of these stands are ranging between *2* years old and *50* years old, while the corresponding biomass rate is included between *5* and *150 tons/ha*. For every forest age, a statistical description of the forest is derived from ground-truth measurements. This description provides the input parameters required in the theoretical model. They include density of trees per m^2, trunk radius inside and outside the crown, height of the trunk under first branch and total height of the trunk, radius, length and density *per m* of primary and secondary branches [2]. In addition, a clear-cut map is also available. For some of the clears-cuts, ground truth measurements including surface soil moisture in a *(0-5 cm)* layer have been performed, as well as surfaces profiles providing surface correlation function, rms height and correlation length.

3.2 SIR-C/XSAR data

The different passes of *SIR-C/XSAR* mission over Les Landes provided us with *L* and *C* bands Single Look Complex (*SLC*) full polarimetric data at *24.5°* and *26.4°* of incidence, completed by *X* band *VV* responses. In addition, *L* and *C* bands *HH* and *HV* responses are available for *54 °*. As an example, in the case of *26.4° SLC* images, the nominal range resolution is *17.94 m*, and the nominal azimuthal resolution is *8.18 m*. Line spacing and pixel spacing are *5.22 m* and *6.66 m* respectively. In this paper, only *SIR-C* data have been addressed. *XSAR* data analysis remains a topic for further studies. The calibration of data has been assessed, based on the comparison with calibrated data acquired during *Maestro1* (1989) and *Mac-Europe* (1991) campaigns [9].

4. THEORETICAL MODELING

This section summarizes a four-layer Radiative Transfer (RT) model constructed for the modeling of pine forest [7]. The four layers include a crown layer, a trunk layer, an understory layer, and a ground interface. The ground is modeled by a random rough surface. The Kirchhoff approximation is used to calculate the scattering from the random rough surface. The permittivity of soil is calculated from the soil model [10] with 0.2 g/cm^3 volumetric water content. The trunk is modeled as a tilted circular cylinder, branches and needles are modeled as circular cylinders where finite cylinder approximation is utilized [11]-[12]. The permittivity of trunk, branches and leaves is calculated using dielectric model [13] with gravimetric moisture content of 0.55 g/cm^3. The scattering properties of structured pine trees are taken into account in the model by incorporating the branching model [6] into the phase matrix of the RT equation. The vector radiative transfer equation for the specific intensity in each scattering region is of the form :

$$\cos\theta \frac{d\bar{I}(\theta,\phi,z)}{dz} = -\bar{\bar{\kappa}}_e(\theta,\phi,z).\bar{I}(\theta,\phi,z) + \int_{4\pi} d\Omega' \bar{\bar{P}}(\theta,\phi,\theta',\phi').\bar{I}(\theta',\phi',z) \qquad (1)$$

where the Stokes vector $\bar{I}$ contains information regarding field intensity and phase relation of the two orthogonal polarisations and is defined as :

$$\bar{I} = \begin{pmatrix} I_h \\ I_v \\ U \\ V \end{pmatrix} = \frac{1}{\eta} \begin{pmatrix} \langle |E_h|^2 \rangle \\ \langle |E_v|^2 \rangle \\ 2\,Re\langle E_v E_h^* \rangle \\ 2\,Im\langle E_v E_h^* \rangle \end{pmatrix} \qquad (2)$$

In (2), the subscripts *h* and *v* represent the horizontal and vertical polarisations, respectively. The bracket $<\bullet>$ denotes ensemble average over the size and orientation distributions of scatterers and $\eta = \sqrt{\mu_o/\varepsilon_o}$ is the free space impedance. The extinction matrix $\bar{\bar{K}}_e$ represents the

attenuation due to both scattering and absorption, and can be obtained through the optical theorem in terms of forward scattering functions. The phase matrix $\overline{\overline{P}}(\theta,\phi,\theta',\phi')$ characterizes the scattering of the Stokes vector from (θ',ϕ') direction into (θ,ϕ) direction. The phase matrix can be formulated in terms of the scattering functions of the randomly distributed discrete scatterers.

The radiative transfer theory having the advantage of its simple formulation, the key issue is to evaluate the extinction matrix and the phase matrix accurately. For the tree which has specific structure, it is important not only to take into account the angular distribution of branches, but also their relative position which affects the overall scattering characteristics. In order to account for the coherent scattering effect due to clustered vegetation structures in the radiative transfer theory, the phase matrix based on the branching model of vegetation needs to be formulated. In pine forest, most of the scatterers are of cylindrical shape. Hence the main subject to be considered is clusters consisting of cylinders. For a cylinder cluster that has one center cylinder and N branching cylinders, the total backscattering function $f_{\alpha\beta}$ is :

$$f_{\alpha\beta} = f_{0\alpha\beta} + \sum_{n=1}^{N} f_{n\alpha\beta}\, e^{i\phi_n} \qquad (3)$$

where $f_{\alpha\beta}$ is the $\alpha\beta$-*th* element of the scattering matrix for the center cylinder, and $f_{n\alpha\beta}$ is $\alpha\beta$-*th* element of the scattering matrix for the n-th branching cylinder. $\alpha,\beta,\gamma,\delta$ represent horizontal or vertical polarisation.

Assuming all the branching cylinders are identical and independent of each other, and their relative position is independent of their scattering properties (which depends on size and orientation), the correlation of f is given by :

$$\left\langle f_{\alpha\beta} f^*_{\gamma\delta} \right\rangle = \left\langle f_{0\alpha\beta} f^*_{0\gamma\delta} \right\rangle + N\left\langle f_{n\alpha\beta} f^*_{n\gamma\delta} \right\rangle + 2N.Re\left[\left\langle f_{0\alpha\beta}\right\rangle\left\langle f^*_{n\gamma\delta}\right\rangle\left\langle e^{-i\phi_n}\right\rangle\right] + N(N-1)\left\langle f_{m\alpha\beta}\right\rangle\left\langle f^*_{n\gamma\delta}\right\rangle\left\langle e^{-i\phi_n}\right\rangle\left\langle e^{-i\phi_m}\right\rangle \qquad (4)$$

The relative phase of the n-th branch with respect to the center cylinder is defined as :

$$\phi_n = (\bar{k}_i - \bar{k}_s).\bar{r}_n \qquad (5)$$

where and $\overline{k_i}$ and $\overline{k_s}$ are the incident and scattered wave vectors, respectively. $\overline{r_n}$ is the location of the n-th branching cylinder relative to the center cylinder.

In (4), the third and fourth terms are the coherent scattering terms. It can be seen that the incoherent scattering approximation is valid when the average of the random phase factor $\left\langle e^{i\phi_j}\right\rangle$ is small so that the coherent terms are negligible compared to the incoherent terms. However, this is not true for vegetation structures with scale lengths comparable to the wavelength.

Using the above approach, the crown is modelled as a 4-scale cluster constituted of trunk, primary branches, secondary branches and needles [7]. Since the effective permittivity of vegetation layer is very close to air, there is no reflection between different layers except at the ground boundary. The radiative transfer equations with the boundary conditions can be solved iteratively for polarimetric backscattering coefficients [14].

5. DATA ANALYSIS

Based on the biomass map, *22* forest stands, ranging between *2* years old and *50* years old, and *5* areas of soil surfaces have been selected over the Nezer test site. Each of these areas covers at least *2000* pixels in the basic *SLC* image. For each of the *27* selected areas, the covariance matrix defined as below has been derived :

$$\left(\overline{\overline{C}}\right) = \begin{pmatrix} < f_{hh}\cdot f^*_{hh} >< f_{hh}\cdot f^*_{hv} > & < f_{hh}\cdot f^*_{vv} > \\ < f_{hv}\cdot f^*_{hh} >< f_{hv}\cdot f^*_{hv} > & < f_{hv}\cdot f^*_{vv} > \\ < f_{vv}\cdot f^*_{hh} >< f_{vv}\cdot f^*_{hv} > & < f_{vv}\cdot f^*_{vv} > \end{pmatrix} \qquad (6)$$

where *fhh*, *fvv* and *fhv* are the complex backscattering coefficients for *HH*, *VV* and *HV* polarisations respectively. The covariance matrix provides estimations of $\sigma^o hh$, $\sigma^o vv$, $\sigma^o hv$ (i.e. <*fhh fhh**>, <*fvv.fvv**>, <*fhv.fhv**>), as well as correlation terms between polarisations.

5.1 Intensity data analysis

5.1.1 Backscattering at 26.4° of incidence angle

Fig. 1-a-b-c and ***Fig. 2-a-b-c*** display experimental and theoretical backscattering coefficients $\sigma^o hh$, $\sigma^o vv$ and $\sigma^o hv$ for both forest stands and clear-cut areas at *L* band and *C* band respectively, at the incidence of *26.4°*. Clear-cuts correspond to areas for which biomass is lower than *5 tons/ha*. In the case of *L* band, the copolarized backscatter is relatively less sensitive to variations of biomass content (i.e. age of forest) (**Fig. 1-a-b**). The discrepancies between theory and experiment may be explained only partly by calibration uncertainties. The average $\sigma^o hh$ for forest is *-9.87 dBm²/m²*, whereas the average of $\sigma^o vv$ is *-10.74 dBm²/m²*. For *L* band *HV* (**Fig. 2-c**), the increase of backscatter responses as a function of forest age is more significant. Modeling results show the same trend. The backscatter response becomes saturated around *-15 dBm²/m²* for forest stands with biomass higher than *50 tons/ha*. The dynamic range of $\sigma^o hv$ measurement is about *6 dB*.

At *C* band, we observe a decrease of $\sigma^o hh$ and $\sigma^o vv$ when forest is growing older (***Fig.2-a-b***), and reaches about *50 tons/ha*. The response gets saturated around *-8 dBm²/m²*. This behavior is consistent with the theoretical model which takes into account a direct contribution from the soil for young forest stands. The soil contribution does not show up at *L* band since at lower frequency backscattering from the soil is not significantly higher than

backscattering from the vegetation. Its effect on the crosspolarized backscattering is also negligible (**Fig.2c**).

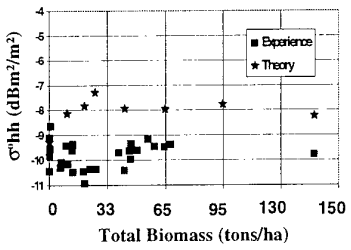


Fig.1-a : Experimental and theoretical $\sigma^{o}hh$ *for forest stands and soils surfaces. L band, 26.4°*

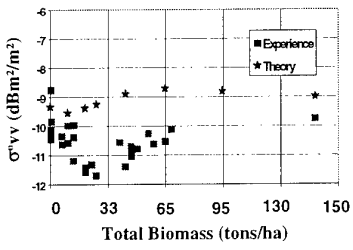


Fig.1-b : Experimental and theoretical $\sigma^{o}vv$ *for forest stands and soils surfaces. L band, 26.4°*

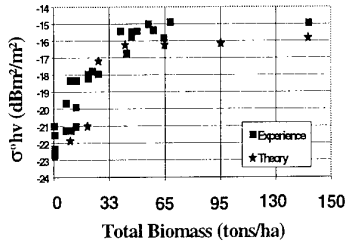


Fig.1-c : Experimental and theoretical $\sigma^{o}hv$ *for forest stands and soils surfaces. L band, 26.4°*

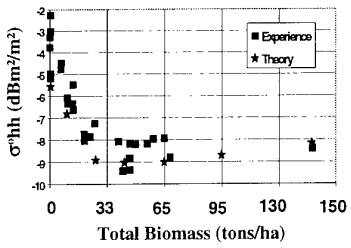


Fig.2-a : Experimental and theoretical $\sigma^{o}hh$ *for forest stands and soils surfaces. C band, 26.4°*

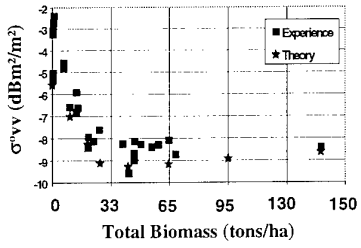


Fig.2-b : Experimental and theoretical $\sigma^{o}vv$ *for forest stands and soils surfaces. C band, 26.4°*

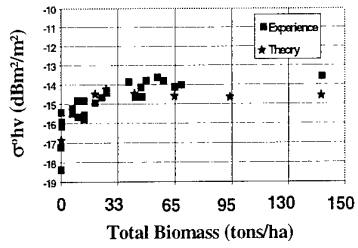


Fig.2-c : Experimental and theoretical $\sigma^{o}hv$ *for forest stands and soils surfaces. C band, 26.4°*

Once the theoretical modeling has been validated by experimental intensity data, it is further used in simulations in order to determine the major contributions in the backscattering process, for a given configuration (incidence, frequency, polarisation), and for various stages of the growth. As an example, ***Fig. 3*** displays the major backscattering mechanisms for *C* band *VV, 26.4°*. For young forest stands, a direct contribution from the soil is expected at low incidence. This direct contribution has been calculated using the Kirchhoff approximation [15], with ground parameters estimated during the field measurement campaign (moisture content of the understory soil is *22%*, r.m.s. height is *1.5 cm*, correlation length is *10 cm*, and the shape of the surface correlation function is exponential). The direct contribution from the soil drops when the forest gets older. Within the crown, the needle contribution is shown to be dominant, for forest stands with biomass larger that *50 tons/ha*. The reasons are the large quantity of needles, and their size which is comparable with wavelength at *C* band.

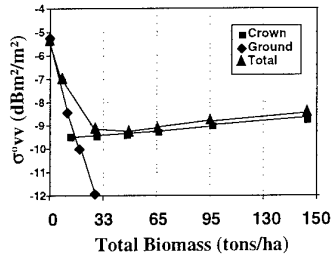


Fig. 3 : Decomposition of major backscattering mechanisms. C band, 26.4°, VV

5.1.2 : Backscattering at 54° of incidence angle

Only $\sigma^{o}hh$ and $\sigma^{o}hv$ responses are available at this incidence. ***Fig. 4*** and ***Fig. 5*** display experimental $\sigma^{o}hh$ and $\sigma^{o}hv$ at *L* and *C* band respectively. *L* band *HH* and *HV* responses show an increasing trend with regards to the biomass rate. In both cases, the dynamic range in the forest response is about *6 dB*, when biomass ranges between *0* and *65 tons/ha*. Among all the data available at this time, *L* band *HH* or *HV* at *54°*, and *L* band *HV* at *26.4°* have been found to be the radar configurations the most sensitive to the biomass content (i.e. forest age). These results are consistent with previous studies based on *AIRSAR* data analysis [1]-[2], indicating an enhanced sensitivity of the radar backscatter to biomass for *P* or *L* band, *HH* or *HV* polarisation. ***Fig. 6*** presents the backscattering mechanism decomposition for *L* band *HH*. The backscattering process is found to be mainly created by the crown scattering and the crown-ground interaction. The crown scattering is characterized by a strong dynamic range, and is consequently directly responsible for the increase in the *L* band *HH* response when forest is growing older. The model has been subsequently used to specify the contributions of each part of the crown (primary, secondary branches, needles) in the backscattering process. For *L* band *HH* or *L* band *HV*, it is found that the backscattering comes mainly from primary and secondary branches. The comparison between ***Fig. 3*** and ***Fig. 6*** highlights the complementarity between data acquired under diffrent angles of incidence which indicate dominant backscattering mechanisms of different types.

C band *HH* (***Fig.5***) is almost independent of forest age, and presents a backscatter response stable around *-11.5 dBm²/m²*. Unlike the *26.4°* case, the soil contribution does not appear for young forest stands, which explains the stability of the response. *C* band *HV* shows the same behavior, with a value close to *-16 dBm²/m²*. Whatever the polarisation, the needle contribution is shown to be dominant in the backscattering process.

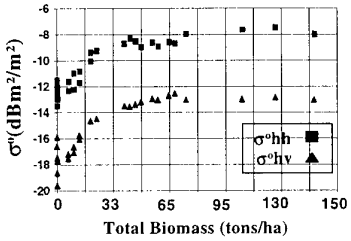


Fig.4 : Experimental and theoretical $\sigma^{o}hh$ for forest stands and soil surfaces, *LHH 54°*

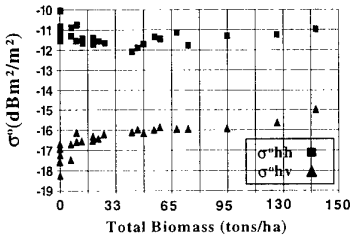


Fig.5 : Experimental and theoretical $\sigma^{o}hh$ for forest stands and soil surfaces, C*HH, 54°*

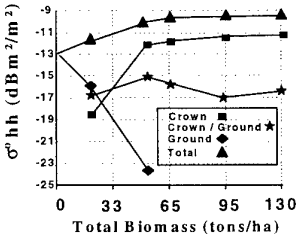


Fig.6 : Decomposition of major backscattering mechanisms, L band HH, 54°

5.1.3. Wave attenuation through the canopy :

Using the theoretical approach previously described, the objective of this section is to estimate the wave attenuation through the canopy, for a given configuration of the radar signal (frequency, incidence, polarisation). This study has been initiated during Mac-Europe campaign (1991), which provided P, L and C band data. P band (λ=68 cm) is of particular interest, since it is the lowest available frequency and consequently the one with the best penetration capabilities. **Table 1** presents the estimated P, L and C band one-way attenuation for forest stands of different ages at 25° and 50° of incidence. For each forest age, the average tree height and density of trees per hectar have been derived from ground data. The density of trees per hectare decreases versus forest age, while the total biomass increases. The drop of density versus forest age finally results in an total attenuation slowly varying versus forest age.

The major trends obtained through these simulations can be summarised in the following points : *1)* The attenuation increases versus frequency (i.e. the penetration depth gets larger when frequency decreases). *2)* For a given frequency and polarisation, the attenuation increases versus angle of incidence (the path of the wave within the forest canopy gets larger when the incidence increases). *3)* VV attenuation is slightly higher than HH attenuation.

age (years)	*freq*	*Att.H (dB) 25°*	*Att.V (dB) 25°*	*Att.H (dB) 50°*	*Att.V (dB) 50°*	*height (m)*	*density (tr./ha)*
6	P	1.342	1.929	1.770	2.827	5.18	1140
	L	2.964	3.022	4.189	4.3		
	C	13.414	13.476	19.092	19.163		
14	P	1.703	2.36	2.395	3.688	9.66	840
	L	3.404	3.522	5.002	5.170		
	C	10.435	10.502	15.104	15.190		
22	P	2.070	2.709	2.976	4.243	12.66	610
	L	3.566	3.721	5.436	5.535		
	C	10.313	10.365	14.388	14.637		
30	P	2.137	2.702	3.219	4.384	15.52	450
	L	3.618	3.767	5.650	5.746		
	C	9.718	9.761	13.587	13.673		
46	P	1.794	2.064	2.751	3.320	21.05	240
	L	2.947	3.048	4.635	4.75		
	C	6.344	6.397	8.899	9.003		

Table 1 : Simulated P, L and C band one-way attenuation (in dB) for forest stands of different ages at 25° and 50°.

By way of validation, two *2.5 m* trihedral corner reflectors have been positioned within a forest stand of age 18 years during Mac-Europe campaign under an incidence of 50°. Identical reflectors were simultaneously located on open fields. The comparison between their respective responses permitted to estimate the wave attenuation through the canopy. The experimental one-way attenuations are given in **Table 2**. The comparison between theory and experiment is acceptable, although P band attenuation appears to be under-estimated by the model. One reason could be the small amount of reflectors to make a reliable estimation

of the attenuation, especially at low frequency. A statistical approach based on the responses of a significant amount of corners reflectors would have been more appropriate.

freq	*Att.H (dB) 50°*	*Att.V (dB) 50°*
P	*1.5*	*1.65*
L	*6.35*	*7.3*
C	*15.6*	*14.9*

Table 2 : Experimental P, L and C band one-way attenuation for a 18 years old forest canopy. Incidence : 50°.

5.2 Polarimetric data analysis

In addition to the study of traditional backscattering coefficients ($\sigma^o hh$, $\sigma^o vv$, $\sigma^o hv$), the information content directly related to the polarimetric aspect of the measurement has been taken into consideration. Two polarimetric discriminators are tested :

- The depolarisation ratio :

$$r = \frac{< f_{hv} \cdot f_{hv}^{*} >}{< f_{vv} \cdot f_{vv}^{*} >}$$

- The normalised degree of coherence between *HH* and *VV* signals :

$$\rho = \frac{< f_{hh} \cdot f_{vv}^{*} >}{\sqrt{< f_{hh} \cdot f_{hh}^{*} > . < f_{vv} \cdot f_{vv}^{*} >}}$$

5.2.1 The depolarisation ratio

The variations of $\sigma^o hv/\sigma^o vv$ show that it constitutes a good discriminator between clear-cut areas and forest zones, either for *L* or C band. For both frequencies, forests are characterized by a ratio higher than *-11 dBm²/m²* at *26.4°* of incidence. As an example, ***Fig. 7*** displays *C* band experimental and theoretical variations of *r* versus biomass. In a previous study [16], the depolarisation ratio has already been proved to be a good discriminator for identifying trees from other types of surfaces. A consistent physical explanation underlying the efficiency of *r* is to consider it as a normalized ratio of the cross-polarized intensity, which is mainly generated by multi-path scattering mechanisms. High value of *r* results from a large amount of volume scattering with respect to surface scattering, which indicates presumably the presence of a canopy or a dense vegetated medium. Moreover, the effects of miscalibration on normalized terms such as *r* tend to be reduced. When the forest grows from *8 tons/ha* to *50 tons/ha*, the increase of $\sigma^o hv/\sigma^o vv$ is *-6.5 dBm²/m²* at *L* band, and *-5.5 dBm²/m²* at *C* band for *26.4°* of incidence.

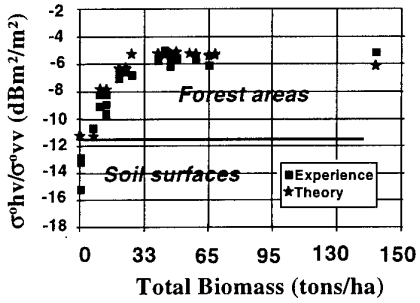


Fig. 7 : Experimental and theoretical $\sigma^o hv/\sigma^o vv$ for forest stands and soil surfaces. C band, *26.4°*

5.2.2 The degree of coherence

The degree of coherence between *HH* and *VV* polarisations, ρ, is also a parameter of interest. It is conceptually very different from the depolarisation ratio mentioned above. It is derived from a coherent measurement between co-polarized signals and does not require crosspolarisation information. On the contrary, the depolarisation coefficient is based on both co- and crosspolarized information, but does not require coherence information between measurements. ***Fig. 8*** represents experimental and theoretical variations of ρ for both forest stands and clear-cut areas at *C* band. Bare soil surfaces are characterized by a degree of coherence close to *1* in magnitude. When forest stands are considered, the magnitude of ρ decreases to a value of *0.35* for older stands. A value of *0.85* in the magnitude of ρ appears to be an appropriate threshold to separate bare soils responses from forest responses. *L* band correlation coefficient indicates a similar trend [17].

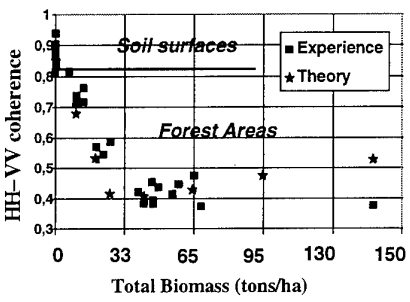


Fig. 8 : Experimental and theoretical ρ for forest stands and soil surfaces. C band, *26.4°*

In order to capture the behavior of ρ at *C* band over the Nezer on a per-pixel basis, *L* and *C* band coherence maps of the Nezer test site have been constructed. Prior to coherence mapping, a multilook processing is applied on the basic *SLC* image (*4* looks in azimuth, *2* looks in range). For every pixel of the resulting image, the local degree of coherence is further estimated throughout a moving window enclosing the pixel under consideration. The window size is adjustable. A trade-off is encountered between accuracy of the local estimation of ρ and the loss in resolution. To that extent, a window size of *3*3* pixels was found to be appropriate. Considering the generated *8* looks image, the local estimation of ρ is consequently made by averaging polarimetric responses of *72* pixels of

the basic *SLC* image. The amount of independent samples is nevertheless reduced, due to the oversampling of the image. L and C band coherence maps, showing an enhanced contrast between classes, compared with the corresponding color composite images generated from intensity information [18]. Based on these maps, the unsupervised extraction of forest stands lower than 33 tons/ha has been achieved with an accuracy higher than 85%, whereas forest/clear-cut segmentation indicated an accuracy of 87%.

The theoretical modeling of polarimetric backscattering from a forest canopy [7]-[8] is helpful to highlight the efficiency of ρ for forest / non forest mapping in the configuration *C* band, *26.4°* over Les Landes test site. The main reason appears in ***Fig.6***, where a direct contribution from the soil is expected for young forest stands at low incidence angle. As already discussed, the corresponding backscattering intensity of this contribution is very likely soil moisture and roughness dependent. In order to quantify this dependence, the Integral Equation Model (IEM) [15] has been used to calculate the direct contribution from the soil in the *ERS-1* configuration (which is very close to *C* band *VV*, *26.4°*), for varying moisture and roughness conditions (**Fig. 9**). Assuming exponentially shaped surface correlation functions, this contribution ranges between *-3.5 dBm²/m²* and *-12 dBm²/m²* when soil moisture varies between *5% (ε_r~3)* and *40% (ε_r~20)*, r.m.s height between *0.8 cm* and *1.5 cm*, and surface correlation length between *8 cm* and *12 cm*. The spreading of soil backscattering creates a zone of confusion which makes the biomass retrieving critical for the regenerating forest (e.g. forest stands for which biomass rate is lower than *40 tons/ha*).

When bounded to the above-mentionned soil conditions, the *HH-VV* correlation coefficient at *C* band indicates a much higher stability : compared to forests, soil surfaces (as well as water surfaces) have been experimentally found to exhibit large degree of coherence at any roughness and moisture conditions. In addition, the fact that ρ is a normalised parameter makes it less sensitive to miscalibration effects, especially to those related to radiometric uncertainties, as well as terrain local slope effects (the latest statement needs nevertheless further investigations to be fully assessed). For these reasons, ρ is consequently expected to exhibit more stability than intensity based discriminators when applied in varying conditions.

Using ρ at *C* band on a per-pixel basis over a large Landes area, the *C* band coherence map is established (**Fig. 10-a**),. In comparison, **Fig. 10-b** displays the corresponding *4*2* multilooked *C* band *VV*, *26.4°* image (~ *ERS-1* configuration). When comparing **Fig. 10-a** with **Fig. 10-b**, the contrast between forested and non forested areas is found to be enhanced over the coherence image. The upper left part, including a collection of soil surfaces with different roughness conditions, illustrates drastically different behaviors between *C* band coherence and intensity information. The mean grey tone ranging from very dark to very bright over this sub-area in **Fig. 10-b** is stable around bright tones (i.e. high coherence) in **Fig. 10-a**. Moreover, the forest zone located at the middle of the image shows up more homogeneously in **Fig. 10-a** than in **Fig. 10-b**.

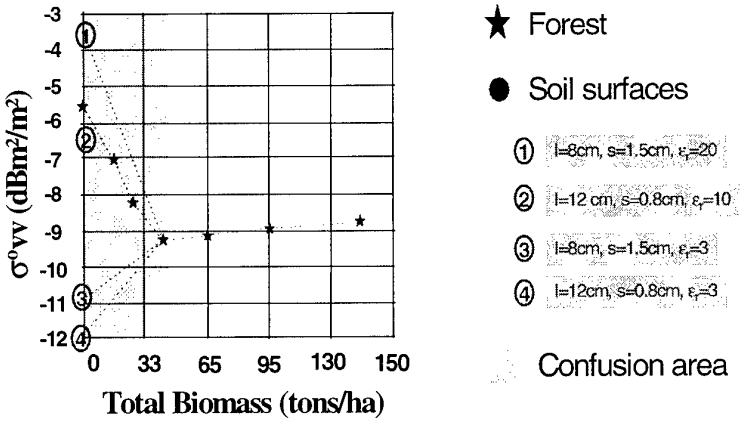


Fig. 9 : Simulated ERS-1 data (C-VV, 23°) for varying moisture and roughness conditions



Fig. 10-a : C band coherence map of the Landes test site. 24.5°

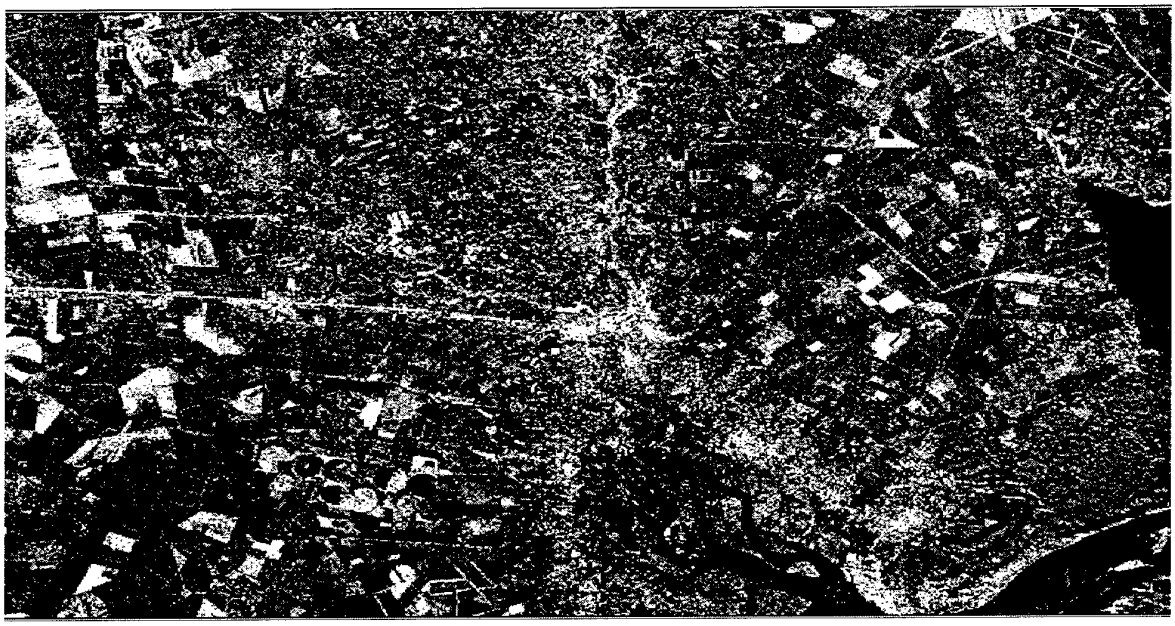


Fig. 10-b : C band VV map of the Landes test site. 24.5°

5.3 Use of polarisation synthesis

In the previous parts, L band HV configuration has been found to offer the highest sensitivity to forest age (i.e. biomass content) in the range [0 tons/ha, 100 tons/ha]. Besides, C band degree of coherence between HH and VV signals at low incidence angle (26.4°) led to the best results regarding forest/non forest segmentation.

Using polarisation synthesis techniques, the objective of this part is to demonstrate the consistency of the above-mentioned configurations regarding the choice of transmitting and receiving polarisations, but also to investigate the polarisation dependence of backscattering from forested and non forested areas. It consists in deriving the target response for any set of transmitting and receiving polarisation, from the knowledge of the polarimetric response derived in a linear polarisation basis $(\hat{h}, \hat{v})$. For every configuration, the degree of coherence between copolarised signals is also calculated.

5.3.1 Synthesis of intensity information

The knowledge of the full polarimetric information, included either in the 3*3 covariance matrix $[\overline{\overline{C}}]$ or in the 4*4 Muller matrix $[\overline{\overline{M}}]$, permits the application of polarisation synthesis techniques in order to

maximize/minimize the contrast between various types of targets. The simulated incident Stokes vector is given by : ${}^t\tilde{S}^i = (1, \cos 2\psi_i \cdot \cos 2\tau_i, \sin 2\psi_i \cdot \cos 2\tau_i, \sin 2\tau_i)$, where ψ_i and τ_i are respectively the orientation with respect to $\hat{h}$ direction, and the ellipticity of the polarisation ellipse ($\tau_i = 0°$ is a linear polarised wave, whereas $\tau_i = \pm 45°$ is a circular polarised wave) [19]. The free space backscattered energy is given by $\tilde{S}^{fs} = \left[\overline{M}\right].\tilde{S}^i$ where $\tilde{S}^{fs}$ is the free space backscattered Stokes vector. Considering a receiving polarisation $\tilde{S}^{rec}$, the synthesised intensity response is linked to the measured power at the terminal of the receiving antenna, propotional to : $P^{rec} = {}^t\tilde{S}^{rec} \cdot \left[\overline{M}\right] \cdot \tilde{S}^i$

The choice of optimal polarisations according to a given objective has been extensively discussed in the past. A great amount of work has been done either to determine the optimal polarisations of incoherently reflected waves [20], [21], or to determine the optimal transmitting and receiving polarisations for maximizing the polarimetric contrast between two targets [22]. Regarding the contrast maximisation issue, the basic principle is a two-step process : *1)* The incident polarisation $\tilde{S}^i$ is chosen in such a way that the ratio P^{fs1}/P^{fs2} is maximum, where P^{fs1} and P^{fs2} are respectively the free space backscattered power density coming from *Target 1* and *Target 2*. This step is subject to alternative methods, which are either to derive an optimisation based on total intensities, or to consider the ratio between the adjustable intensity parts only, namely the polarised parts of the backscattered waves. Regarding the issue of optimal polarisations for incoherently reflected waves, a numerical comparison between the two approaches [20]-[21] showed that the resulting mathematical and numerical problems turn out to be quite similar. *2)* Once $\tilde{S}^i$ is determined, the receiving polarisation $\tilde{S}^{rec}$ is adjusted so that P^{rec2} is minimum. The interest of this approach is a total decoupling between the choice of the transmitting wave polarisation and the one of the receiving wave. Another way of maximising the contrast is simultaneously to look for $\tilde{S}^i$ and $\tilde{S}^{rec}$ such as P^{rec1}/P^{rec2} is maximum. The two methods may lead to different results. Their comparison in the frame of remote sensing applications is in progress.

5.3.2 Synthesis of coherence information

In addition to intensity-based information (which are shown to be critically dependent on site and measurement conditions, including calibration uncertainties), polarisation synthesis can be further applied to coherence information. The generalized *pp-qq* degree of coherence [23] is defined as :

$$\rho_{pp-qq} = \frac{\langle S_{pp} \cdot S_{qq}^* \rangle}{\sqrt{\langle |S_{pp}|^2 \rangle \cdot \langle |S_{qq}|^2 \rangle}}$$

where $(\hat{p}, \hat{q})$ is a basis of orthogonal elliptical polarisations. The notion of *contrast optimisation* between *Target 1* and *Target 2* has been assessed using the degree of coherence. The selected criterion is a maximisation of the difference between $|\rho_1|$ and $|\rho_2|$ estimated over the two extended targets under consideration. The search for other types of criteria remains nevertheless a topic for further studies.

5.3.3 Applications

The contrast optimisation procedure has been applied to SIR-C/XSAR data. Over the Nezer test site, 4 different targets have been considered to assess the polarisation synthesis aspect. *Target 1* is a 46 years old forest (characterized by a biomass rate of 150 tons/ha). *Target 2* is a low vegetated area (clear-cut), *Target 3* is a smooth soil surface (r.m.s. height is 1.2 cm) and *Target 4* is a rough soil surface (r.m.s. height is 2.5 cm).

Forest / non forest segmentation

At C band and low incidence angle (24°), polarisation synthesis applied to the degree of coherence confirmed that $\rho\ hh - vv$ provided the best results for forest/non forest segmentation (the result of optimisation using *Target 1* and *Target 2* led actually to a basis of polarisations with ellipticity of 1°).

Forest stands mapping

Once the segmentation forest/non-forest is achieved using C band $\rho\ hh - vv$, the objective is to investigate the more appropriate set of transmitting and receiving polarisations to enhance intensity contrast (considering the adjustable intensity part only) between *Target 1* and *Target 2*, and consequently stretching out the dynamic between forest stands of different ages, in order to set up a classification of forested areas. Using 24°, L band data, the result of this optimisation is given by:

$\psi^i = 0.5°,\ \tau^i = -0.5°,\ \psi^{rec} = 90°,\ \tau^{rec} = -1°$

The optimal transmitting polarisation is a linear polarisation in the $\hat{h}$ direction, whereas the receiving polarisation is linear in the $\hat{v}$ direction. The result of the optimisation consequently confirms LHV as the most appropriate configuration for biomass estimation. Taking *Target 3* or *Target 4* instead of *Target 2* in the contrast optimisation process led to a similar result. Using this configuration, a biomass map in the range [0, 100 tons/ha] has been set up over the Nezer test site [24]. The image was first segmented into two classes (forest / non forest) using C band $\rho\ hh - vv$. The original SIR-C image is 12 looks, which has been further processed with a Kuan filter applied with a 9*9 moving window.

Roughness mapping

In order to derive a roughness mapping over non forested areas, as a preliminary step to a moisture inversion process, contrast optimisation has been subsequently applied to *Target 3* and *Target 4*, which have the same moisture content. In this case, the optimisation result led to :

$\psi^i = 43°, \tau^i = -37°, \psi^{rec} = 40°, \tau^{rec} = -30°$ $\tilde{S}^i$ and $\tilde{S}^{rec}$ are fully elliptical polarisations, close to the circular configuration. It should be noted that circular polarisations provides similar results. **Fig. 11a-b** display respectively the 24°, L band filtered image over a bare soils area for this synthesis, and the corresponding filtered LHV image. The different roughness states over the bare soil surfaces clearly show up in **Fig. 11-a**, whereas the intensity level is almost constant in the LHV configuration (**Fig. 11-b**). A similar result has been presented in [19], considering lava flows of different roughness. In a parallel study [23], the degree of coherence expressed in a circular polarisation basis has been chosen as a roughness estimator.

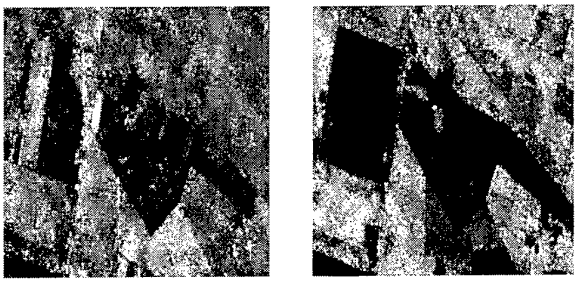


Fig. 11 : L band, 24° image over bare soils of varying roughness
(a) $\psi^i = 43°, \tau^i = -37°, \psi^{rec} = 40°, \tau^{rec} = -30°$
(b) $\psi^i = 0.5°, \tau^i = -0.5°, \psi^{rec} = 90°, \tau^{rec} = -1°$ (HV)

6. CONCLUSION

In this work, polarimetric properties of forest stands from the Landes Forest in South Western France have been investigated through the analysis of multi-incidence (*26.4°, 54°*) polarimetric data acquired by *SIR-C/XSAR* in April and October *1994*. Data analysis has been mainly based upon a theoretical modeling using radiative transfer theory, in conjunction with a branching model taking into account coherent scattering effects due to the clustered vegetation structures. The model has been further applied *1)* to the determination of major contributions in the backscattering mechanism for a given configuration of the incident wave (incidence, frequency, polarisation). *2)* to the estimation of wave attenuation through the forest canopy. Simulated results have been partly validated by experiment.

Besides the use of intensity information (i.e. backscattering coefficients σ^ohh, σ^ovv, σ^ohv) for clutter characterization, the information content of the polarimetric measurements has been assessed, conjointly with the use of polarisation synthesis techniques. For specific applications (e.g. the enhancement of roughness effect), the use of non linear polarisations has been shown to critically improve the sensitivity of radar signal. The work devoted to polarisation synthesis techniques is still in progress, especially for the application of man-made target detection and identification. To another extent, the information content of the vectorial aspect of the radar signal can be assessed through the use of interferometric data. A study based on ERS-1 repeat-pass interferometric data for forest environment studies is also in progress.

The technique developed over Les Landes test site is also tested over a collection of various forest types, especially tropical forest areas in Brazil (Manaus Cabaliana, Sena Madureira), where extensive deforestation activities are in progress.

7. REFERENCES

[1] Le Toan, T., A. Beaudoin, J. Riom and D. Guyon, « Relating forest biomass to SAR data », *IEEE Trans. on Geoscience and Remote Sensing, vol.30, n°.2*, March 1992.

[2] Beaudoin, A., T. Le Toan, S. Goze, E. Nezry, A. Lopes, E. Mougin, C.C. Hsu. H.C Han, J.A. Kong and R.T. Shin, « Retrieval of forest biomass from SAR data », *Int. J. of Remote Sensing, vol. 15, n°14, pp. 2777-2796*, 1994.

[3] Dobson, M.C., F.T. Ulaby, T. Le Toan, A. Beaudoin, E.S. Kasischke, N. Christensen, « Dependence of radar backscatter on coniferous forest biomass», *IEEE Trans. on Geoscience and Remote Sensing, vol.30, n°.2, pp. 412-415,* March 1992.

[4] Beaudoin, A., T. Le Toan and Q.H.J Gwyn, « SAR observations and modeling of the C band backscatter variability due to multiscale geometry and soil moisture », *IEEE Trans. on Geoscience and Remote Sensing, vol.28, no.5*, September 1990.

[5] Sarabandi, K., « Derivation of phase statistics of distributed targets from the Mueller matrix », *Radio Science, vol. 27, no. 5, pp 553-560*, September-october 1992.

[6] Yueh, S.H., J.A. Kong, J.K. Jao, R.T. Shin and T. Le Toan, « Branching model for vegetation », *IEEE Trans. on Geoscience and Remote Sensing, vol.30, no.2, pp. 390-402*, March 1992.

[7] Hsu, C.C., H.C. Han, R.T. Shin, J.A. Kong, A. Beaudoin and T. Le Toan, « Radiative transfer theory for polarimetric remote sensing of pine forest », *International Journal of Remote Sensing, Vol. 15, no. 14, pp. 2943-2954*, September 1994.

[8] Hsu, C.C., L. Wang, J.A. Kong, Souyris J.C., T. Le Toan « Theoretical modeling for microwave remote sensing of forest », *Proceedings IEEE Workshop « Retrieval of bio- and geophysical parameters from SAR data for land applications »*, Toulouse, 10-13 October 1995.

[9] Souyris J.C., Le Toan T., Beaudoin A, Hsu C. C., Kong J.A., "On the use of extended targets for calibration : results from Maestro-1 data and Mac-Europe data", *Proceedings of CEOS calibration/validation working group, ESTEC, Noordwijk, The Netherlands,* September 1993.

[10] Hallikainen M.T., F.T. Ulaby, M.C. Dobson, M.A. EL-Rayes and L.K. Wu, "Microwave dielectric behaviour of wet soil. Part I : Empirical models and experimental observations", *IEEE Transactions on Geoscience and Remote Sensing, Vol. GE-23, no.1,* January 1985.

[11] Van de Hulst, H. C., "Light Scattering by Small Particles", Wiley, New York, 1957.

[12] Karam, M.A., A.K.Fung, and Y. M. M.Antar, "Electromagnetic scattering from a layer of finite length, randomly orientated, dielectric, circular cylinders over a

rough interface with application to vegetation", *International Journal of Remote Sensing, Vol. 9, No. 6, 1109-1134,* 1988.

[13] Ulaby, F. T., M.A.El-Rayes, `"Microwave dielectric spectrum of vegetation - part II: dual - dispersion model", *IEEE Trans.Geosci Remote Sensing, Vol. GE-25, No. 5, 550--557,* 1987.

[14] Tsang, L., J.A.Kong, and R. T. Shin, *Theory of Microwave Remote Sensing, Wiley-Interscience, New York,* 1985.

[15] Fung A.K., Z. Li, K.S. Chen, "Backscattering from randomly rough dielectric surface", *IEEE Transactions on Geoscience and Remote Sensing, Vol 30, n°2, 356-369,* 1992.

[16] Kong, J.A., A.A. Swartz, H.A. Yueh, L.M. Novak and R.T. Shin, "Identification of terrain cover using the optimum polarimetric classifier", *Journal of Electromagnetic Waves and Applications, 2, 2, 171-194,* 1988.

[17] Souyris J.C., Le Toan T., Hsu C.C., Kong J.A., « Assessment of SIR-C/XSAR polarimetric data for the estimation of forest parameters », *Conference proceedings Third international workshop on radar polarimetry, pp. 636-645, Nantes (France),* March 1995.

[18] Souyris J.C., Le Toan T., Hsu C.C., Kong J.A., « Inversion of Landes Forest biomass using SIR-C/XSAR data », *Proceedings International Symposium « Retrieval of bio- and geophysical parameters from SAR data for land applications »,* Toulouse, France, October 17-20, 1995.

[19] : Zebker, H.A., J.J. Van Zyl, D.N. Held, « Imaging radar polarimetry from wave synthesis », *Journal of Geophysical Research, vol. 92, no. B1, pp. 683-701,* January 1987.

[20] : Van Zyl, J.J., C.H. Papas, C. Elachi, « On the optimal polarisations of incoherently reflected waves », *IEEE Transations on Antennas and Propagation, vol. AP-35, No. 7,* July 1987.

[21] : Kostinski, A.B., B.D. James, W.M. Boerner, « Optimal Reception of Partially Polarized Waves », *J. Opt. Soc. Am. A, Vol. 5, No. 1,* January 1988.

[22] : Kostinski, A. B., W.M. Boerner, « On the polarimetric contrast optimisation », *IEEE Transations on Antennas and Propagation, Vol. AP-35, No. 8,* August 1987.

[23] : De Carolis, G., F. Mattia, G. Pasquariello, F. Posa, P. Smacchia, N. Floury, T. Le Toan, J.C. Souyris, « Soil moisture estimation from multipolarisation and multifrequency SAR data », *Proceedings IEEE Workshop « Retrieval of bio- and geophysical parameters from SAR data for land applications »,* Toulouse, 10-13 October 1995.

[24] : Souyris J.C., T. Le Toan, N. Floury, L. Thomasson, C.C. Hsu, J.A. Kong, « Use of polarisation synthesis for deforestation studies based on SIR-C/XSAR data analysis », *Proceedings IGARSS' 96,* Lincoln, Nebraska, USA, 27-31 May 1996.

9. ACKNOWLEDGMENTS

This work was supported by the Direction des Recherches et Etudes Techniques (DRET) under contract n° 94-091, the Centre National d'Etudes Spatiales (CNES) and NASA contract n° 958461. The authors thank Jerôme Bruniquel for image filtering work.

Paper 20

E. Krogager (DE)

Since the choice of basis is rather arbitrary, the HH, HV, UV elements of the linear basis represents a rather arbitrary choice of features. One of the great advantages of utilizing fully polarimetric data is that the data can easily be transformed to any other general basis, e.g. a circular basis, or decomposition methods can be applied in order to extract various features. Have you tried or considered using such other parameters for your analyses?

Autoadaptive Scene Identification in Multi-Spectral Satellite Data

P. Boekaerts, E. Nyssen, J. Cornelis

Department of Electronics (ETRO), Research Group IRIS
Vrije Universiteit Brussel (VUB)
Pleinlaan, 2 - 1050 Brussels, Belgium
e-mail: pkboekae@etro1.vub.ac.be

SUMMARY.

A non-supervised scene identification scheme for multi-spectral satellite data is presented. The identification of a scene is equivalent to the assignment of meaningful labels to image regions belonging to the scene. Multi-spectral scene region detection is reduced to the problem of finding an algorithm that performs a data reduction on composite images while (optimally) preserving scene region information. A self-organising one-dimensional feature map applied to random segments of individual METEOSAT channels has already been shown to meet the requirements of such algorithm for mono-spectral cloud identification. The use of two- and three dimensional feature maps cannot be generalised for bi- and tri-spectral scene analysis but it is shown that conventional clustering algorithms can be used for that purpose. The latter is illustrated for tri-spectral scene analysis in METEOSAT composite images of the visible, infrared and water vapour channels. A study of the segment size indicates that small segment sizes are sufficient and even better than large segment sizes for multi-spectral scene detection. This is explained in terms of the distribution properties of the so called segment space and the structural features of the code vectors (code segments). Decreasing the number of code segments used to reduce the information content of METEOSAT composite images results in a systematic loss of scene information. The proposed method for multi-spectral scene identification can be applied to data sets obtained after the fusion of spectral information of different sensors with different resolution and opens interesting perspectives for automated object tracking in temporal sequences of satellite data.

1. INTRODUCTION.

Remote sensing plays a crucial role in the observation of the Earth, e.g. for the triggering and modelling of physical processes in the atmosphere and on the Earth surface (land, ocean and lakes). The accuracy of the determination of local phenomena or structures in the atmosphere or on the Earth surface depends both on the resolution of the observations and on the number of spectral bands in which the radiances are measured. The demand for higher level scene identification facilities in remote sensing data resulted in the fact that an increasing number of satellites (e.g. SPOT, LANDSAT, METEOSAT, ...) are transmitting an increasing amount of spectral information observed at the top of the atmosphere with an ever increasing resolution. While the current METEOSAT satellite series collect and transmit for example daily +- 1.5 Giga Byte (including three spectral channels every half hour) to observe the climatological condition of the atmosphere, the METEOSAT Second Generation satellites (MSG) will collect daily in the next decade up to 19 Giga Byte (including 11 spectral channels every 15 minutes). This evolution has important consequences for the technology used to transmit and archive remote sensing data, as well as for the speed and the degree of automation with which these data must be processed (in real time).

The generation of secondary data products from remote sensing data covers the identification of scenes in many applications, e.g. the identification of cloud/surface types in narrow band METEOSAT data for the derivation of wind vectors or in broad band ERBE or ScaRaB data for the derivation of Earth Radiation Budget measurements, the characterisation of crop coverage in SPOT data for vegetation studies ...

This paper presents an autoadaptive multi-spectral scene identification scheme for the analysis of remote sensing data in general. The proposed method can be applied to a variable number of spectral observations with a variable resolution and to radiation measurements of different sources. Some results on tri-spectral land surface/cloud identification in METEOSAT data are discussed and it is shown that the method can be used for the triggering of scenes in temporal sequences of data.

Paper presented at the AGARD SPP Symposium on "Remote Sensing: A Valuable Source of Information", held in Toulouse, France from 22-25 April 1996, and published in CP-582.

2. MULTI-SPECTRAL SCENE IDENTIFICATION.

2.1. Multi-spectral segment analysis.

The spectral sensitivity of the different sensors of a radiometer on board of satellites is chosen in such a way that the scenes of interest observed by the radiometer are characterised by typical multi-spectral distributions. The spectral distributions of different scenes are, in reality, always overlapping however. This implies that a single multi-spectral measurement (on the level of one pixel corresponding to a given geographical location or footprint on the earth) is not sufficient to identify a given scene. Individual multi-spectral observations may be shared by different scenes types. This is the reason why multi-spectral histogram analysis (e.g. multi-spectral Gaussian curve fitting in tri-spectral METEOSAT data or Maximum Likelihood Estimation in bi-spectral broad band ERBE and ScaRaB data) deals with a scene separability problem. If we accept that a scene in multi-spectral radiation data can be defined as a spatially coherent region with a characteristic (set of) local spatial distribution(s) of multi-spectral measurements, then we must accept that the scene separability problem can be solved by analysing the spatial properties of these multi-spectral measurements. In the case of tri-spectral measurements (with the same resolution), a scene will appear in the corresponding tri-spectral composite image as a region with typical local color distributions (each color in a tri-spectral composite image represents a unique tri-spectral measurement). The *local* nature of these color distributions implies that we don't need to analyse the whole scene in order to be able to decide to which scene type a given multi-spectral measurement (or color pixel) belongs. We assume that the spatial color distribution observed in the local neighbourhood of a pixel of the composite image is sufficient to make that decision. Such neighbourhood is defined by a rectangular sub-region in the (multi-spectral) composite image, characterised by a given size (dimension) and a given location in the image, and is referred to in this paper as a (multi-spectral) segment. The corresponding approach to scene analysis is called (multi-spectral) *segment analysis*. A tri-spectral segment with dimension (S1*S2) contains S= S1*S2 color variables (or 3xS grey variables) and can therefore be represented as a point in a S-dimensional discrete color space (or a 3xS dimensional grey value space), the *segment space*. The segment space contains more generally the multi-spectral information content of all possible segments present in multi-spectral image data if we move a segment with constant size over the image. This can be done in two ways: by moving the segment pixel by pixel or block by block. In the latter case, we do not consider overlapping segments and the covered segment distribution is only a sub-distribution of the one obtained after moving the segment pixel by pixel. The pixel or block segment distribution of a multi-spectral image can be regarded as a point distribution in the segment space. Position information of the image segments however is lost and the construction of the segment space may, for that reason, be considered as a lossy image data transformation in which meso-scale information of scenes is discarded (the image cannot be reconstructed from the point distribution). This property of the segment space is compatible with the basic assumption that scenes can be identified within one multi-spectral segment, independently of the position of the segment in the image. If the assumption is true that scenes are characterised by local multi-spectral distributions, the information content of the segment space must be sufficient to perform multi-spectral scene analysis, and a given image region corresponding to a typical scene type must appear as a (set of) local cluster(s) in the point distribution. Multi-spectral scene analysis in the segment space is therefore reduced to the analysis of point distributions and based on the assumption that local point distributions (clusters) in the segment space correspond to typical scene types in the original multi-spectral data. The segment size will not only affect the dimension of the segment space, but also the separability of the clusters corresponding to typical scene types. The segment size is for this reason an important experimental variable to be optimised for any scene identification scheme in the segment space.

2.2. Multi-spectral scene identification task.

The identification of a given scene in multi-spectral data consists of the localisation of (the) image region(s) belonging to the scene and the labelling of the(se) regions. If we accept the definition of a scene given in section 2.1., then the task of multi-spectral scene detection consists of assigning one (or a set of unique) multi-spectral value(s) to spatially coherent regions with similar local spectral distribution(s). This is equivalent to assigning one (or a set of unique) value(s) to the segments covering a given scene and hence, if we accept the properties of the segment space discussed in section 2.1., to a data reduction or coding of local point distributions in the segment space.
Once we have detected the image regions belonging to a given scene, then the task of scene labelling consists of an image transformation in which we re-define the unique value(s) assigned to different scene types. The labelling of scenes may be a complex transformation based on different types of information, e.g. the application of scene radiation models.

2.3. Multi-spectral segment coding.

The (optimal) representation (or coding) of local point distributions in the segment space covers the problem of vector quantisation used for data compression and encoded data transmission and deals with the question how to find a set of prototypes (code segments) that

optimally matches a given (unknown) segment distribution.

Consider an arbitrary S-dimensional vector distribution V with a given probability density function:

$$\mathbf{P}(\mathbf{s}) \mid \mathbf{s} = \mathbf{s}[i,j]$$

The indices i, j are used to treat the corresponding S-dimensional vector of a given multi-spectral image segment as a two-dimensional data structure.
The problem of encoding may be defined as a mapping of the segment distribution on a set of S-dimensional code vectors (code segments):

$$V(\mathbf{P}(\mathbf{s})) \rightarrow W = \{\mathbf{c}_{r1}, \mathbf{c}_{r2} \ldots \mathbf{c}_{rn}\}$$

This transformation introduces an index mapping function $F(\mathbf{s})$ that maps a given segment on the index of the corresponding code segment:

$$F(\mathbf{s}) = r_i$$

With r_i being the index of the centroid of a subset of V for which $\mathbf{c}_{ri}$ is coding.

An approximation for such mapping can be found by minimising the expectation value of the squared error between the segments of the whole segment distribution and the corresponding code segments $\mathbf{c}_{F(s)}$:

$$\left[\mathbf{E}_W = \int |\mathbf{s} - \mathbf{c}_{F(s)}|^2 P(\mathbf{s}) d\mathbf{s}\right]_{\min} \qquad (1)$$

Different types of solutions for the corresponding discrete minimisation problem can be found in literature. Methods based on a local gradient descent technique (applied to a random sampling of the segments from the distribution) include discrete competitive adaptation (or "learning") known as the "best match principle" or "δ-adaptation" and the extended version of self-organising topological feature maps (Kohonen maps). The final solution of these algorithms, i.e. the set of code segments that represents a global minimum of (1), is a data compression tool that can be used for encoded data transmission. Any segment of the distribution can be represented by a simple index of the corresponding code segment:

$$\mathbf{s} \rightarrow r_i \rightarrow \mathbf{c}_{ri}$$

Self-organising one-dimensional topological feature maps has been shown to be an appropriate tool for mono-spectral scene identification [6], but the use of multi-dimensional topological feature maps for multi-spectral scene identification could not be generalised due to practical limitations (i.e. unreasonable computation requirements). [5, 6]

More conventional vector quantisation algorithms, e.g. K-means clustering, suffer from the same constraint as δ-adaptation [5], i.e. the code segments are not interrelated and code segment indices can therefore not be used as consistent scene identification labels (the code segment indices represent random labels in both cases). It is shown in the next paragraph however that other properties of the code segments of these encoding algorithms can be used for the purpose of classification with a minimal loss of scene identification performance.

2.4. K-means clustering and multi-spectral scene classification.

In δ-adaptation, the best matching code segment is updated after each random selection of a segment from the distribution with a (decreasing) factor ε towards the selected segment, i.e.

$$\Delta \mathbf{c}_r \mid_{\varepsilon \rightarrow 0} = \varepsilon\, \delta_{r,F(s)} (\mathbf{s} - \mathbf{c}_r)$$

$$\text{with} \quad \delta_{r,F(s)} = 1 \quad \text{for} \quad F(s) = r$$
$$\delta_{r,F(s)} = 0 \quad \text{for} \quad F(s) \neq r$$

K-means clustering requires, in contrast to δ-adaptation, the a priori knowledge of the whole segment distribution (which is the case in most remote sensing applications) and all the code segments are updated at once after a sequential scan of all the segments belonging to the distribution by replacing them with the mean segment values of the segments for which the code segments fulfilled the best matching principle:

$$\mathbf{c}_r = \frac{1}{\sum \delta_{r,F(s)}} \sum \delta_{r,F(s)} x S \qquad (2)$$

K-means reduces for this reason the computational complexity of the adaptation procedure used in δ-adaptation. It can be shown that in the case of the choice of an Euclidean distance measure as a matching criterion between two segments, i.e.

$$D(S1, S2) = \sqrt{\sum_{i,j} (S1_{[i,j]} - S2_{[i,j]})^2}$$

the iterative repetition of the adaptation procedures (2) results in the minimisation of the squared distance error between the whole segment distribution and the code segment set.

In section 2.1., it was explained that we consider a multi-spectral segment with odd segment size in order to be able to decide to which scene the central pixel of the segment belongs. In equation (2), we can observe that during the last adaptation, the central pixel of each code segment is replaced by the mean multi-spectral measurement of the central pixels of all the segments for which the code segment is coding. This means that the central pixel of a given code segment can be used, after convergence of the adaptation procedure, as a (consistent) scene classification label representing the mean multi-spectral value of the scene for which the code segment is coding. If the central pixel of each segment of a given multi-spectral image is replaced by the central pixel of the corresponding code segment obtained after autoadaptation, a high resolution classified image is obtained in which each (or a set of) multi-spectral value(s) represent(s) a given scene. The resulting scene identification procedure quarantees consistent scene labels but does not quarantee that the number of scene labels equals the total amount of code segments used to analyse the data. Different code segments may indeed have the same central pixel value (different scenes may have the same mean multi-spectral value). The reduction of the computational cost for multi-spectral scene analysis by using K-means clustering results for this reason in a (small) degradation of the scene identification performance compared to the use of multi-dimensional topological feature maps.

3. RESULTS.

3.1. METEOSAT tri-spectral image data.

The proposed method for multi-spectral scene identification is illustrated for tri-spectral cloud and surface analysis in METEOSAT radiometric data. The European geostationary METEOSAT satellites observe the earth with an imaging radiometer in three spectral channels: the solar spectrum (VIS) between 0.4 and 1.1 μm (with a resolution of 2.5 km x 2.5 km; one VIS image contains 5000x5000 pixels of 8 bit/pixel), in the infrared window region (IR) between 10.5 and 12.5 μm and in the water vapour (WV) absorption band between 5.7 and 7.1 μm (both with a resolution of 5 km x 5 km, 2500x2500 pixels image data, 8 bit/pixel).[1] The VIS channel measures reflected solar radiation from clouds and the earth's surface (e.g. land, sea), the IR channel measures thermal radiation in an atmospheric window region emitted from clouds and surface and the WV channel observes the earth's atmosphere in a strong absorption band and is sensitive to the atmospheric column water vapour in the upper troposphere above 600 hPa.[2] The raw METEOSAT images are geometrically rectified in real time at EUMETSAT, Darmstadt. These rectified images are used as the basic data input for multi-spectral scene identification after bit-wise inversion of the IR and WV channels (to obtain a consistent scene appearance in the three channels). Although the proposed autoadaptation procedure discussed in section 2.4 allows the use of multi-spectral data of different resolution and from different sources, we reduced the resolution of the VIS channel by a factor two (giving the same resolution as the IR and WV channel) in order to be able to represent tri-spectral METEOSAT data as RGB composite images. The corresponding composite images and scene identification results are represented as grey images in the present publication.

3.2. Study of the segment size.

The result of autoadaptive, tri-spectral cloud and Earth surface detection in a tri-spectral METEOSAT window composed of the VIS, IR and WV channels (slot 17 of 94-03-7, see Fig. 1a) with a code set of 40 code segments is illustrated in figure 1c for a segment size of 3x3 pixels (S=3) and in Fig. 1d for a segment size of 5x5 pixels (S=5). The shown METEOSAT window (covering the South-West part of Madagascar) represents only a fraction of the region to which the code sets were adapted and was selected in order to show the local consistency of the scene identification results. By inter comparing Fig. 1c. and Fig. 1d., we can draw the conclusion that an increase of the segment size results in an enhanced spatial coherence of the detected cloud and surface regions. Fig. 1b represents the results obtained after a direct coding of the RGB histogram (N=40 code colors), which is equivalent to the limit situation of scene identification with a segment size of 1 pixel. In Fig. 1b., cloud and surface regions are fragmented and are composed of different scene types. Fig.1 (a, b, c, d) illustrate also that the segment size must be chosen in function of the resolution of the data. Very fine cloud structures observed in the left upper corner of Fig. 1a. are vanishing systematically when the segment size is increased. This can be explained in terms of the statistical properties of the segment space and the internal structure of the code segments. An analysis of the internal structure of adapted code segments reveals that we can divide them in two different types: code segments containing sharp color transitions, coding for edge information in the image, and nearly homogeneous code segments, coding for the color distributions in the bulk of cloud and surface regions. Fine cloud and surface regions with a similar dimension of that of the code segments will therefore appear as edge information and will be represented as scattered local point distributions in the segment space. This implies that a large amount of code segments are needed to code for these scenes and for edge information in general (see also next paragraph). As a conclusion we can state that an increase of the segment size used for multi-spectral

scene identification results in a higher separability of scenes and therefore in a higher accuracy of scene detection if, and only if the number of code segments are increased sufficiently.

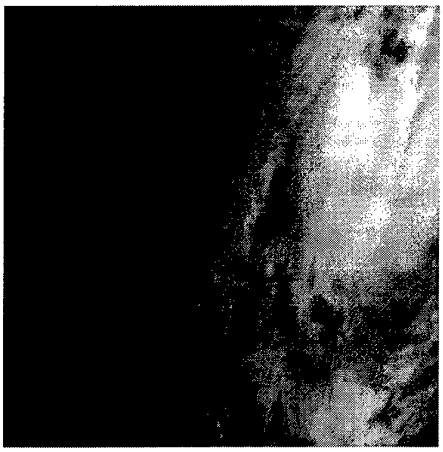


Fig.1a Date 94-03-07, slot 17

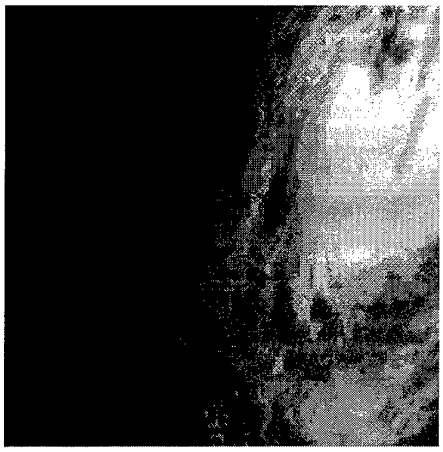


Fig.1b Date 94-03-07, slot 17: N=40, S=1

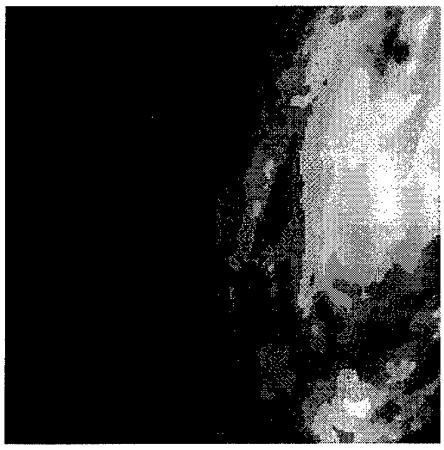


Fig.1c Date 94-03-07, slot 17: N=40, S=3

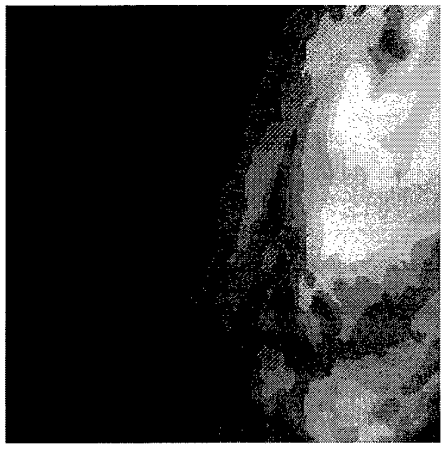


Fig.1d Date 94-03-07, slot 17: N=40, S=5

3.3. Study of the number of code segments.

Fig's. 2(a, b, c) illustrate the effect of the number of code segments on autoadaptive cloud and surface detection in the same tri-spectral METEOSAT window of section 3.2. Decreasing the number of code segments used to represent the segment distribution of the tri-spectral data results in a systematic loss of scene information. The number of code segments influences directly the density of code segments assigned to typical local point distributions in the segment space. If the segment distributions of two different cloud types are well separated in the segment space, then a limited number of code segments is sufficient to represent these distribution regions. To obtain consistent cloud region detection for scene segment distributions that are very close to each other (while belonging to different scene types) or that are scattered in the segment space (while belonging to one scene type), a sufficient number of code segments must be assigned. By increasing the number of code segments, we also increase the number of code segments assigned to well separated segment distributions belonging to one scene type however, resulting in a fragmentation of these regions with

(unique) sub labels. This situation complicates the final labelling of regions (see section 2.2.) but may not be confused with the erroneous fragmentation observed in Fig.1b (S=1).



Fig.2a Date 94-03-07, slot 17

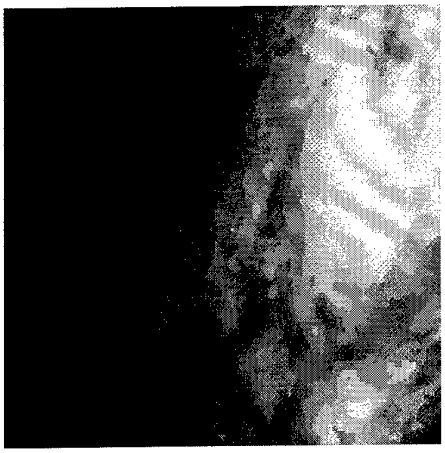


Fig.2b Date 94-03-07, slot 17: N=40, S=3



Fig.2c Date 94-03-07, slot 17: N=10, S=3

3.4. Multi-spectral scene analysis in temporal sequences of image data.

The characteristics of the local spectral distributions of scenes in multi-spectral image data may vary in temporal sequences if the images are not acquired under the same environmental conditions and geometrical setup of the sensors (viewing angle).
The stability of code segments, i.e. the degree to which code segments, adapted to a given data set acquired at a given time, may be used for scene identification in a data set acquired at an other time, will depend on the stability of the local spectral distributions of the scenes observed in the time sequences. METEOSAT satellites are geostationary satellites, so the viewing angle of cloud and earth surface regions varies only from the Earth position at nadir to the edges of the Earth disk and does not vary between slots taken at different time intervals. The position of the sun is an important environmental variable in time sequences of METEOSAT slots however. The position of the sun varies slowly however within the half hour sample frequency of METEOSAT data during day and varies slowly during subsequent days at a given time of the day (due to seasonal variations). The stability of code

segments (N=40, S=3) has been verified for these conditions and is illustrated in Fig's. 3b and 4b.
Fig. 3b represents the scene identification obtained in METEOSAT slot 19 of 94-03-07 with a code segment set adapted to slot 17 of the same day (see Fig's. 1a or 2a). Although both slots are acquired with a time interval of one hour, cloud and Earth surface regions are properly detected in slot 19. A similar observation could be made for a time delay of 25 hours (1 day + 1 hour), see Fig.'s 4a and 4b.

Note that stable code segments allow the tracking of cloud formations in temporal sequences of METEOSAT slots, e.g. the tracking of the centre of the hurricane Kelvina displayed in the right part of Fig's. 2a, 3a, 4a.

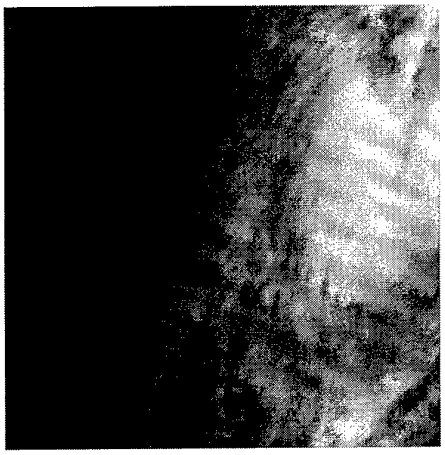


Fig.3a Date 94-03-07, slot 19



Fig.3b Code set 94-03-07, slot 17: N=40, S=3

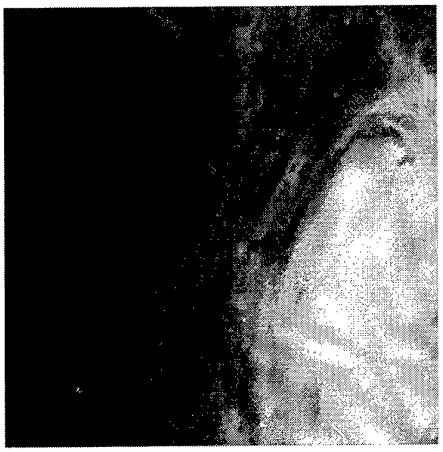


Fig.4a Date 94-03-08, slot 19



Fig.4b Code set 94-03-07, slot 17: N=40, S=3

4. CONCLUSION.

The present study has shown that a conventional vector quantisation algorithm, i.e. K-means, applied to the multi-spectral segment distribution of remote sensing data, e.g. tri-spectral METEOSAT data, can be used for consistent multi-spectral scene identification. The proposed method can be applied to multi-spectral data with different resolution and from different sensors. The statistical properties of local point distributions in the segment space can be used to explain the experimental evidence that both the segment size and the number of code segments control the accuracy with which different types of scenes can be identified. If the local spectral distribution(s) of the multi-spectral measurements within a given scene are stable enough in time, then the

presented method allows automated object tracking in temporal sequences of multi-spectral image data.

5. ACKNOWLEDGEMENTS.

This research was funded by the Federal Office for Scientific, Technical and Cultural Affairs (DWTC/SSTC) of the Science Policy Office of Belgium, as part of the Telsat III Program (Program Teledetection, contract reference T3/03/41).

The author thanks Mr. J. Leber from the European Space Operations Centre in Darmstadt for the METEOSAT data and Mr D. Crommelynck from the Royal Meteorological Institute of Belgium for his support.

6. REFERENCES.

1. J. Schmetz, K. Holmlund, J. Hoffman, B. Strauss, B.Mason, V. Gaertner, V., A. Koch, L. van de Berg, "Operational Cloud-Motion Winds from Meteosat Infrared Images", J. Appl. Meteorol., vol. 32, No. 7, 1993.

2. J. Schmetz, L. van de Berg, "Upper Tropospheric humidity observations from Meteosat compared with short-term forecast fields", Geophysical Research Letters, vol. 21, No. 7, pp. 573-576., 1994.

3. Meteosat Exploitation Project , MIEC Processing, ESA STR-224, 1987.

4. P. Boekaerts, E. Steenput, M. Acheroy, P. Van ham, J. Cornelis, "Cloud detection in Multi-Spectral NOAA-AVHRR data using Multi-Layer Perceptrons: a pilot study", Space Scientific Research in Belgium, Volume III, Earth Observation, Part 1, p.173-179, Federal Office for Scientific, Technical and Cultural Affairs, Editor DWTC, 1995.

5. P. Boekaerts, "Autoadaptive cloud identification in Meteosat images", ESA SP-1183, ISBN 92-9092-331-8, September 1995.

6. P. Boekaerts, E. Nyssen, J. Cornelis, "Autoadaptive mono-spectral cloud identification in METEOSAT satellite images", European Symposium on Satellite Remote Sensing, Conference on Image and Signal Processing for Remote Sensing II, EOS/SPIE Vol. 2579, pp 259-271, November 1995.

Modeling and Wave-Oriented Processing of Scattering Data from Complex Environments

LOCATION AND IDENTIFICATION OF BURIED OBJECTS BY GROUND PENETRATING RADAR

Leopold B. Felsen
Department of Aerospace and Mechanical Engineering and
Department of Electrical, Computer and Systems Engineering
Boston University
110 Cummington Street
Boston, MA 02215
Also: Polytechnic University (Emeritus), Brooklyn, NY 11201 USA

Abstract
Inverse algorithms for processing data due to wave scattering from complex environments can be strengthened by incorporating relevant robust wave physics in the inversion scheme. Discriminants (footprints) for object classification and identification are best documented in the configuration (space-time)-spectrum (wavenumber-frequency) phase space. Illustrations here emphasize ground penetrating radar (GPR) applications. Because of the complexity of this target-background environment, the processing strategy generally relies heavily on numerical, model-based and statistical techniques. A model is proposed here which utilizes wave-based analytic techniques to reduce the "size" of the overall problem. The model is structured around high resolution pulsed-beam propagators for transporting the incident signal to and from the target through lossy dispersive soil environments. Attention is given to those portions of an overall problem which are good candidates for wave-based analytic methods and those portions which are better served by other methods.

I. Background Perspectives

The detection, classification and identification of hidden objects is a subject area which is steadily gaining importance for military as well as civilian applications. Here, "hidden" is taken as a generic designation that includes objects buried underground, embedded in concrete or other materials, hidden in foliage, etc., as well as objects submerged in water channels. Electromagnetic waves are used for interrogation, with emphasis on buried objects in the presence of background clutter. However, the proposed techniques could also be applied to acoustic coupling to buried objects, with the possibility of inducing elastic responses (e.g., resonances) in certain materials.

In the development and testing of inverse algorithms for processing data generated by scattering from a hidden target, it is becoming recognized that to improve accuracy, speed, and robustness, the underlying wave physics should be built into the inversion schemes; by thus enlarging the prior-knowledge base, the dimensionality of the subsequent processing is reduced. Parametric inversion of data implies extraction of characteristic parameters (footprints) which are discriminants for object classification and identification, and they are usually established by beginning with the forward problem for various model environments. Wave-generated footprints are best catalogued in the configuration (space-time) - spectrum (wavenumber-frequency) phase space because wave phenomena and their local features are most completely characterized by their configurational as well as spectral attributes; both are given equal weight in the phase space setting. While an awareness of the time-frequency phase space subdomain has begun to be felt within the electromagnetics and acoustics wave scattering communities because of the strong emphasis placed on it during the past decade by signal processors [1-10], the studies by Carin and Felsen [9, 11, 12] were among the first to apply phase space concepts also to the space-wavenumber domain. From these phase-space investigations, some important insights were gained concerning robustness in the presence of additive system noise and deterministic (target) - stochastic (clutter) interaction. Thus before choosing a signal-processing strategy for a particular scattering scenario, it is advisable to perform a complete parametrization of the underlying wave physics, thereby facilitating choices as to whether a particular parameter should be treated as deterministic or random.

The studies initiated by Carin and Felsen (see [9] and the references cited therein) have been continued, with Carin and his research group focusing primarily on the development of numerical modeling and model-based signal processing algorithms, and Felsen concentrating primarily on analytic techniques for classifying basic wave propagation and scattering issues. The latter are addressed here within the context of ground penetrating radar (GPR).

II. Modeling Strategy

A. Structure

To detect and classify buried targets by radar the modeling strategy seeks to parametrize how various soil background environments modify the field incident on the target, and how these environments modify the field scattered from the target before it reaches the observer. If the effect of the background environment can be accounted for by a "good" propagation model, the target interrogation by the background-modified signal can proceed in the local environment near the target. The decomposition into the background domain and the target domain highlights the individual footprints of these distinct scattering phenomenologies, which must then be combined appropriately to parametrize the target-background interaction.

Paper presented at the AGARD SPP Symposium on "Remote Sensing: A Valuable Source of Information", held in Toulouse, France from 22-25 April 1996, and published in CP-582.

In essence, the propagation algorithm projects the signal from an initial surface S_I near the transmitter onto a surface S_T near the target (Fig. 1). The target domain encompasses the interior of S_T, and the scattering from the target in its local environment can be addressed analytically, when warranted, or purely numerically; in the latter case, the size of the numerical problem is much reduced over what it would have been if the *entire* problem, starting from S_I, were treated numerically. In this context, the analytic propagators (for which we choose beams in the frequency domain and pulsed beam wavepackets in the time domain, see Sec. IIE) could be regarded as long-range elements and might be embedded accordingly in numerical discretization codes.

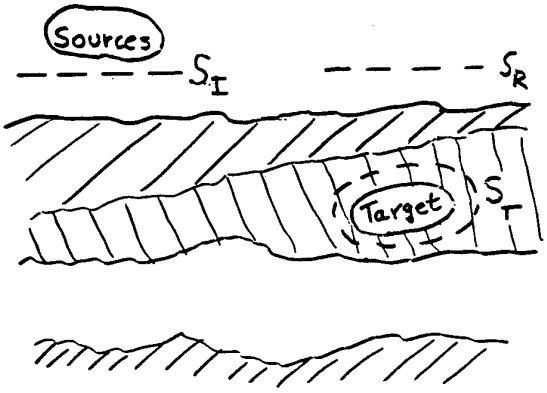


Fig. 1 - Physical configuration of lossy, dispersive layers with a target. S_I: Initial field aperture; S_T : Target aperture; S_R : Received field aperture.

More specifically, the modeling strategy advocated here is based on the following constructs:

1. Represent the incident field due to a given source "aperture" distribution as a superposition of beam basis fields. *Rigorous algorithms to this effect are available.*

2. Propagate these basis beams individually through the lossy dispersive soil environment into the vicinity of the target. *Establishing the rules of propagation by asymptotic and related techniques is partly known and partly new.*

Because the beams act like well-collimated local probes, they can be propagated similar to ray fields through heterogeneous media provided that the scales of variation in such media and on their boundaries are large compared to the beam width. However, unlike ray fields, beam fields do not fail near shadow boundaries, caustics, and other ray-optical transition regions.

3. Synthesize the resulting total field near the target by beam superposition.
This process projects the initial aperture distribution onto an "aperture" near the target, with full account taken of the distortion of the signal by the soil environment. The aperture distribution near the target furnishes the ambient incident field for target interrogation and classification; note that this aperture surface need not be planar but can be chosen to suit the target interrogation scheme.

4. Extract the target footprints.
The previous steps have accounted for the distortion of the original signal by the soil background, and therefore furnish the modified signal which is incident on the target from the equivalent aperture distribution near the target. The scattering from the target in its "local environment" represents one of the building blocks in the overall scheme. The local environment depends on the problem conditions. For example, it might incorporate the dominant interaction mechanisms of primary target-scattered fields with environmental features (like nearby interfaces) that induce secondary scattering. Note that for target classification and identification (ID), the classifier need not necessarily be the *shape* image of the target. Target "footprints" in the phase space, such as resonances, are frequently more effective than information based on shape. Because the problem scale for target ID has been much reduced by projecting from the initial data plane onto the aperture near the target, phase space modeling and footprinting of the target scattering process by direct numerical methods may be an attractive option as noted earlier.

5. Construct the scattered fields at the receiver.
Here, the previous steps are followed in reverse order: the scattered field on the aperture near the target is decomposed into beams, the beams are backpropagated through the soil environment (the conventional designation of "backpropagation" is actually forward propagation in the reverse direction), and they are recombined at the receiver.
6. Calibration of the beam propagator algorithms.
The asymptotic beam propagator solutions need to be calibrated for accuracy. This is to be done by comparison with numerically generated reference data for selected model environments. We have previously pursued this strategy for various simple canonical scattering configurations ([9], etc.) which are cited in the list of references.

B. Implementation

As in all parametrization schemes, the implementation depends on the choice of the "parameters." This choice is generally problem dependent. Thus, in multilayered environments, wave objects that emphasize *individual* interface effects sample the layering via a succession of multiply reflected *progressing* waves or wavefronts, whereas wave objects that emphasize the *collective* effect of several interfaces sample the entire configuration via a series of *oscillatory* wavefields (modes or resonances). The *progressing* (ray or beam-type) wavefields furnish a *local* sensor of the layering whereas the *oscillatory* (mode or resonance) wavefields furnish a *global* sensor of the composite, which can, however, be synthesized formally by summing over all of the ray field multiples. The most versatile combination encompasses a self-consistent *local-global* (ray-mode or wavefront-resonance) *hybrid* [46-48]. For the pulsed signals which are of particular interest in our applications, the *progressing* wave formulation furnishes the natural basis, from which oscillatory events can by synthesized when appropriate.

C. Parametrizing complexity - the importance of scales

The GPR problem belongs to the category of complex problems. One of the characteristics of *complexity* is the presence of many *scales* that play a role in the parametrization. Relevant for the multilayer environment is the ratio of the scales that characterize the layers to the wavelength scales in the incident signal spectrum. With respect to resolution, large (d_n/l_{en}) favors local probing, while small (d_n/l_{en}) favors global probing; here, d_n is the nth layer thickness and l_{en} is the wavelength in the nth layer

material. For the inverse problem, windowed transform processing of the scattered field data adds to the above the scale w representative of the window widths. When the target is included in the scenario, the scales of the target are added to those of the background environment and they influence the choice of the "most promising" target sensor. Finally, when scales in the environment or on the target are irregular, statistical parametrizations with deterministic-stochastic interactions etc, play a role.

D. Learning the Rules

To address the issues raised in the preceding section it is evidently prudent to escalate from "simplicities" to cumulative complexity. Judiciously chosen model problems provide the laboratory for quantification. These model problems should be analytically tractable and simulate conditions to be encountered in GPR applications. Details are given in Sec. III.

E. Beam parametrization of radiation from aperture distributions

Because the representation of radiation from frequency and time domain distributed aperture fields in terms of beams is a basic constituent in the problem strategy, we give here a brief review. We summarize the *continuous* beam basis representation but refer for the corresponding discrete representation on a Gabor lattice to the relevant publications [13].

E.1 Formulation

We begin by first addressing beam propagation in a homogenous medium; extensions for the case of a lossy dispersive, layered medium are discussed subsequently. In the frequency domain, suppressing a time dependence $\exp(-i\omega t)$, assume that field data $u_O(x_O)$ is specified along an aperture in the z=0 plane of a two-dimensional (x,z) coordinate space; x_O tags the x-coordinate in the aperture plane. The procedural steps enumerated next are straightforward (see, for example, [13]) and they lead to the final expression in (1) below. By applying the windowed Fourier transform (or some other sampling window like the Morlet wavelet transform) along the x_O coordinate, one generates the phase space distribution $U_O(x_O',\xi_O')$, where ξ_O is the Fourier spectral wavenumber corresponding to x_O, and (x_O',ξ_O') represents the location of the window center in the (x_O,ξ_O) phase space. The phase space spectral distributions can be propagated away from the aperture plane to an observer plane z > 0 by applying the spectral propagator $\exp[i(k^2-\xi^2)^{1/2}z]$, where $k=\omega/c$ is the wavenumber in the medium and c is the wave propagation speed. The resulting phase space propagators excited by the transformed aperture field are beam-like fields $Y(x,z;x_O',\xi_O')$ which are localized around (x_O',ξ_O'). This leads to the following representation of the field u(x,z) and its spectrum $\hat{u}(\xi,z)\equiv\hat{u}(\xi_O,z)$ [13],

$$u(x,z) = (\omega/2\pi)\, N_x^{-2} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} U_O(x_O',\xi_O') \cdot Y(x,z;x_O',\xi_O')\, dx_O'\, d\xi_O' \qquad (1a)$$

$$\hat{u}(\xi_O,z) = (\omega/2\pi) N_x^{-2} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} U_O(x_O',\xi_O')\hat{Y}(\xi_O,z;x_O',\xi_O') \cdot dx_O'\, d\xi_O' \qquad (1b)$$

where N_X is a normalization constant, while the beam propagator $Y(x,z;x_O',\xi_O')$ and its spectrum $\hat{Y}(\xi_O,z;x'_O,\xi'_O)$ are given by

$$Y(x,z;x_O',\xi_O') = (\omega/2\pi) \int_{-\infty}^{\infty} d\xi_O \hat{Y}(\xi_O,z;x_O',\xi_O') \cdot \exp(i\omega\xi_O x) \qquad (2a)$$

$$\hat{Y}(\xi_O,z;x_O',\xi_O') = \hat{w}(\xi_O-\xi_O')\exp[-i\omega[(\xi_O-\xi_O')x_O'-\zeta_O z]],$$
$$\zeta = [(1/c^2) - \xi_O^2]^{1/2} \qquad (2b)$$

In (2b), for the windowed Fourier transform, $\hat{w}(\xi_O) = (2\pi/\omega)^{1/2}\exp(-\omega\xi_O^2/2\alpha)$ is the spectrum of the Gaussian window $w(x_O) = \exp(-\omega\alpha x_O^2/2)$, with α defining the window width. The aperture domain relations are recovered from (1) and (2) by setting z=0. The formulation above treats the aperture excitation in terms of line source radiators (Kirchhoff formulation) that tie the field at (x,z) to the spatial coordinate x_O [13]. The axis, on which the beam propagator has its maximum, emerges from $x_O=x_O'$ in the aperture plane at an angle $\theta_O=\sin^{-1}(c\xi_O')$ relative to the z-axis. The representation in (1) describes the field as a continuous spectrum of shifted beams, whose origin and axis direction are determined by x_O' and ξ_O', respectively. The transform $U_O(x_O',\xi_O')$ in the z=0 plane specifies the excitation strength of these beams and provides the local matching of these beam fields to the aperture field u_O. The phase space window width determines the degree of confinement (collimation) of these excitation strengths to the vicinity of the axis of the beam propagators. By asymptotic (saddle point) evaluation of the spectral integral in (2a), with (2b), one obtains an explicit approximation Y_S for Y which behaves in the far zone of the (x,z) domain like a conventional Gaussian beam (Fig. 2). Saddle point asymptotics can also be applied to evaluation of the integrals that determine $U_O(x_O',\xi_O')$. With the integrand in (1a) thus approximated asymptotically, the field u(x,z) can be synthesized approximately as well. In asymptotic transition regions where saddle points are not isolated, uniform asymptotic techniques need to be utilized. Details pertaining to the above are documented in ref. [13]. This reference also contains the corresponding discretized representation based on the Gabor lattice in the phase space.

In the time domain, the asymptotic propagators in (2a) are pulsed beam wave packets which are locally matched to the space-time aperture distribution, and the rigorous transforms corresponding to the transform in (1) are the windowed forward and inverse radon transforms. The pulsed beams sample time resolved sections of the transverse beam profile in Fig. 2 at locations ct along the beam axis z_b, with spatial spread around z_b=ct determined by the pulse width. For details, see [14], which again contains also the Gabor representation.

Having thus parametrized the radiation from planar aperture distributions, the local matching of the asymptotic basis beams can be generalized to curved aperture distributions .

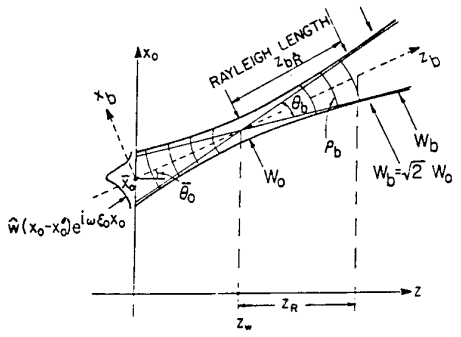


Fig.2 - Time harmonic beam field Y_s and beam parameters in the (x,z) configuration domain, phase matched to the windowed aperture profile in the $(x_0,0)$ plane. (x_b,z_b) are beam-based coordinates, z_w locates the beam waist of width W_0,ρ_b is the phase front radius of curvature, and $W_b=\rho_b\theta_b$ is the corresponding beam width. For pulsed beams, the wave packets are localized around the wavefronts which progress along the beam axis with speed c.

The above summary conveys that the formal frequency and time domain machinery for expanding aperture fields (either primary as produced by a source of radiation, or secondary as represented by scattering data from a target) in terms of a continuous or discrete distribution of beam-type basis fields is well established, as are the first-order asymptotic techniques to extract the wave physics from these formal representations. Thus, these algorithms can be used to calibrate the asymptotics against numerical reference data, and to assess the effects of window choice on the phase space resolution of the beam basis elements.

E.2. Applications and extensions

Having parametrized the launch conditions in free space, the basis beam propagators can be allowed to interact with various environments. In this endeavor, use can be made of the large reservoir of beam tracking algorithms that have been developed in various disciplines, especially in the regime based on paraxial, quasi-plane-wave, etc. approximations for the beam models. Of special interest is the rigorous complex source technique [13-34] whose power resides in the fact that it rigorously converts frequency and time domain Green's functions generated by elementary spatial and space-time sources, respectively, at *real* locations into steady state and pulsed beam excitations in the same environment when these source locations are made *complex* . The conversion applies to exact as well as asymptotic representations and has been utilized in a broad range of applications. Because it is rigorously based (i.e., the complex source beams are rigorous solutions of the wave equation), the method has made it possible to critically assess the effects of a *priori* assumptions in other models in the literature.

After individual phase space beams have been tracked through an environment, they need to be combined according to (1a) or its discretized Gabor version [13] for synthesis of the field u(x,z). Because these beams, which are generated by an extended aperture, arrive from different directions and have therefore sampled different portions of the environment, the asymptotically approximated individual beams may not be adequate and uniform methods may be needed. In this connection, the Gabor basis may offer advantages. In references [27] and [28], it has been demonstrated that narrow Gabor basis beams can explore complex noncanonical propagation environments [28] with remarkable accuracy.

III. Providing the Tools

With these observations as background, let us now consider application to the GPR problem. To provide tools for implementation, several "canonical" constituent problems require attention. Two basic problem categories are highlighted which are relevant not only for the GPR problem but also for other areas of application. Emphasis is on the time domain, especially for excitation by pulsed beams which model space-time resolved (i.e., wideband) wavepackets.

The first problem category is concerned with how lossy dispersive media, in bulk, affect the pulsed beam signal. The second category is concerned with how environments comprised of plane homogeneous layers affect the pulsed beam signal. Note again that beams in both the frequency and time domains a) are useful models individually for studying the wave phenomenology when spatially confined fields interact with environments, and b) can represent , by superposition, any prescribed excitation. For item b), the basis beam parameters can be chosen so as to yield well-collimated probes which can sample local inhomogeneities in an environment if the variations of these inhomogeneities are small over the beam width; with respect to localization, beams may be regarded as generalized rays (in fact, beams are complex rays [15-28]) which are not beset by the "catastrophes" that nonuniform ray theory encounters at caustics, shadow boundaries, etc. The outcomes from these investigations, i.e. exploration of the soil environment as such, are relevant for subsequent studies of buried targets in that environment.

A. Pulsed beam interaction with homogenous lossy dispersive materials

This is a basic electromagnetic problem that relates to materials of all types. One of the most thorough studies of transient fields in particular homogenous lossy dispersive media has been carried out by Oughston and collaborators and has been documented in a monograph [35]. This monograph presents theorems and rigorous alternative representations pertaining to general space-time dependent fields in such media (part I of [35]); the asymptotic reduction of the formal representation integrals via saddle point techniques has been carried out for the special case of pulsed plane wave excitation of a single resonance Lorentz medium (part II of [35]). The intricate asymptotics identifies and quantifies the relevant wave physics that gives rise to the distortions of the pulsed plane wave signal input, and furnishes valuable insight into the relation between the phenomenology and the model parameters as well as the input pulse shape for the plane wave case in that special medium. This still leaves the general *phenomenology* associated with a *pulsed beam* input (i.e., temporal *and* spatial localization) as a separate problem.

For the simplest nondispersive lossy model $\varepsilon=\varepsilon_r + i\sigma/\omega$, where ε_r and σ are positive real, the pulsed beam propagation problem has recently been addressed by Heyman, Tijhuis and Boersma [49], using the complex space-time source point approach. Their results furnish the exact and asymptotically approximated propagators in such a medium, and parametrize the effect of the loss factor σ on the pulse shape. For more general dispersive models, when these are required, it is unlikely that one can develop exact closed form expressions. Now, a generalization of the Oughston formulation [35] will be appropriate. However instead of aiming at descriptions that approach the level of sophistication achieved in the

plane wave pulse problem in [35] for a *particular* medium, it suffices to seek *generic* parameters that are tied to the *beam* parameters and the *more general* lossy dispersive $\varepsilon(\omega)$ under consideration. One may start with the pulsed beam formal spectral integrals which involve the transverse spatial wavenumber $\mathbf{k}_t$ and the temporal frequencies ω for lossy (i.e. complex) $\varepsilon(\omega)$. By performing sequential saddle point asymptotics, these integrals can be reduced to approximate forms that highlight the dependence on loss and beam parameters. This should yield the pulsed beam phenomenology for the well developed dispersive regime. For the precursors at short time scales, i.e., near the wavefronts, alternative asymptotics will be explored [36]. These studies should suffice to extract generic phenomenology, which can be compared with the complex space-time ray asymptotics performed in [16], [50]. Calibration of the asymptotics against numerically computed results for relevant test models is an option that can be considered in this context.

Interpretation of the space-time asymptotic results should be carried out in terms *of transient plane wave spectra* parametrized by the *spatial* wavenumbers, thereby parametrizing the entire wave process *directly* in the time domain. In this connection, the complex space-time source point technique [24,26,42], which deals directly with the time domain, is suggestive. After the problem in the unbounded homogeneous medium has been settled, adaptation to weakly inhomogeneous media can be handled by adiabatic asympotics.

B. Synthesizing focusing beam and incident aperture profiles

Synthesis of input pulsed beam and aperture profiles for *focusing* the beam energy at a particular space-time location in the lossy, dispersive environment, is a problem of special importance. Recent studies have dealt with collimated ultra-wideband/short pulse aperture profiles in free space [37,43]. It is shown there how to synthesize the input profile by frequency-dependent beam apertures so as to render the Fresnel distance frequency insensitive. This yields better collimation than a constant-width aperture where that distance depends on frequency. Also relevant are previous asymptotic studies of plane wave pulse compression (i.e., space-time focusing) in lossless dispersive media [38].

The challenge is to extend these concepts to pulsed beam apertures in lossy dispersive media. It is suggestive to initially address the $(\varepsilon_r + i\sigma/\omega)$ model, for which the (pulsed beam parameter)-(medium-parameter) connection [49] takes on a simple form. For more general lossy medium models, the complex space-time ray asymptotics in [50] are appropriate, which reveal considerably more complex phenomenology, even for plane wave pulses. It remains to be seen what relevant approximate information can be extracted via pulsed beam asymptotics; due to the beam localization, the algorithms may actually simplify.

C. Pulsed beam interaction with layered lossy dispersive materials

In this model, lossy dispersive materials are arranged in homogenous parallel layers so as to simulate soil stratification. The objective is to explore how layering affects an incident pulsed beam signal; it is assumed that via the bulk studies in Sec. A, the effects of loss and dispersion as such are understood. For normal incidence, the relevant mechanism is single and multiple reflection at the boundaries, with the multiply reflected field confined to the vicinity of the normal plane of incidence. This mechanism can be understood by first replacing the layers with a half space and establishing the reflection and transmission coefficients for a single interface; the multiple reflection mechanism is then built up sequentially. If the multiple reflections can no longer be resolved individually, collective summation into modal resonances in the plane-of-incidence cross section becomes an alternative descriptor. The extension from the bulk studies in Sec. A to this normal incidence case is expected to be straightforward.

For oblique incidence of the pulsed beam signal, the phenomenology changes drastically. Multiple reflections between the layers now propagate laterally along the layers, thereby producing energy transport (guiding) in the lateral direction. Moreover, depending on the layer composition, a beam incident obliquely beyond the critical angle at an interface will produce a strong reflected and weak transmitted field which, under lossless conditions, would be evanescent (decay exponentially). The global (collective) effects of the layer-trapped multiples can by expressed in terms of trapped and leaky modes which propagate along the layer direction and shed energy back into the exterior of the layered half space (the air region above ground can be regarded as an infinitely wide layer). The time domain leaky modes (which are sustained, damped oscillatory waveforms with time-dependent wavenumbers) are new wave objects which were first explored (to the best of our knowledge) in [39-41] (it should be recalled that extrapolation to the time domain of the frequency domain leaky mode phenomenology fails under short pulse conditions since the frequency dependent leakage angles then vary over a wide interval). In [39-41], the configuration consists of a *nondispersive, lossless* grounded dielectric slab excited by a pulsed horizontal dipole element inside the slab, with the fields observed outside the slab. The time domain leaky mode phenomenology, including the mechanisms of excitation and leakage, as derived by rigorously based asymptotics, is schematized in Fig. 3. The analysis revealed a curious anomalous leaky mode behavior when the problem conditions permit Brewster angle (perfect) transmission; at the instant of time when the leaky mode is matched to the Brewster angle, the radiated leaky mode field undergoes a highly singular transition which has not yet been traced out asymptotically but which can be avoided by employing a hybrid wavefront-resonance algorithm[41].

The results in [39-41] pertain only to the special problem conditions noted above. Therefore, much can be learned from this canonical configuration with respect to the GPR problem conditions. Further studies should include: *A. Lossless nondispersive case* - 1. replacing the pulsed dipole element by a pulsed beam; 2. allowing all combinations of inside-outside source and observer locations (although the radar is always in the exterior half space, energy can be coupled into a layer from the inside by virtual sources which model scattering from a strong soil inhomogeneity); 3. examining the effect of lateral waves excited by a wavefront incident at the critical angle (this regime was avoided in [39-41]; 4. creating a dielectric gap by having $\varepsilon_{int}>\varepsilon_{ext}$ (i.e., no trapped modes). *B. Lossy dispersive case* - introducing loss and dispersion into the items listed under A. These investigations should be performed selectively according to need. Concerning calibration of the analytic asymptotics against numerical reference solutions, this does not necessarily have to be performed on the idealized homogenous plane layer model but can be carried out, by local adaptation of the beam asymptotics, on noncanonical configurations

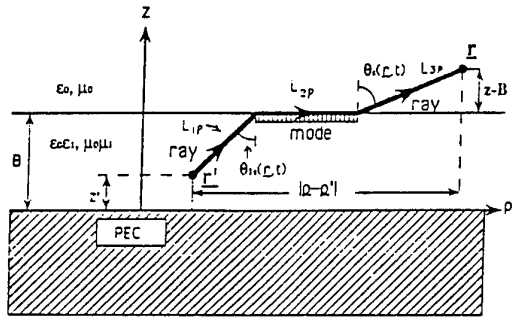


Fig. 3. Physical configuration of horizontal electric current element located at r′ inside a grounded dielectric layer, with the observer r outside the layer. Also shown is the ray-mode interpretation of the space-time evolution of a leaky mode field excited by the pulsed dipole. Paths L_{1p} and L_{3p} describe trajectories of *ray* fields in the slab and exterior bulk media, respectively, whereas L_{2p} describes the trajectory of the pth leaky mode. The ray field wave vectors along the ray paths L_{1p} and L_{3p} are phase matched to the leaky mode wave vector along L_{2p}, with the space-time dependent leaky mode launch and detachment angles θ_{1s} and θ_s adjusted continuously so as to satisfy also the space-time dependent leaky mode dispersion relation.

provided that deviations from the ideal prototype occur over scales that are large compared to the beam widths. It should be noted that these extensions to lossy dispersive conditions are nontrivial, can give rise to substantial changes in phenomenology, and may require more sophisticated asymptotics than those which extracted the time domain leaky modes for the lossless nondispersive case.

D. Including the target

The methodology for scattering from a target buried in the soil background has been detailed in Sec. II. The relevant problem sees the target in its *local* environment. The simplest local environment is a thick layer whose boundaries are so far removed from the target as to allow modeling in unbounded lossy space. The next escalations in "complexity" involve a target buried near the surface of a lossy half space, a target buried inside a layer with nearby boundaries, and so on. The matching from the background - corrected asymptotic incident field produced by the beam basis functions on the target domain boundary S_T to whatever basis is adopted for the "interior" problem can be handled to a lowest order of approximation by Kirchhoff-type assumptions wherein fields incident on one side of S_T furnish the initial conditions for outgoing fields on the other side, and vice-versa. It should be emphasized again that via the progressing pulsed - beam formalism adopted here, the asymptotic beam tracking can be performed, in principle, through any background (bulk material and/or interface) configuration provided that the previously stated "slow variability conditions" with respect to the beam widths are satisfied.

D.1 Weak-contrast targets -- Born-type scattering

For weak nondispersive inhomogeneities in a homogeneous lossless nondispersive background, recent studies have shown how frequency domain diffraction tomographic methods for object reconstruction can be generalized to interrogation by pulsed plane wave inputs [51], and thereby take advantage of the resolution capabilities of the short pulse signal. Going further, work now in progress shows that interrogation via *pulsed beams* optimizes the *local* reconstruction process [52, 53]. It is interesting to ask whether and how the procedure in [52, 53] can be applied to effect weakly-scattering-object reconstruction in a *lossy* environment.

IV. Summary

There has been a growing awareness that recourse to phenomenology can play an important role when devising processing schemes for data from complex scattering environments. Without *some* phenomenological (i.e., wave physics) input into algorithms that rely entirely on non-wave-oriented statistical and other parametrizations, an opportunity may have been lost to reduce the complexity of the image processing by ignoring *a priori* information. If this premise is accepted for the overall GPR problems, the question is where, in the modeling chain, phenomenology makes sense, and what *measures* of that phenomenology (local or global, fine or coarse, etc.) should be utilized. The decision is evidently problem dependent, and is also influenced by the other parametrization schemes in the process.

The objective in this paper has been to parametrize the wave phenomenology for the GPR soil background environment by employing beam propagators as the basic synthesizing wave objects, separating the background domain from the target domain, and connecting the two domains via wave-based assumptions. Only "regular" background conditions have been considered, which do not require statistical measures. What portions of the overall problem can be phrased around this background and can then be perturbed by statistical measures depends on the overall problem conditions. The phenomenological hierarchy has been addressed here in detail, with due consideration of local vs. global characterizations, dispersive vs. nondispersive effects, etc., and the analytic framework for calibrating the beam basis algorithms has been stated.

Clearly, implementation of the algorithm by beam tracking through the overall complex environment is computationally massive, and is *not* advocated here (unless it makes sense). Instead, by having listed the various anticipated phenomenologies and having stated how they can be analytically calibrated, an *intelligent* choice can be made as to "which aspect" and "what measure" of relevant wave physics should be embedded within the overall modeling strategy that includes non-wave-oriented (numerical and model-based) techniques. This decision requires the intimate coupling of the wave-oriented and non-wave-oriented portions. Only the former aspect has been discussed in this write-up., and the list of references cited here is limited accordingly. The latter aspect has been pursued in other investigatoins (not cited here), but a sample of those performed by L. Carin et al. has been documented in [54-56]. Collaboration is continuing to provide the interactive framework for blending these alternative modeling strategies into an overall package with better potential than each alone [57-59].

V. References

[1] Moghaddar and E.K. Walton, "Time-frequency distribution analysis of scattering from waveguide cavities," *IEEE Trans. Antennas Prop.*, vol. 41, pp. 677-679, May 1993.

[2] L. Carin, L.B. Felsen, S.U. Pillai, D. Kralj, and W.C. Lee, "Dispersive modes in the time domain: analysis and time frequency representation," *IEEE Microwave and Guided Wave Letts.*, Jan. 1994.

[3] D.R. Kralj, M. McClure, L. Carin, and L.B. Felsen, "Time domain wave-oriented data processing for scattering by nonuniform trunated gratings," *J.*

Optical Soc. America A, vol. 11, pp. 2685-2694, Oct. 1994.

[4] H. Kim and H. Ling, "Wavelet analysis of backscattering data from an open-ended waveguide cavity," *IEEE Microwave and Guided Wave Letts.*, vol. 2, pp. 140-142, April 1992.

[5] H. Kim and H. Ling, "Wavelet analysis of radar echo from finite size targets," *IEEE Trans. Antennas Propagat.*, vol. 41, pp. 200-207, Feb. 1993

[6] S.R. Cloude, P.D. Smith, A. Milne, D.M. Parkes, and K. Trafford, "Analysis of time domain ultrawideband radar signals," in Ultra-Wideband Short-Pulse Electromagnetics, Plenum: New York (H.L. Bertoni, L. Carin, and L.B. Felsen, Eds.), pp. 445-456, 1993.

[7] G.C. Gaunaurd, H.C. Strifors, A. Abrahamsson, and B. Brusmark, "Scattering of short EM-pulses by simple and complex targets using impulse radar," in Ultra-Wideband Short-Pulse Electromagnetics, Plenum: New York (H.L. Bertoni, L. Carin, and L.B. Felsen, Eds.), pp. 437-444, 1993.

[8] E.K. Walton and A. Moghaddar, "Time frequency-distribution analysis of frequency dispersive targets," in Ultra-Wideband Short-Pulse Electromagnetics, Plenum: New York (H.L. Bertoni, L. Carin, and L.B. Felsen, Eds.), pp. 423-436, 1993.

[9] L. Carin and L.B. Felsen, "Ultra-widebanded, short-pulse electromagnetics at Polytechnic University," *IEEE Antennas and Propagat. Magazine*, June 1994.

[10] L. Cohen, "Time-frequency distributions - A review," *Proc. IEEE*, vol. 77, pp. 941-981, July 1989.

[11] M. McClure, D.R. Kralj, T.-T. Hsu, L. Carin and L.B. Felsen, "Frequency domain scattering by nonuniform truncated arrays: wave-oriented data processing for inversion and imaging," *J. Optical Soc. America A*, vol. 11, pp. 2675-2684, Oct. 1994.

[12] L.B. Felsen and L. Carin, "Wave-oriented processing of scattering data," *Elect. Letters*, vol. 29, pp. 1930-1932, Oct. 28, 1993.

[13] B.Z. Steinberg, E. Heyman, and L.B. Felsen, "Phase space beam summation for time-harmonic radiation from large apertures," *J. Opt. Soc. Am. A*, vol. 8, pp. 41-59, 1991.

[14] B.Z. Steinberg, E. Heyman, and L.B. Felsen, "Phase space beam summation for time-dependent radiation from large apertures: continuous parametrization," *J. Opt. Soc. Am. A*, vol. 8, pp. 943-958, 1991.

[15] L.B. Felsen, "Complex-source-point solutions of the field equations and their to the propagation and scattering of Gaussian beams," Symposia Mathematica, Istituto Nazionale di Alta Matematica, vol. XVIII, Acad Press, London and New York, 40-56, 1976.

[16] K.A. Connor and L.B. Felsen, "Gaussain pulses as complex-source-point solutions in dispersive media," *Proc. of the IEEE Issue on Rays and Beams*, Vol. 62, pp. 1614-1615, Nov. 1974.

[17] S.Y. Shin and L.B. Felsen, "Gaussian beams in anisotropic media," *Applied Physics*, vol. 5, pp.239-250, 1974.

[18] L.B. Felsen, "Rays, modes, and beams in optical fiber waveguides." *Optical and Quantum Electronics*, vol. 9, pp. 189-195, May 1977.

[19] S.Y. Shin and L.B. Felsen, "Gaussian beam modes by multipoles with complex source points,," *IEEE Trans. Antennas Propagat.*, pp.189-195, May 1977.

[20] G. Ghione, I. Montrosset and L.B. Felsen, "Complex ray analysis of radiation from large apertures with tapered illuminations," *IEEE Trans. on Antennas and Propagation*, vol. AP-32, pp.689-693, 1984.

[21] L.B. Felsen, "Geometrical theory of diffraction, evanescent waves, complex rays and Gaussian beams," *Geophys. J. Roy. Astron. Soc.*, vol. 79, pp.77-888, 1984.

[22] Y.Z. Ruan and L.B. Felsen, "Reflection and transmission of beams at a curved interface," *J. Opt. Soc. Am. A*, vol. 3, April 1986.

[23] H. Ikuno and L.B. Felsen, "Complex rays in transient scattering from smooth targets with inflection points," *IEEE Trans. Antennas Propagat.*, vol. AP-36, pp.1272-1280, Sept. 1988.

[24] E. Heyman and L.B. Felsen, "Propagating pulsed beam solutions by complex source parameter substitution," *IEEE Trans. Antennas Propagat.*, vol. AP-34, pp.1062-1065, 1986.

[25] L.B. Felsen, "Systematic study of fields due to extended apertures by Gaussian beam discretization" *IEEE Trans. Antennas Propagat.*, vol. AP-37, pp.884-892, 1989.

[26] E. Heyman and L.B. Felsen, "Complex source pulsed beam fields," *J. Opt. Soc. Am. A*, pp. 806-817, 1989.

[27] J.Maciel and L.B. Felsen, "Gaussian beam analysis of propagation from an extended plane aperture distribution through dielectric layers: I- Plane layer,"*IEEE Trans. Antennas Propagat.*, vol. AP-38, pp.1607-1617, 1990.

[28] J.Maciel and L.B. Felsen, "Gaussian beam analysis of propagation from an extended plane aperture distribution through dielectric layers: II-Circular cylindrical layer,"*IEEE Trans. Antennas Propagat.*, vol. 38, pp.1618-1624, 1990.

[29] E. Heyman and L.B. Felsen, "Real and complex spectra- A generalization of WKJB seismograms," *Geophys. J. Roy. Astron. Soc.*, vol. 91, pp.1087-1126, 1987.

[30] L.B. Felsen and S. Zeroug, "Beam parametrization of weak debonding in a layered aluminum plate," Review of Progress in Quantitative NDE, Plenum Press, pp.195-202, 1990.

[31] L.B. Felsen and S. Zeroug, "Ultrasonic beam method for localized weak debonding in a layered plate," *J. Acoust. Soc. Am. A*, vol. 90, pp.1527-1538, 1991.

[32] I.T. Lu, L.B. Felsen, J.M. Klosner, "Beams and modes for scattering from weak bonding flaws in a layered aluminum plate," *J. Acoust. Soc. Am.*, vol. 88, pp. 496-504, 1990.

[33] I.T. Lu, L.B. Felsen and J.M. Klosner, "Beam-to-mode conversion in an aluminum plate for ultrasonic NDE applications," *ASME J. of Engineering Materials and Tech.*, vol. 112, pp.236-240, 1990.

[34] I.T. Lu, L.B. Felsen, J.M. Klosner and C. Gabay, "Beams and modes for scattering from weak bonding flaws in a layered aluminum plate," , *J. Acoust. Soc. Am. A*, vol. 88, pp. 496-504, 1990.

[35] K.E. Oughston and G.C. Sherman, "Electromagnetic Pulse Propagation in Causal Dielectrics," Springer-Verlag, New York, 1994.

[36] L.B. Felsen, "Transients in dispersive media, Part I: Theory,"*IEEE Trans. Antennas Propagat.*, vol. AP-19, pp.424-432, 1971.

[37] E. Heyman and T. Melamed, "Certain considerations in aperture synthesis of ultra-wideband/short-pulse radiation," *IEEE Trans. Antennas Propagat.*, vol. AP-42, pp.518-525, 1994

[38] L.B. Felsen, "Asymptotic theory of pulse compression," *IEEE Trans. Antennas Propagat.*, vol. AP-19, pp.424-432, 1971.

[39] L.B. Felsen and F. Niu, " Spectral analysis and synthesis options for short pulse radiation from a point dipole in a grounded dielectric layer," *IEEE Trans. Antennas Propagat.*, vol. AP-41, pp. 747-754, June 1993

[40] F. Niu and L.B. Felsen, "Spectral analysis and synthesis options for short pulse radiation from a point dipole in a grounded dielectric layer," *IEEE Trans.*

Antennas Propagat., vol. AP-41, pp. 762-769, June 1993

[41] F. Niu and L.B. Felsen, "Asymptotic analysis and numerical evaluation of short-pluse radiation from a point dipole in a grounded dielectric layer," *IEEE Trans. Antennas Propagat.*, vol. AP-41, pp.762-769, June 1993

[42] E. Heyman, "Pulsed beam propagation in inhomogeneous medium," *IEEE Trans. Antennas Propagat.*, vol. AP-42, pp.311-319, 1994.

[43] A.M. Shaarawi, I.M. Besieris, R.W. Ziolkowski and S.N. Sedky, "Generation of approximate focus-wave-mode pulses from wideband dyanmic Gaussian apertures," *J. Opt. Soc. Am. A*, vol. 12, pp. 1954-1964, 1995.

[44] E. Heyman, B.Z. Steinberg and L.B. Felsen, "Spectral analysis of focus wave modes," *J. Opt. Soc. Am. A*, vol. 4, pp. 2081-2091, 1987.

[45] E. Heyman and L.B. Felsen, "Comments on 'Nondispersive waves: interpretation and causality'," *IEEE Trans. Antennas Propagat.*, vol. AP-42, pp.1668-1670, 1994.

[46] L.B. Felsen, "Progressing and oscillatory waves for hybrid synthesis of source excited propagation and diffraction," *IEEE Trans. Antennas and Propagat.*, AP-32, 775-796 , 1984.

[47] I.T. Lu and L.B. Felsen, "Ray, mode and hybrid options for transient source excited propagation in an elastic layer," *Geophys. J. Roy. Astron. Soc* . 86, 177-201 , 1986.

[48] E. Heyman and L.B. Felsen, "Traveling wave and SEM representations for transient scattering by a circular cylinder," *J. Acoust. Soc. Am.* 79(2), 230-238 , 1986.

[49] E. Heyman, A. Tijhuis and J. Boersma, "Spherical and collimated pulsed fields in conducting media," *presented at the URSI Electromagnetic Theory Symposium*, St. Petersburg, Russia, August 1995.

[50] K.A. Connor and L.B. Felsen, "Complex space-time rays and their application to pulse propagation in lossy dispersive media," *Proceedings of the IEEE, Special Issue on Rays and Beams*, Vol. 62, No. 11, 1586-1598 , 1974.

[51] T. Melamed, Y. Ehrlich and E. Heyman, "Short-pulse inversion of inhomogeneous media," *submitted to Inverse Problems.*

[52] T. Melamed, E. Heyman and L.B. Felsen, "Local spectral analysis for short-pulse inversion of inhomogeneous media: Part I--spectral properties," *in preparation.*

[53] T. Melamed, E. Heyman and L.B. Felsen, "Local spectral analysis for short-pulse inversion of inhomogeneous media: Part II--reconstruction," *in preparation.*

[54] S. Vitebskij and L. Carin, "Moment-method modeling of short pulse scattering from , and the resonances of, a wire buried in a lossy, dispersive half space", *IEEE Trans. Antennas Propagat.*, vol. AP-43, Nov. 1995

[55] S. Vitebskij, K. Sturgess and L. Carin, "Short pulse scattering from buried perfectly conducting bodies of revolution" *IEEE Trans. Antennas Propagat.*, to appear

[56] S. Vitebskij and L. Carin, "Resonances of perfectly conducting wired and bodies of revolution buried in a lossy dispersive half space", submitted to *IEEE Trans. Antennas Propagat.*

[57] L. Carin, L.B. Felsen and T.T. Hsu. "High frequency fields excited by truncated arrays of nonuniformly distributed filamentary scatterers on an infinite dielectric slab: Parameterizing (leaky-mode) - (Floquet-mode) interaction", *IEEE Trans. Antennas Propagat.* vol. 44, pp. 1-11, 1996

[58] T.T. Hsu, M.R. McClure, L.B. Felsen, and L. Carin, "Wave-oriented processing of scattered field data from a plane-wave-excited finite array of filaments on an infinite dielectric slab", *IEEE Trans. Antennas Propagat.*, vol. 44, pp. 352-360, 1996

[59] L. Carin, L.B. Felsen, D.R. Kralj, H.S. Oh, W.C. Lee, and S.U. Pillai, "Wave-oriented signal processing of dispersive time domain scattering data", submitted to *IEEE Trans. Antennas Propagat.*

Paper 24

K. Anderson (US)

Can you describe the "collective ray'?

Author's reply:

If ray fields are transmitted through a dielectric slab with nearby laterally displace source and receiver, the initial rays arrive from different directions and must therefore be treated individually. However, after a sufficiently large number of internal reflections, neighboring rays arrive almost co-directionally and can therefore be treated approximately as parallel.

This leads to a sum of the form:

$$\sum_{n=N}^{\infty} e^{ik_\epsilon n(\ell_{12}+\ell_{21})} R_1^n R_2^n = \underbrace{e^{ik_\epsilon N(\ell_{12}+\ell_{21})}}_{\text{Phase of } N\text{th ray}} \underbrace{\frac{1}{1-e^{ik_\epsilon(\ell_{12}+\ell_{21})}R_1R_2}}_{\text{plane wave slab transmission coeff}} \underbrace{(R_1R_2)^N}_{\text{reduced by the missing } N \text{ rays}}$$

collective weighting of Nth ray accounts for all $n \geq N$ rays

A GENERIC IMAGE PROCESSING TOOL : TOP DOWN BAYESIAN IMAGE ANALYSIS

S. Dewitte[1,2],J.M. Mangen[3], J. Cornelis[1], D. Crommelynck[2]
1: Vrije Universiteit Brussel, Department of Electronics, Pleinlaan 2, B-1050 Brussels, Belgium
2: Royal Meteorological Institute of Belgium, Department of Aerology, Ringlaan 3, B-1180 Brussels, Belgium
3: Royal Military Academy,Department of Electricity, Renaissancelaan 30, B-1040 Brussels,Belgium

steven@oma.be

Abstract: Most of the classical image processing algorithms, especially in the field of pattern recognition, are bottom-up algorithms, i.e. they start with working on a large amount of local data (pixels, features, micro segments, ...) and end with combining the information gained from these data in the later stages to obtain a smaller amount of global information (classes, segments,...). This way of working is not very efficient, neither in computing time, nor in human development effort. The inefficiency in computing time resides in the use of a large amount of local data to obtain a smaller set of data. The inefficiency in development effort lies in the specificity of the algorithm to a certain problem which is a consequence of starting the information extraction at the local level.
As an alternative to the specific bottom-up algorithms, this work presents a generic top-down algorithm, i.e. a general purpose image analysis algorithm which starts with analyzing a small amount of global data and refines these analysis results to a large amount of local information.
The top down mechanism used, is inspired by the Laplacian pyramid and by pel-recursive motion estimation techniques: an analysis is done on non-overlapping square segments of which the dimensions are recursively divided by two. The recursion starts with analyzing one segment, which is the entire image (top of the pyramid) and ends with an analysis of every pixel as a separate segment (bottom of the pyramid).
As refinement mechanism the theorem of Bayes is used. The analysis result of a child segment, which is part of an already analyzed parent segment, is determined as the most probable result, calculated according to the theorem of Bayes. The parent analysis is used to compute the a priori probabilities, the local analysis within the child segment is used to compute the a posteriori probabilities. In this way refinement by local analysis is combined with regularisation by larger scale analysis.
This leads to an 'empty' generic image analysis tool, which can be 'filled' with a specific analysis type to obtain a specific image analysis algorithm. To demonstrate this, different image analysis algorithms are obtained with very little effort by choosing different specific analysis schemes: smoothing filtering, image interpolation, motion estimation. The usefulness of these different algorithms is demonstrated for the processing of satellite images: subpixel interpolation of ERBE images by the use of METEOSAT images, METEOSAT image compression, METEOSAT time interpolation.

1. INTRODUCTION

Because the image usually represents a number of physical objects, with varying sizes, shapes and internal structures, some correlation is present between the image pixels [1], e.g. the presence of locally similar greyvalues for the pixels within one object, or the presence of an oriented edge for the pixels laying at the border between two distinct objects. Since an image can be build up by objects of different sizes and since large objects can have fine microstructure, a complete description at all scales [2], i.e. some kind of multi-resolution description, is needed.
Now one can search for general multi-resolution image descriptions, which are suited for all image processing tasks, or for large subclasses of them. If compactness of description is important, as e.g. in image compression, the wavelet transform [3] is a good candidate.
The wavelet transformed data (wavelet coefficients) are essentially a set of lowpass filter output data and different sets of highpass filter output data describing the image completely [4]. The highpass output data are subsampled critically, i.e. subsampled so that no redundant wavelet coefficients (coefficients which can be calculated out of other coefficients) are left. This critical subsampling unavoidably causes the wavelet transform to be shift-variant [5], i.e. the way a pixel is transformed varies with its position in the image.
As a conclusion, the critical subsampling makes the wavelet transform well suited for image compression [6],[7] (because of the redundancy reduction), but badly suited for image analysis, especially for pattern recognition [8],[9] (because of shift variance).
This does not mean, however, that all properties of the wavelet transform are bad for image analysis. It provides a natural basis for a top down analysis of the image, i.e. a progressive coarse to fine scale analysis in which previously found large scale analysis results can serve as context to restrain analysis at smaller scales. This kind of analysis is naturally robust with respect to noise and model errors, because at large scale a global analysis is made exploiting large numbers of pixel values, and because at small scales (where the number of analysed pixel values is small and could become dominated by noise or model errors) the analysis is regularized by the low noise large scale analysis.

Paper presented at the AGARD SPP Symposium on "Remote Sensing: A Valuable Source of Information", held in Toulouse, France from 22-25 April 1996, and published in CP-582.

In this paper we present a general image processing methodology, called top down Bayesian analysis, a top down mechanism inspired by the wavelet transform, but which does not have the drawback of shift-variance, and which tries to be more general (but less efficient in computing) than an image analysis based on wavelet coefficients.
The method analyses the original image pixel values in non-overlapping square segments whose linear size is recursively divided by two, as in the 2D separable wavelet transform. The division in segments is only used to define the analysis regions, not to transform the image. The actual analysis is done on the original image pixel values, not on transformed data. In this way, a maximal separation is obtained between the top-down mechanism (wavelet inspired) and the actual image analysis (not wavelet inspired).
The analysis starts with analyzing one segment, which is the entire image (top of the pyramid) and ends with an analysis of every pixel as a separate segment (bottom of the pyramid).
As a refinement mechanism the theorem of Bayes is used. The analysis result of a child segment, which is part of an already analyzed parent segment is determined as the most probable result, calculated according to the theorem of Bayes. The parent analysis is used to compute the prior probabilities, the local analysis within the child segment is used to compute the a posteriori probabilities. In this way refinement by local analysis is combined with regularisation by larger scale analysis.
Throughout the paper, METEOSAT satellite images will be used as main test images. In section 2, we investigate the validity of a top-down analysis for these kind of images by investigating the wavelet compression [10] results. This allows to reformulate the intuitive notion of "natural images" to the mathematical notion of positive wavelet decay coefficient [5][11][12]. In section 3, we give a mathematical definition of a top down Bayesian image analysis. This leads to an 'empty' generic image analysis tool, which can be 'filled' with a specific analysis type to obtain a specific image analysis algorithm. In section 4, we define three specific analysis types: smoothing interpolation, non-smoothing interpolation and motion estimation. We work out the mathematical formulation for these three analysis types using Gaussian distribution assumptions. In section 5, we show an example of application in the processing of METEOSAT images for each of the three analysis types of section 4. Non-smoothing interpolation is used to compress cloud-only segments of METEOSAT images. Smoothing interpolation is used in the subpixel interpolation of ERBE data with METEOSAT images. Motion estimation is used for the time interpolation of METEOSAT images.
We conclude with a review of the obtained results.

2. Wavelet compression

2.1 Wavelet compression as an exploitation of general image properties

The 2D discrete wavelet transform decomposes [3],[4] an image in a series of basefunctions which have local properties in both the spatial and frequency domains.
The basefunctions can be grouped in a set of lowpass functions and different sets of highpass functions.
The coefficients of the lowpass functions contain the information to obtain a low resolution (smoothed) approximation of the transformed image. The coefficients of one set of highpass functions contain the differential information to pass from an approximation at a low resolution to en approximation at a higher resolution.
As the number of lowpass coefficients is small and the values of the highpass coefficients are mostly small, quantisation and coding of the wavelet coefficients in a similar way[1] as the quantisation and coding of the DCT (Discrete Cosine transform) coefficients in the standard JPEG (Joint Photographic Expert Group) compression method, gives good compression results [6]. Usually the results are better than for JPEG thanks to the absence of blocking effects.
Even better compression results can be obtained when the localization of the wavelet functions is exploited. For natural images the largest part of the information is contained in the coefficients of the highpass functions, localized along the edges of the objects in the image. Along the edges the highpass coefficients at all scales are large and proportional to the edge intensity. But, at places where the image is sufficiently smooth, i.e. everywhere far away from the edges and dirac-like impulses, highpass coefficients are relatively small and small scale highpass coefficients have even smaller amplitudes than large scale highpass coefficients with the same localization [5][11][12]. This is exploited in [7] by introducing a special symbol, the zerotree symbol, for coding local groups of highpass coefficients which are decreasing with decreasing scale. In [10] this coding has been successfully applied to the quantified wavelet coefficients of METEOSAT images, yielding systematically better compression results than JPEG compression applied to the same image.

2.2 Wavelet compression of METEOSAT images

The scheme described above can be applied directly to Meteosat images, giving good results when compared to other compression methods, like JPEG for example. This is illustrated in Figure 1.

[1] strong quantisation of the highpass coefficients, weak quantisation of the lowpass coefficients

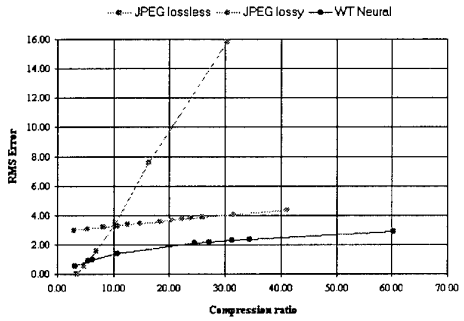


Figure 1 : RMS error versus compression ratio for METEOSAT IR images for lossless JPEG, lossy JPEG and neural wavelet compression

However, Meteosat images are composed of a series of measurements (the infrared image is a collection of measurements of the temperature), and, as such, meteorologists do not accept that high local distortions appear due to the compression process. Therefore, specific schemes to lower the value of the maximum absolute error have to be developed.
Several methods can be applied to optimize such a criterion: in [13], a neural network is used to implement a gradient descent of a cost function equal to the maximum absolute error; in [14], the largest errors are coded using a Huffman like algorithm. The results obtained using this kind of algorithm are compared to those of a 'regular' wavelet-based algorithm in Figure 2 (on the left side: regular algorithm; on the right side: optimized algorithm): the large white dots correspond to an absolute error of 9 gray levels, the large black ones to 8, medium black to 7 and small black to 6.

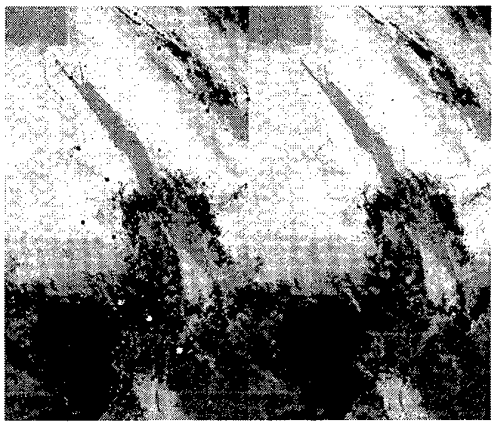


Figure 2 : comparison of errors for Red Sea area in METEOSAT image. Left: normal wavelet compression, maximum absolute error: 9 counts (white dots) right: neural wavelet compression, maximum absolute error: 6 counts (small black dots)

Finally, the same results are compared to those of JPEG (JPEG lossy with a varying quality factor, and JPEG lossless using a varying Point transform - i.e. coding only part of the input samples bits).

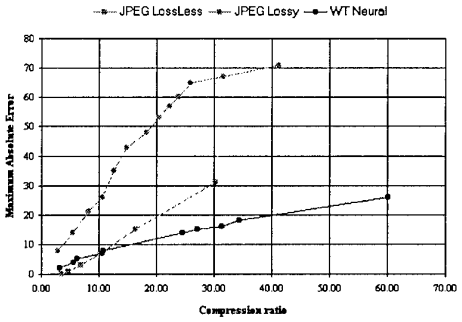


Figure 3 : maximum absolute error versus compression ratio for METEOSAT IR images for lossless JPEG, lossy JPEG and neural wavelet compression

2.3 Wavelet compression of cloud-only METEOSAT images

We tried to repeat this for the METEOSAT derived cloud-only images, but we obtained totally different results. The first two graphs of Figure 4 show the compressed image RMS-error versus file size for the direct wavelet compression (wavelet transform + quantisation + zerotree coding) and for the JPEG compression of the cloud-only image of Figure 5. Only for low compression (right part of the graphs) the wavelet compression gives better results than JPEG. For higher compression (points more to the left), obtained by stronger quantification of the wavelet highpass coefficients, the wavelet compression behaves very badly.
A quick analysis of the error images (not shown here) leads us to the hypothesis that this is due to the strong artificial edges in the image. When the highpass coefficients are quantified strongly, the errors in the neighborhood of these edges rise very quickly. Moreover, as the zerotree coding focuses on coding smooth image zones efficiently, its compression performance decreases for the edges. To test if the loss in compression performance is really due to the strong edges and to try to do something against it, we adapted our wavelet compression method as explained in 5.1.

3. TOP DOWN BAYESIAN ANALYSIS AS A GENERAL IMAGE ANALYSIS TOOL

In this section we give a mathematical definition of top down Bayesian image analysis.
We have an input image with height n, width m and p bands

$$\vec{I}(i,j)=\begin{bmatrix} i_1(i,j) \\ \vdots \\ i_k(i,j) \\ \vdots \\ i_p(i,j) \end{bmatrix} \quad \begin{matrix} i=1,\ldots,n \\ j=1,\ldots,m \end{matrix} \qquad \text{Equation 1}$$

A segment with linear size s and topleft corner (i_c,j_c) is defined as

$$S_s(i_c,j_c)=\left\{\vec{I}(i,j) \middle| i_c \le i < i_c+s, j_c \le j < j_c+s\right\}$$

Equation 2

A grid with segment size s, topleft corner (i_g,j_g), height N_g and width M_g is defined as

$$G_s(i_g,j_g,N_g,M_g)=\bigcup_{\substack{1\le k_i\le N_g \\ 1\le k_j\le M_g}} S_s(i_g+(k_i-1)s, j_g+(k_j-1)s)$$

Equation 3

The bottomright point is (ig+Ng.s-1,jg+Mg.s-1)
The centered grid is defined as the grid with maximal size that fits in the image and that is maximally centered.

$$G_S^c=G_s(i_g,j_g,N_g,M_g)$$

with

$$(1,1)\le(i_g,j_g)<(s,s) \qquad (a)$$

$$(n,m)-(s,s)<(i_g+N_g s-1, j_g+M_g s-1)\le(n,m) \qquad (b)$$

$$(0,0)\le\left((i_g,j_g)-(1,1)\right)-\left((n,m)-(i_g+N_g s-1, j_g+M_g s-1)\right)\le(1,1) \qquad (c)$$

Equation 4

Equation 4-a means that the top and left border are smaller than s. Equation 4-b means that the bottom and right borders are smaller than s. Equation 4-c means that the grid is maximally centered in the image.
A top down Bayesian analysis produces an output image with height n, width m and q bands

$$\vec{O}(i,j)=\begin{bmatrix} o_1(i,j) \\ \vdots \\ o_q(i,j) \end{bmatrix} \quad \begin{matrix} i=1,\ldots,n \\ j=1,\ldots,m \end{matrix} \qquad \text{Equation 5}$$

by analyzing the input image iteratively on centered square grids

$$G_{s_r}^c=G_{s_r}(i_{g,r},j_{g,r},N_{g,r},M_{g,r}) \qquad \text{Equation 6}$$

with iteratively halved segment sizes

$$s_r=2^{R-r} \quad r=0,\ldots,R \qquad \text{Equation 7}$$

r is the iteration counter, $s_R=2^R$ is the start segment size, the end segment size is 1, the number of iterations is R+1.
For every iteration an intermediate output image with height Ng,r, width Mg,r and q bands

$$\vec{O}_r(i,j)=\begin{bmatrix} o_1(i,j) \\ \vdots \\ o_q(i,j) \end{bmatrix} \quad \begin{matrix} i=1,\ldots,N_{g,r} \\ j=1,\ldots,M_{g,r} \end{matrix} \qquad \text{Equation 8}$$

is generated where the vector $\vec{O}_r(i_r,j_r)$ contains the analysis for the segment

$$S_{s_r}(i_{g,r}+(i-1)s_r, j_{g,r}+(j-1)s_r)=S_{s_r}(i_c(i),j_c(j))$$

Equation 9

The iteration starts (r=0) by taking the maximum likelihood estimation,

$$\vec{O}_0(i,j)=\underset{\vec{O}_0}{\text{argmax}}\ P(S_{s0}(i_c(i),j_c(j))\|\vec{O}_0) \quad \begin{matrix} i=1,\ldots,N_{g,0} \\ j=1,\ldots,M_{g,0} \end{matrix}$$

Equation 10

i.e. the parameter vector $\vec{O}$ for which the realization of the segment is the most probable, as analysis for the segment.
For every consecutive (r>0) iteration, this analysis is refined by taking the Bayesian estimation,

$$\vec{O}_r(i,j)=\underset{\vec{O}_r}{\text{argmax}}\ P\left(\vec{O}_r \| S_{s_r}(i_c(i),j_c(j))\right)$$

$$=\underset{\vec{O}_r}{\text{argmax}}\ P(S_{s_r}(i_c(i),j_c(j))\|\vec{O}_r)P(\vec{O}_r) \quad \begin{matrix} i=1,\ldots,N_{g,r} \\ j=1,\ldots,M_{g,r} \end{matrix}$$

Equation 11

i.e. the most probable parameter vector for the segment $S_{sr}(i_c(i),j_c(j))$ is taken as analysis result vector for the segment.

The Bayesian estimations $\vec{O}_r(i,j)$, obtained by maximizing the product of the a posteriori probability $P(S_{sr}(i_c(i),j_c(j))\|\vec{O}r)$ and the a priori probability $P(\vec{O}r)$ are a refinement of the previously obtained $\vec{O}_{r-1}(i,j)$. Let $\vec{O}_{r,p}(i,j)$ be the linear interpolated version of $\vec{O}_{r-1}(i,j)$,

$$\vec{O}_{r-1}(i,j) \quad \begin{matrix} i=1,\ldots,N_{g,r-1} \\ j=1,\ldots,M_{g,r-1} \end{matrix} \Rightarrow_{\textit{int erpolation}} \vec{O}_{r,p}(i,j) \quad \begin{matrix} i=1,\ldots,N_{g,r} \\ j=1,\ldots,M_{g,r} \end{matrix}$$

Equation 12

then

$$P(\vec{O}_r(i,j))=P\left(\vec{O}_r(i,j)\|\vec{O}_{r,p}(i,j)\right) \quad \begin{matrix} i=1,\ldots,N_{g,r} \\ j=1,\ldots,M_{g,r} \end{matrix}$$

Equation 13

At the last iteration (r=R) the segment size becomes one ($s_R=1$), every individual pixel becomes a segment (Ng,R=n;Mg,R=m) and the intermediate output becomes the final output.

$$\vec{O}(i,j)=\vec{O}_R(i,j) \quad \begin{matrix} i=1,\ldots,n \\ j=1,\ldots,m \end{matrix} \qquad \text{Equation 14}$$

This concludes the definition of the general top down Bayesian image analysis. A specific image analysis is obtained by describing explicitly the content of the input and output band images, and by making explicit assumptions about the distributions $P(Ssr()\|\vec{O}r)$ and $P(\vec{O}r\|\vec{O}r,p)$. This will be done in the next section for the specific cases of interpolation and motion estimation.

4. EXAMPLES OF SPECIFIC IMAGE ANALYSES

4.1 Non-smoothing interpolation

In the case of non-smoothing interpolation we consider a single band input image I(i,j). The pixels are classified in q classes.

$$\text{class of pixel } (i,j) = C(i,j) \in \{1,\ldots,q\} \qquad \text{Equation 15}$$

There are known and unknown pixel values indicated by a mask image M(i,j)

$$M(i,j)=0 \Leftrightarrow I(i,j) \text{ unknown} \qquad i=1,\ldots,n$$
$$M(i,j)=1 \Leftrightarrow I(i,j) \text{ known} \qquad j=1,\ldots,m \qquad \text{Equation 16}$$

As output we want the interpolated pixel values for all classes for all pixels

$$\vec{O}(i,j)=\begin{bmatrix} o_1(i,j) \\ \vdots \\ o_c(i,j) \\ \vdots \\ o_q(i,j) \end{bmatrix} \quad o_c(i,j)=\text{interpolated value for class c for pixel (i,j)}$$

Equation 17

with the restriction that for a known pixel, the interpolated value for the class of the pixel must be exactly (non-smoothing) equal to the known pixel value.

$$I(i,j) = o_{C(i,j)}(i,j) \quad \text{if} \quad M(i,j) = 1$$

Equation 18

This can be solved by a top down analysis where

$$\vec{O}_r(i,j)=\begin{bmatrix} o_1(i,j) \\ \vdots \\ o_c(i,j) \\ \vdots \\ o_q(i,j) \end{bmatrix} \quad \begin{matrix} i=1,\ldots,N_{g,r} \\ j=1,\ldots,M_{g,r} \end{matrix}$$

Equation 19

contains the estimated mean grey values per class and per segment $S_{sr}(i_c(i),j_c(j))$.

We define S_c as the known pixels of class c of segment S

$$S_c=\left\{\vec{I}(i,j) \middle| M(i,j)=1, C(i,j)=c, \vec{I}(i,j)\in S\right\} \qquad \text{Equation 20}$$

We suppose that the realization of the segment pixel values for the different classes are statistically independent

$$P(S\|O)=\prod_{c\in\{1,\ldots,q\}} P(S_c\|o_c) \qquad \text{Equation 21}$$

and that the different pixel values of one class are statistically independent realizations of the same Gaussian distribution

$$P(S_c\|o_c)=\prod_{I(i,j)\in S_c}\frac{1}{\sqrt{2\pi\sigma_{c,k}^2}}\,e^{\frac{-(I(i,j)-o_c)^2}{2\sigma_{c,k}^2}}$$

Equation 22

in which $\sigma^2_{c,k}$ is the variance calculated from the k pixel values in S_c.

Equation 10 becomes

$$o_c=\underset{o_c}{\operatorname{argmin}} \sum_{I(i,j)\in S_c}(I(i,j)-o_c)^2 = \frac{1}{k}\sum_{I(i,j)\in S_c} I(i,j) \equiv o_{c,m} \qquad c=1,..,q$$

Equation 23

in which k is the number of pixel values in S_c. The optimal estimation is the measured mean value $o_{c,m}$ of the known pixels of the class c over the segment.

For the a priori probability Equation 13 we also assume stochastic independence between classes

$$P(O\|O_p)=\prod_{c\in\{1,\ldots,q\}} P(o_c\|o_{c,p}) \qquad \text{Equation 24}$$

(in which $o_{c,p}$ is defined as in Equation 12) and Gaussian distributions

$$P(o_c\|o_{c,p})=\frac{1}{\sqrt{2\pi\sigma_{c,p}^2}}\,e^{\frac{-(o_c-o_{c,p})^2}{2\sigma_{c,p}^2}}$$

Equation 25

$\sigma_{c,p}$ is estimated as the difference between the prior estimation for the mean $o_{c,p}$ and the measured mean $o_{c,m}$

$$\sigma_{c,p}=\left|o_{c,p}-o_{c,m}\right| \qquad \text{Equation 26}$$

Using Equation 22 and Equation 25, Equation 11 combined with Equation 13 becomes

$$o_c=\underset{o_c}{\operatorname{argmin}}\left(\sum_{I(i,j)\in S_c}(I(i,j)-o_c)^2+\frac{\sigma_{c,k}^2}{\sigma_{c,p}^2}(o_c-o_{c,p})^2\right) = \frac{o_{c,m}+\lambda o_{c,p}}{1+\lambda} \text{ with } \lambda=\frac{k\sigma_{c,k}^2}{\sigma_{c,p}^2}$$

Equation 27

$\sigma_{c,k}$ is the maximum likelihood estimation of the standard deviation of the known pixels for class c

$$\sigma_{c,k}=\sqrt{\frac{1}{k}\sum_{I(i,j)\in S_c}(I(i,j)-o_{c,m})^2} \qquad c=1,..,q$$

Equation 28

We get in particular

$$k=1 \Rightarrow \sigma_{c,k}=0, \lambda=0, o_c=o_{c,m} \qquad \text{Equation 29}$$

which ensures that Equation 18 is valid.

4.2 Smoothing interpolation

Here we present a smoothing variant of the non smoothing interpolation in 4.1. This smoothing variant is useful for the case where the known pixel values are contaminated by noise. In these cases it is interesting to replace known pixel values by some smoother varying interpolated values, in order to average out the noise. Therefore we drop Equation 18 and replace Equation 27 by

$$o_c=\frac{o_{c,m}+\lambda o_{c,p}}{1+\lambda} \text{ if } k\geq k_{min} \qquad c=1,..,q$$
$$=o_{c,p} \text{ if } k<k_{min}$$

Equation 30

i.e. we introduce a minimum number k_{min} of measurements needed to perform the update of Equation 27. If the number of measurements is smaller than this minimum number, we suppose that the measured mean is too much contaminated by noise and we prefer not to use it and stop the refinement of the mean estimation.

4.3 Motion estimation

In the case of motion estimation we have given a two band input image

$$\vec{I}(i,j)=\begin{bmatrix} i_1(i,j) \\ i_2(i,j) \end{bmatrix} \quad \begin{matrix} i=1,...,n \\ j=1,...,m \end{matrix} \qquad \text{Equation 31}$$

containing image values $i_1(i,j)$ at time t_1 and image values $i_2(i,j)$ at time t_2. For a given time t between t_1 and t_2

$$t=(1-f)t_1+ft_2 \quad 0\le f\le 1 \qquad \text{Equation 32}$$

we want to obtain the two band output image

$$\vec{O}(i,j)=\begin{bmatrix} o_1(i,j) \\ o_2(i,j) \end{bmatrix} \quad \begin{matrix} i=1,...,n \\ j=1,...,m \end{matrix} \qquad \text{Equation 33}$$

where $o_1(i,j)$ is the i-component (top to bottom) of the displacement and $o_2(i,j)$ is the j-component (left to right) component of the displacement from i_1 to i_2 measured at time t and at location (i,j).
This problem can be solved by a top down Bayesian analysis where

$$\vec{O}_r(i,j)=\begin{bmatrix} o_1(i,j) \\ o_2(i,j) \end{bmatrix} \quad \begin{matrix} i=1,...,N_{g,r} \\ j=1,...,M_{g,r} \end{matrix} \qquad \text{Equation 34}$$

contains the estimated mean motion components for segment $S_{sr}(i_c(i),j_c(j))$ at time t.
We suppose that for every segment Ssr at time t the image deformation from t1 to t2 can be modeled as a rigid motion with addition of white Gaussian noise

$$i_1(i-f\,o_1, j-f\,o_2) = i_2(i+(1-f)\,o_1, j+(1-f)\,o_2)+n(i,j) \qquad \text{Equation 35}$$

with

$$n(i,j) \bullet N(0,\sigma^2_n) \qquad \text{Equation 36}$$

and n(i,j) and n(i',j') are statistically independent if (i,j)≠(i',j') so that

$$P(S|\|(o_1,o_2)) = \prod_{(i,j)\in S} P_n\left(i_1(i-f\,o_1,j-f\,o_2)-i_2(i+(1-f)o_1,j+(1-f)o_2)\right)$$

$$= \prod_{(i,j)\in S} \frac{1}{\sqrt{2\pi\sigma_n^2}} e^{\frac{-\left(i_1(i-f\,o_1,j-f\,o_2)-i_2(i+(1-f)o_1,j+(1-f)o_2)\right)^2}{2\sigma_n^2}} \qquad \text{Equation 37}$$

where P_n is the Gaussian noise distribution.
Equation 10 becomes

$$(o_1,o_2) = \underset{(o_1,o_2)}{\text{argmin}} \sum_{(i,j)\in S} \left(i_1(i-f\,o_1,j-f\,o_2)-i_2(i+(1-f)o_1,j+(1-f)o_2)\right)^2$$
$$\equiv \underset{(o_1,o_2)}{\text{argmin}}\, D_S(o_1,o_2) \qquad \text{Equation 38}$$

with $D_S(o_1,o_2)$ the Euclidean distance between i_1 and i_2 for segment s at time t. This theoretical deduction of Euclidean distance matching agrees well with the experimental finding that among various matching techniques, the Euclidean distance matching has the best performances [15].
For the prior probability Equation 13 we assume a Gaussian distribution

$$P((o_1,o_2)\|(o_{1,p},o_{2,p})) = \left(\frac{1}{\sqrt{2\pi\sigma_p^2}}\right)^2 e^{\frac{-\left((o_1-o_{1,p})^2+(o_2-o_{2,p})^2\right)}{2\sigma_p^2}} \qquad \text{Equation 39}$$

where $(o_{1,p}, o_{2,p})$ is the a priori estimation for (o_1,o_2)
Equation 11 becomes

$$(o_1,o_2) = \underset{(o_1,o_2)}{\text{argmin}}\left[D_S(o_1,o_2)+\lambda\left((o_1-o_{1,p})^2+(o_2-o_{2,p})^2\right)\right] \qquad \text{Equation 40}$$

with

$$\lambda = \frac{\sigma_n^2}{\sigma_p^2} \qquad \text{Equation 41}$$

We take

$$\sigma_n^2 = \frac{D_S(o_{1,p},o_{2,p})}{k} \qquad \text{Equation 42}$$

with k equal to the number of pixels in a segment. The typical $k.\sigma_n^2$ allowed is of the order of the Euclidean distance for the prior estimation and

$$\sigma_p^2 = \frac{2(s_r/2)^2}{k} \qquad \text{Equation 43}$$

i.e. the typical motion refinement allowed is of the order of half the segment size.

5. EXAMPLES OF APPLICATION OF SPECIFIC ANALYSES

5.1 Wavelet compression of an image segment as application of non-smoothing interpolation

As mentioned in 2.3 direct wavelet compression applied to cloud-only images gives bad results. To adapt our wavelet compression method to the presence of strong edges we first code the binary cloud mask (Figure 6) by a lossless compression method. For simplicity we use the standard LZW compression algorithm, although this is not specially adapted to image coding.
We then remove the strong edges in the cloud-only image (Figure 5) by applying the non smoothing interpolation of 4.2, considering the non cloud pixels as unknown cloud pixels. The result is an interpolated texture image, as shown in Figure 7.

Reconstruction of the original image is done by simple multiplication of the cloud mask with the decompressed texture image.
This texture image is compressed by the same wavelet compression algorithm as used in 2.2. Graph 3 of Figure 4 shows the RMS-error versus compressed image file size for wavelet compression of the interpolated texture image. Now compression results comparable to those in [10] are obtained. This shows that the bad compression results of 2.2 were effectively due to the strong edges. The interpolation strongly improves the compression as the useful part of the image is now better compressible thanks to the removal of edges. The number of extra bits needed to code the interpolated part is very low as the interpolated part is very smooth.
The total compressed file size for the original image is the size of the compressed texture image plus the size of the compressed cloud mask. Graph 4 of Figure 4 shows this total file size versus RMS-error for the image in Figure 5. As can be seen, even with the double overhead of coding the useless pixels in the interpolation part of the texture image, and of coding the cloud mask, better

results are obtained than for JPEG for high compression.
At high compression's (leftmost points of graphs 2 and 3 of Figure 4) the cloud mask overhead begins to play an important role. For the leftmost points we have an original file size of 262 Kbytes, a compressed texture image file size of 22 Kbytes and an compressed cloud mask size of 8 Kbytes, so a total compressed file size of 31 Kbytes, which means a compression ratio of 8.5 . The RMS-error is 4.4 counts. The reconstructed image is shown in Figure 8. Even better compression results could be obtained when a better adapted compression for the cloud mask would be used (e.g. run length coding) and when useless wavelet coefficients (wavelet coefficients which only affect the interpolated part of the texture image) would not be coded.

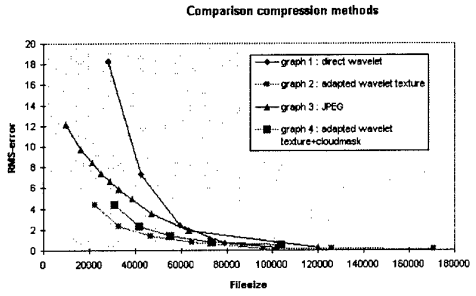


Figure 4 comparison of compression methods

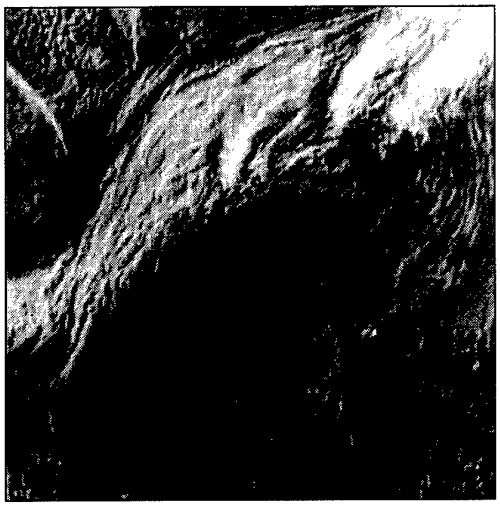


Figure 5 : cloud-only image used as test image for compression methods



Figure 6 : cloud-mask for Figure 5

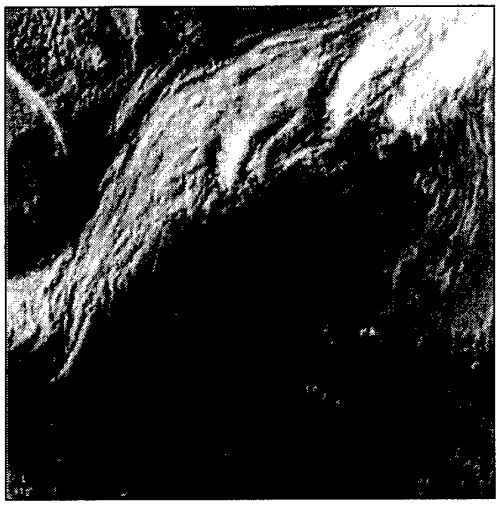


Figure 7 : interpolated texture image for Figure 5

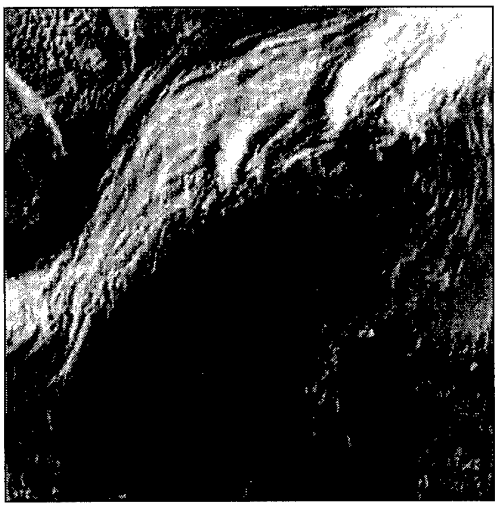


Figure 8 : reconstructed image for Figure 5

5.2 Subpixel interpolation of ERBE images by METEOSAT images as application of a smoothing interpolation

In [16] we decomposed images of the geostationary satellite METEOSAT in clear-sky images with the purpose of completing the ERBE polar satellite measurements with it.
In this section we demonstrate how we can use smoothing interpolation for the combination of METEOSAT clear-sky images and ERBE clear-sky flux measurements. Figure 9 shows the METEOSAT IR clear-sky image for 08/11/86 time slot 30 (14:30-15:00GMT)

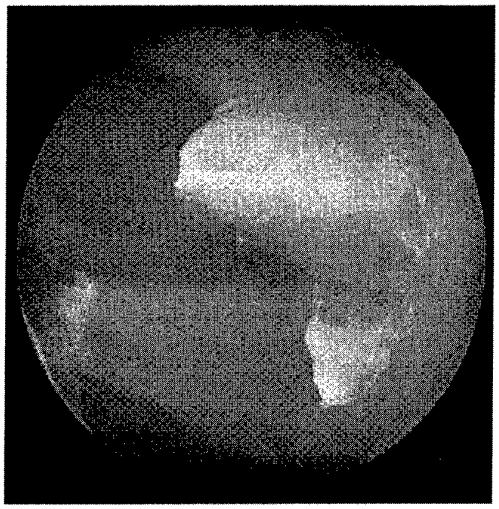


Figure 9 : clear sky METEOSAT IR image for 08/11/86 slot 30 (14:30-15:00GMT)

Figure 10 shows the clear-sky LW flux measurements derived from the ERBE polar satellites collected in METEOSAT format for 08/11/86 during the 3 hour time period centered around METEOSAT time slot 30, i.e. the time period 13:15-16:15 GMT.

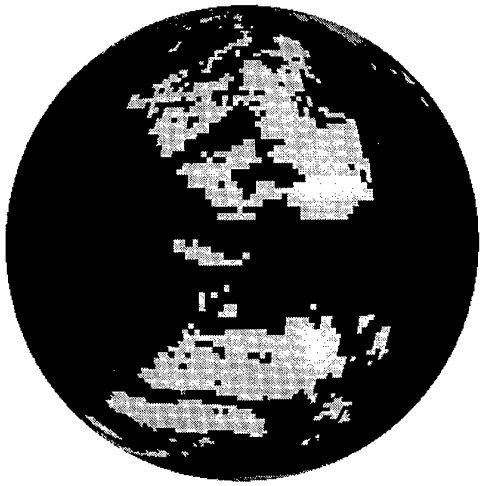


Figure 10 : ERBE longwave flux measurements for 08/11/86 for period 13:15-16:15 GMT

Figure 11 shows the land-ocean mask we will use as auxiliary input.



Figure 11 : land-ocean mask in METEOSAT format used as class image in smoothing interpolation.

The METEOSAT counts can be converted into ERBE flux estimates by combined calibration, narrowband to broadband conversion, and radiance to flux conversion [17],[18]. These three operations can be modeled in first approximation as a multiplication with the respective conversion factors k_{calib}, $k_{nb->bb}$ and $k_{rad->flux}$ so that we obtain a global conversion factor k

$$F_{LW}^{cs} = (k_{rad->flux})(k_{nb->bb})(k_{calib})C_{IR}^{cs} \equiv k\ C_{IR}^{cs}$$

Equation 44

where F_{LW}^{cs} is the clear-sky longwave ERBE flux, C_{IR}^{cs} is the clear-sky METEOSAT count, and k is the global conversion factor from C_{IR}^{cs} to F_{LW}^{cs}.
The global conversion factor k is scene-dependent through the factors $k_{nb->bb}$, which depends on the spectral distribution of the radiation emitted by the scene, and $k_{rad->flux}$, which depends on the angular distribution of this radiation.
For the points for which ERBE clear-sky flux measurements are available (non black points in Figure 10), the conversion factor is estimated as

$$k^{ERBE} = \frac{F_{LW}^{cs,ERBE}}{C_{IR}^{cs}}$$

Equation 45

We interpolate these conversion factors k^{ERBE} using smoothing interpolation with the land-ocean mask of Figure 11 as class image and using k_{min}=1000.

$$k^{ERBE} \xrightarrow[\substack{k_{min}=1000 \\ \text{2 classes: land,ocean}}]{\text{smoothing interpolation}} k^{interpol}$$

Equation 46

Figure 12 shows the image of the obtained interpolated conversion factors $k_{interpol}$.

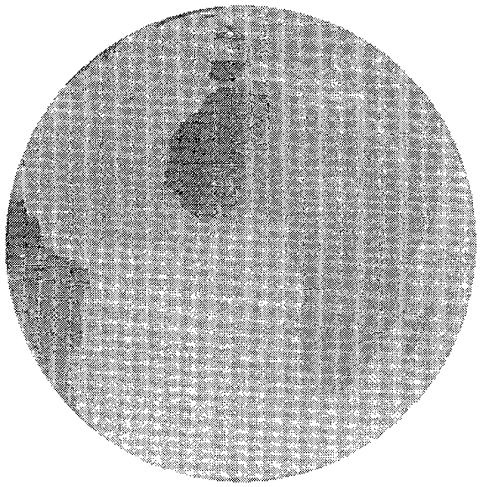


Figure 12 : interpolated METEOSAT IR to ERBE LW conversion factors

We reapply these conversion factors to the METEOSAT counts and obtain the joint ERBE+METEOSAT flux estimation

$$F_{LW}^{cs\,ERBE+METEOSAT} = k^{interpol}\, C_{IR}^{cs}$$

Equation 47

Figure 13 shows the flux estimations obtained from Figure 9 and Figure 12.

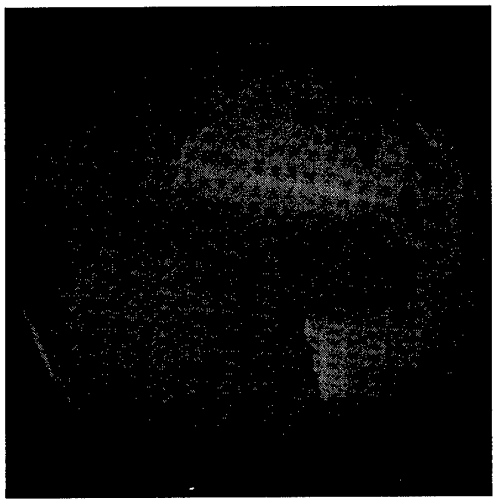


Figure 13 : longwave flux estimation from combined ERBE (Figure 10) and METEOSAT (Figure 9) measurements

This image is a de facto subpixel interpolation of Figure 10 by Figure 9, as its pixel values are close to those of Figure 10 thanks to Equation 45 and as it has the spatial structure of Figure 9 thanks to Equation 47 where $k_{interpol}$ is a smoothly varying factor.

5.3 Motion interpolation of METEOSAT images

In this section we demonstrate how motion estimation can be used for the time interpolation of METEOSAT images.
The METEOSAT IR images over Europe for the time slots 23, 24, and 25 of 04/04/95 are shown in Figure 14, Figure 15 and Figure 16.

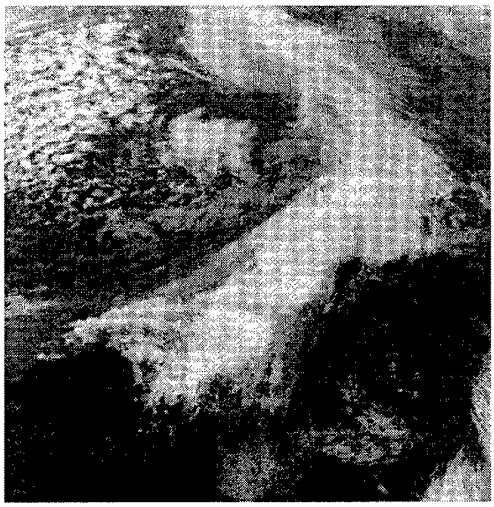


Figure 14 : METEOSAT IR image 04/04/94 slot 23



Figure 15 : METEOSAT IR image 04/04/94 slot 24



Figure 16 : METEOSAT IR image 04/04/94 slot 25

Figure 17 shows the approximation of Figure 15 by simple linear time interpolation between Figure 14 and Figure 16 obtained as

$$C_{24}^{li}(i,j)=0.5(C_{23}(i,j)+C_{25}(i,j)) \qquad \text{Equation 48}$$



Figure 17 : approximation of Figure 15 by simple linear interpolation between Figure 14 and Figure 16.

All the cloud borders which have moved from Figure 14 to Figure 16 are "doubled" in this image, as no motion is taken into account during the interpolation.
Figure 18 shows the error image

$$C_{24}(i,j)-C_{24}^{li}(i,j) \qquad \text{Equation 49}$$

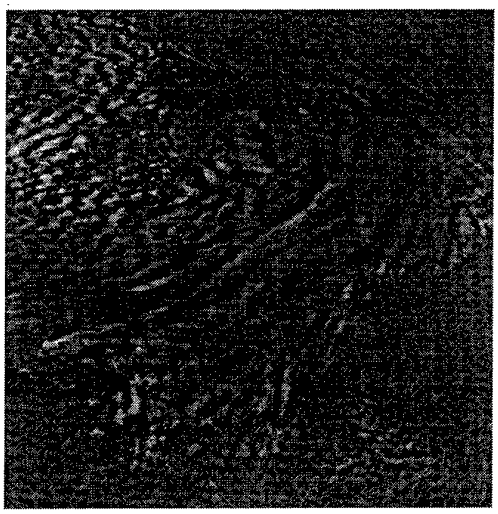


Figure 18 : difference between Figure 15 and Figure 17.

The RMS value of this image is 10.9 counts, i.e. 4.27 % of the full range 255. Typically the highest errors occur at the edges which have moved most. Figure 19 shows the approximation of Figure 15 by motion interpolation between Figure 14 and Figure 16 obtained as

$$C_{24}^{mi}(i,j) = 0.5(C_{23}(i - 0.5m_i(i,j), j - 0.5m_j(i,j)) + C_{25}(i + 0.5m_i(i,j), j + 0.5m_j(i,j)))$$

Equation 50



Figure 19 : approximation of Figure 15 by motion interpolation between Figure 14 and Figure 16.

where m_i and m_j are the i and j direction motion components obtained by applying the top down Bayesian motion estimation of 4.3 to Figure 14 and Figure 16 with f=0.5. The border doubling of Figure 17 does not occur here.
Figure 20 shows the error image

$$C_{24}(i,j)-C_{24}^{mi}(i,j) \qquad \text{Equation 51}$$

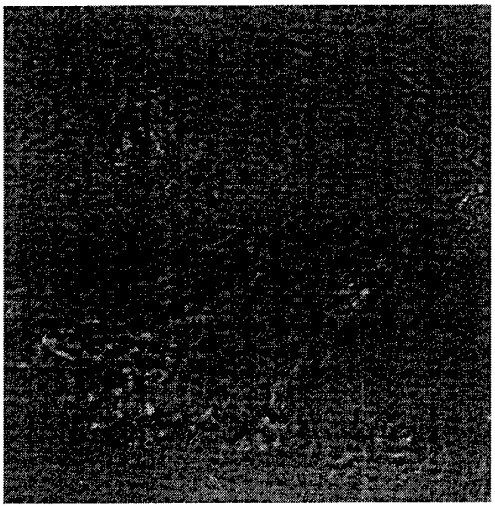


Figure 20 : difference between Figure 15 and Figure 19.

The RMS value of this image is 6.7 counts, i.e. 2.63 % of the full range 255. This means the motion estimation reduces the RMS interpolation error with 39 %. The largest errors remain where the clouds have been deformed too much to stay recognizable, i.e. where the rigid motion hypothesis is violated.

6. CONCLUSION

In this paper we have presented the top down Bayesian analysis method. This method combines the multi-resolution aspects of the wavelet transform and the general applicability of image analysis in windows.
To motivate the conception of this method we started with investigating the wavelet compression results on satellite images. This investigation showed that two kinds of redundancies exist (and are exploited by our wavelet compression method) in natural images: "small scale" variations are smaller then "large scale" variations (this can be called linear redundancy, as it is exploitable by a linear transformation, ideally the Karhunen-Loeve transform, or the wavelet transform in our compression method) and the largest variations are locally concentrated in the most discontinuous parts of the image, typically the borders of the objects in the image, which form a small fraction of the image (this can be called non-linear redundancy as it must be exploited by some other mechanism than a linear transform, e.g. by the zerotree coding in our compression method). Both kind of redundancies favour a top-down analysis.
We defined the top down Bayesian image analysis method mathematically in a general way, i.e. without specifying any specific analysis yet. This stresses the generic nature of the method.
To demonstrate how this unique method can be used for different image processing problems, we worked out its practical application in the three different problems of smoothing interpolation, non-smoothing interpolation and motion estimation. For these three problems we demonstrated a practical application in the processing of METEOSAT images : non-smoothing interpolation was applied in the compression of cloud-only images, smoothing interpolation was used in a ERBE-METEOSAT data fusion for the derivation of clear-sky longwave fluxes with high resolution, and motion estimation was used in the time interpolation of METEOSAT images.

REFERENCES

1 Koenderink, "The structure of images", Biological Cybernetics 50 (1984), pp. 363-370
2 J.G. Daugman, "Complete Discrete 2-D Gabor Transforms by Neural Networks for Image Analysis and Compression", IEEE ASSP, Vol. 36, No. 7, pp. 1169-1179, July 1988
3 S.G. Mallat, "A theory for multi-resolution signal decomposition: the wavelet representation", IEEE Trans. on Pattern Analysis and Machine Intelligence, July 1989, vol. 11 no. 7, pp 674-692
4 M. Vetterli and C. Herley, "Wavelets and filter banks: Relationships and new results", Proc. IEEE ICASSP, Albuquerque, April 1996
5 S.G. Mallat, "Multifrequency Channel Decompositions and Wavelet Models", IEEE ASSP, pp. 2091-2110, december 1989
6 M. Antonini, M. Barlaud, P. Matthieu, I. Daubechies, "Image coding using wavelet transform", IEEE Trans. on Image Processing, December 1993, vol. 1 no. 2,
7 J.M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients", IEEE Trans. on Signal Processing, vol. 41 no. 12, pp. 3445-3462
8 M. Bister, J. Cornelis, A. Rosenfeld, "A Critical View on Pyramid Segmentation Algorithms", Pattern Recognition Letters 11 1990, pp. 605-617, North Holland
9 M. Petrou, N. Fatemi-Ghomi, P.L. Palmer, "Wavelet texture analysis for remote sensing", European Symposium on Satellite Remote Sensing II, SPIE Vol. 2579, pp. 329-340, November 1995
10 J.M. Mangen, S. Grandjean, S. Dewitte, A. Massart, M. Acheroy, J. Cornelis, "Méthode itérative de compression d'images radiométriques par la transformée en ondelettes", First International Symposium "Scientific Imagery and Image Processing", AAAF, April 1995, Cannes, pp. S97-S102
11 S. Mallat, S. Zhong, "Characterisation of signals from multiscale edges", IEEE Trans. on Pattern Analysis and Machine Intelligence, July 1992, vol. 14 no 7, pp. 710-732
12 S. Jaffard, 'Estimation Holderiennes ponctuelles des functions au moyen des coefficients d'ondelettes', Comptes Rendus a l'Academie des Sciences, t. 308, serie I, 79 (1989)
13 J.-M. Mangen, M. Acheroy, "Meteosat image compression using a neural-like architecture", ESA contract report 10031/92/D/IM, July 95.

14 J.-M. Mangen, X. Neyt, A. Massart, M. Acheroy, "CD-ROM archiving of one year of Meteosat data", Image'COM 96, Bordeaux (France).
15 S. Dewitte, M. Vemis, M. Acheroy, J. Cornelis, "Comparison of Matching Techniques for Cloud Motion Wind extraction", proceedings of the First International Symposium on Scientific Imagery and Image Processing, organized by L'Association Aeronautique et Astronautique de France, Cannes, 1995, pp. S40-S60
16 S. Dewitte, E. Nyssen, D. Crommelynck, J. Cornelis , "Multi stage analysis of METEOSAT images", European Symposium on Satellite Remote Sensing II, SPIE Vol. 2579, pp. 170-181, November 1995
17 J. Schmetz, Q. Liu, "Outgoing Longwave Radiation and Its Diurnal Variation at Regional Scales Derived From Meteosat", Journal of Geophys. Research, Vol. 93, No. D9, pp. 11192-11204, September 1988
18 P. Minnis, E.F. Harrison, "Diurnal Variability of Regional Cloud and Clear-Sky Radiative Parameters Derived from GOES data. Part III. November 1978 Radiative Parameters", Journal of Climate and Applied Meteorology, Vol. 23, pp. 1032-1051

Paper 25

F. Lanzl (GE)

How did you get your clear-sky images?

Author's reply:

METEOSAT images form a time series of images taken over different days of the month at exactly the same time of the day. Visible clear-sky imges are obtained by taking for every pixel the minimum value of these time series.

CLOSED FORM EXPRESSIONS OF ALGORITHMIC PERFORMANCE AND COMPLEXITY IN SIGNAL PROCESSING

by

André MARGUINAUD and Pierre BERTRAND

ALCATEL ESPACE
5, rue Noël Pons
92737 Nanterre
France

SUMMARY

In spite of the price deflation of computing power and available simulation tools, closed-form bounds are necessary for system designers to check overall simulation results and to decrease costs.

The mathematical concepts used to optimise the algorithms and their associated bounds are : analytic signal representation, the principle of Maximum Likelihood, Neaman Pearson formulation of requirements, Chernoff bounding, quantification optimisation at each stage of the processing, weighted least squares, Lagrange multipliers, statistical generating functions, etc ...

The application of the above methods by means of analytical expressions of performance and graphics covers the following items encountered essentially in transmission : non linearity impact on system assessment, MFSK and MPSK modulation , soft decoding of convolutional codes , combination of inner and outer codes.

1. INTRODUCTION

The ever increasing complexity of planed systems places a heavy burden on performance verification : exact theory is practically impossible with or without formal computing assistance, and on the other hand , digital simulation is far from being exhaustive and is subject to undetected numerical errors. One way to cope with the above difficulties is to organize the verification task as a set of separated processes, the errors of which are clearly identified in nature and such that their likelihood of occurrence are decorrelated.

Conceptually, verification optimisation is a minimax problem : how to have the best security of assessment at the lowest cost. The general trend consists in deriving an exact theory of all elementary processes in order to build complete theory of the whole system. In practice such theoritical expressions are simply impossible or inextricable so that it is necessary to to use approximations. In order to cope with the resulting uncertainty, it seems appropriate to simulate the smallest feature of the system under study. Very often, an explosion of the number of the elementary operations necessary to obtain a reliable estimation of the mean performance and of its distribution is observed (*cf. centred variances of the main operational parameters*).

A more effective procedure is to exploit the fact that the environmental conditions and the operational requirements are defined up to a limited precision by looking for bounding expressions using simple usual functions : polynomials and exponentials.

Of course the results obtained by such expressions should be confirmed by a few realistic simulation runs in order to detect any material error.

In the next paragraph, we present the general principles and results useful to handle actual systems, then the remaining paragraphs are practical illustrations of this way of thinking.

2. PRESENTATION AND JUSTIFICATION OF ELEMENTARY TOOLS

1) Analytic representation of real signals

We consider only locally continuous and band limited signals to be processed by time sampling techniques. Since all digital processes are finite in time, local continuity leads to uniform continuity. On the other hand, a signal s(t) real or complex is limited to a band ΔB if it can be locally obtained by taking the inverse Fourier transform of a function null outside a compact finite frequency interval of width ΔB.

As explained in [4] these two conditions lead to the following fundamental result :

- the signal s(t) can be locally approximated with a uniform precision either by a trigonometric polynomial, the frequencies of which lie in the above frequency interval ΔB, or by a sequence of complex samples spaced in time by a constant step $\Delta t = \frac{1}{\Delta B}$.

This result is useful most of the time as an existence statement, rather than as an effective procedure.

A pertinent application of this way of thinking is the explicit determination of the optimum weighing which transforms a real sampling into an analytic sampling (cf. [4]) . To optimise or characterise a linear algorithm, it is sufficient to consider a single properly parametrized sine, where often the result can be stated explicitly.

Another significant consequence of the uniform approximation by an exponential polynomial is the fact that the 2 components should have the same amplitude histogram following the law of histogram composition.

2) Probabilistic characterisation of the noise in quadrature

At this step , the received real signal results from the summation of a deterministic (*or useful information*) signal and an additive noise n(t) function of time t. n(t) is considered decorrelated from the deterministic signal. The compulsary analog (*band pass*) filtering of the composite signal before digitizing leads to the result that n(t) can be considered as continuous and band limited.

Paper presented at the AGARD SPP Symposium on "Remote Sensing: A Valuable Source of Information", held in Toulouse, France from 22-25 April 1996, and published in CP-582.

Let Δt be the time step sampling , n(t) is known only by its mathematical expectation <n(t)> = 0 and its variance $\langle n(t)\, n(t+k\,\Delta t)\rangle = \sigma^2\,\delta_{k0}$. The last relation implies that Δt is of the order of the time decorrelation.

The (*optimum*) weighting derived in [4] in order to evaluate the 2 components of the analytic representation x(t), y(t) has the following remarkable structure :

$$(1)\quad \begin{cases} x(t) = \sum\limits_{k\geq 0} (-1)^k\, C_k \left[n\left(t - \dfrac{(2k+1)\,\Delta t}{2}\right) + n\left(t + \dfrac{(2k+1)\,\Delta t}{2}\right)\right] \\ y(t) = \sum\limits_{k\geq 0} C_k \left[n\left(t - \dfrac{(2k+1)\,\Delta t}{2}\right) - n\left(t + \dfrac{(2k+1)\,\Delta t}{2}\right)\right] \end{cases}$$

C_k are real parameters. It must be noted that the numerical values of the C_k are irrelevant for deriving the pertinent statistical properties of the « **analytical** » noise.

$$(2)\quad \begin{cases} \langle x(t)\rangle = \langle y(t)\rangle = 0 \\ \langle x(t)\, x(t+\ell\,\Delta t)\rangle = \langle y(t)\, y(t+\ell\,\Delta t)\rangle = 2\,\sigma^2 \sum\limits_{k\geq 0} (C_k)^2\,\delta_{\ell 0} \\ \langle x(t)\, y(t+\ell\,\Delta t)\rangle = 0 \end{cases}$$

Relation (2) shows that the noise in quadrature is « **statistically independent** » of the real noise and has the same variance. The signal to noise ratio is preserved only if the C_k values are properly chosen.

3) Central limit theorem (CLT)

This theorem has been established in the case of repeated independent measurements by de Moivre in the 17th century ; under wider conditions it has been extended by Lindbergh to any finite sum of statistical variables characterised only by their mean and (*centred*) variance.

The first step is to prove that the mean (*resp. variance*) of the sum of independent variables equals the sum of the means (*resp. variances*) of the elementary variables as stated by (3).

(3)

$$\begin{cases} \langle x\rangle = \left\langle \sum\limits_i x_i \right\rangle = \sum\limits_i \langle x_i\rangle \overset{\Delta}{=} \sum\limits_i m_i \\ \left\langle (x - \langle x\rangle)^2 \right\rangle = \sum\limits_i \left\langle x_i - \langle x_i\rangle^2 \right\rangle = \sum\limits_i \langle x_i^2\rangle - \sum\limits_i \langle\langle x_i\rangle\rangle^2 \overset{\Delta}{=} \sum\limits_i v_i \end{cases}$$

In the case of discrete variables, (3) is easily derived using generating function formalism.

To each elementary variable x_i is associated the real function $G_i(z) \overset{\Delta}{=} \sum\limits_a p(a)\, z^a \overset{\Delta}{=} \exp E_i(z)$ in which p(a) is the probability to obtain the amplitude a in its definition domain.

(4)

$$\begin{cases} G_i \overset{\Delta}{=} G_i(1) = \sum\limits_a p(a) = 1 \\ \qquad \Rightarrow E_i \overset{\Delta}{=} E_i(1) = 0\,,\ E_i' \overset{\Delta}{=} E_i'(1)\,,\ E_i'' \overset{\Delta}{=} E_i''(1) \\ G_i' \overset{\Delta}{=} \left.\dfrac{d\,G_i(z)}{d\,z}\right|_{z=1} = \sum\limits_a a\,p(a)\,z^{a-1}\Big|_{z=1} \overset{\Delta}{=} m_i \\ \qquad \Rightarrow m_i = E_i'(z)G_i(z)\Big|_{z=1} = E_i' \\ G_i'' \overset{\Delta}{=} \dfrac{d}{dz}\left[z\,G'(z)\right]\Big|_{z=1} = \sum\limits_a a^2\,p(a)\,z^{a-1}\Big|_{z=1} = \langle x_i^2\rangle \\ \qquad \Rightarrow v_i = \dfrac{d}{dz}\left[z\,E_i'(z)\,G_i(z)\right]\Big|_{z=1} - (E_i')^2 = E_i' + E_i'' \end{cases}$$

In order to prove (3) from (4) , it remains to show that the generating function of the variable defined as the sum of elementary variables equals the product of their generating function as shown below.

$$\exp\left[E_i(z)+E_j(z)\right] = G_i(z)\,G_j(z) = \left(\sum_a p(a)\,z^a\right)\left(\sum_b q(b)\,z^b\right) = \sum_{a,b} p(a)\,q(b)\,z^{a+b} = \sum_{a,b} p(a,b)\,z^{a+b}$$

The CLT asserts that the summation of variable with given mean and variance converges to a Gaussian distribution as their number increases. For practical applications encountered in signal processing, this convergence is faster than expected.

4) Bounding of discrete distributions

This technique exploits the fact that statistical distributions falls off exponentially for large amplitudes.

Considering a statistical distribution defined by its generating function G(z) = exp E(z), we are interested in deriving an efficient upper bounding of the probability P_L that its random variable be equal or superior to the level L.

$$P_L \overset{\Delta}{=} \sum_{a\geq L} p(a) \;\Rightarrow\; z^L\,P_L < \sum_{a\geq L} p(a)\,z^a \quad \text{when } z>1$$

$$z^L\,P_L < \sum_a p(a)\,z^a = G(z) = \exp E(z)$$

From the above series of inequality we derive the desired upper bound (5).

(5)

$$\mathrm{Ln}\left(P_L\right) < \min_{z>1}\left[E(z) - L\,\mathrm{Ln}\,(z)\right] < \min_{z>1}\left[E(z) - L(z-1)\left(1 - \frac{z-1}{2}\right)\right]$$

In order to connect this bound with the CLT presented above, we consider the Taylor expansion of E(z) about z=1.

$$E(z) = (z-1)m + \frac{(z-1)^2}{2}\,v + \frac{(z-1)^3}{6}\,E'''$$

When $\frac{(z-1)^3}{6}$ E''' is negligible, the lowest upper bound for P_L is obtain for $z-1=\frac{L-m}{L+v}$ and (5) takes the form below where E''' is now maximum of E'''(z) for $1<z<\frac{2L+v-m}{L+v}$.

(6) $$P_L < \exp\left[-\frac{(L-m)^2}{2(L+v)} + \frac{1}{6}\left(\frac{L-m}{L+v}\right)^3 E'''\right]$$

with $L > m > 0$.

m, v and E'''are additive functions of the statistical variables, so that (6) can be used to check quantitatively (*by progressive elimination of the* E''' *term*) the Gaussian tendency as we sum elementary variables.

This general principle is used in all the following paragraphs presenting calculation techniques and models encountered in usual signal processing operations such as channel characterization (*non linerarities, phase noise, quantification*) and signal parameters estimation (*carrier frequency, phase, amplitude, signal to noise ratio*): performances and degradations are expressed in terms of the Gaussian noise power density resulting in the frequency domain : $\sigma^2 = \Delta N_0 B$ where B represents the input signal bandwidth. Thus, the system performances taking several consecutive models into account are derived from the resulting Gaussian statistical distribution.

For a Gaussian variable of mean m and variance v, it is possible to derive a lower bound and an upper bound (7) which have the same structure as (6).

(7)

$$\frac{1}{\sqrt{8\pi\frac{(L-m)^2}{v}}}\exp-\frac{(L-m)^2}{2v} < P_L = \frac{1}{\sqrt{2\pi v}}\int_L^\infty\left[\exp-\frac{(x-m)^2}{2v}\right]dx$$

$$P_L < \frac{1}{\sqrt{2\pi\frac{(L-m)^2}{v}}}\exp-\frac{(L-m)^2}{2v}$$

The left-hand inequality is established for $L-m>\sqrt{v}$ and is obtained by evaluating the area under the tangent of the density distribution at the abscissa L.

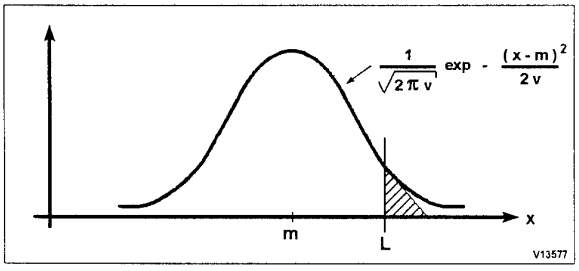


Figure 1 - Bounding of a Gaussian distribution

The second inequality valid for L - m > 0 is obtained by replacing dx by $\frac{x-m}{L-m}dx$ (> dx) and integrating.

$$\frac{1}{\sqrt{2\pi v}}\int_L^\infty\left[\exp-\frac{(x-m)^2}{2v}\right]dx < \frac{1}{\sqrt{2\pi\frac{(L-m)^2}{v}}}\int_L^\infty\left[\exp-\frac{(x-m)^2}{2v}\right]\frac{(x-m)\,dx}{v}$$

5) Non linearity in amplifiers

The general procedure consists in considering that a (*non linear*) actual amplifier delivers a signal proportional to the input signal plus a noise signal.

The (*complex*) gain and delay of the linear part of the above model are determined by minimising the expected powers σ^2 of the above noise signal. This power is then interpreted as a power density ΔN_0 in the frequency domain : $\sigma^2 = B\Delta N_0$ where B represents the input signal bandwidth.

In order to illustrate the method we consider some instances of practical use.

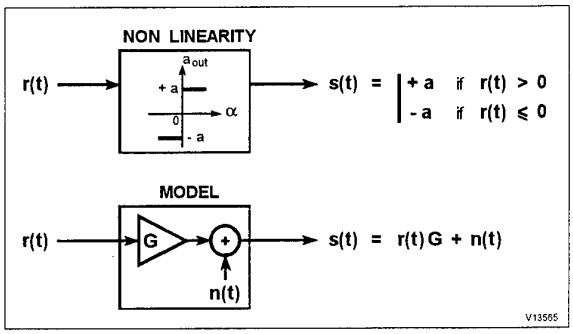


Figure 2 - Hard limiting model

The first situation considered here is the hard limiting of a real signal r(t). The output signal is +a if the input amplitude α is positive, and the output amplitude is -a otherwise. The non-linear noise power is completely defined by the input amplitude distribution density $g(\alpha)$, supposed even. The output noise power is $N = 2\int_0^\infty (a - G\alpha)^2\, g(\alpha)\, d\alpha$ for an output signal S plus noise N power $S+N = a^2$. N is minimum for

$$G = \frac{a\langle|\alpha|\rangle}{\langle\alpha^2\rangle} = \frac{a\int_0^\infty \alpha\, g(\alpha)\, d\alpha}{\int_0^\infty \alpha^2\, g(\alpha)\, d\alpha}.$$

(8) $$N = a^2\left[1-\frac{\langle|\alpha|\rangle^2}{\langle\alpha^2\rangle}\right]$$ and $a^2 = N+S$ implies

$$\frac{S+N}{S} = \frac{\langle\alpha^2\rangle}{\langle|\alpha|\rangle^2} = \frac{a^2}{S}$$

For a Gaussian distribution $\frac{\langle\alpha^2\rangle}{\langle|\alpha|\rangle^2} = \frac{\pi}{2}$ and for the amplitude distribution of a sinewave $\frac{\langle\alpha^2\rangle}{\langle|\alpha|\rangle^2} = \frac{\pi^2}{8}$.

Thus for a Gaussian distribution the amplitude degradation is 2 dB, and for a sinewave this degradation is only 1 dB.

The above result is operationally useful only if the input signal to noise ratio is rather large.

The performance degradation is more severe in the case of the hard limiting of the sum of 2 real signals with different amplitudes since the weak signal undergoes an extra level degradation of several dB as we show next.

As before, we assume the amplitude density $u(\alpha)$ of the lower signal and the amplitude density $f(\beta)$ of the higher signal to be even.

When the lower signal level α is positive, the hard limiter delivers +a when the higher signal level β is greater than $-\alpha$ which occurs with probability $\int_{-\alpha}^{+\infty} f(\beta)\,d\beta$, and -a when the higher signal level β is lower than $-\alpha$ which occurs with probability $\int_{-\infty}^{-\alpha} f(\beta)\,d\beta$. A comparable argument holds when the level α is negative.

For the lower signal the limiting process is characterised by a gain factor G corresponding to the minimum variance of the limiting noise variance N.

$$\frac{1}{2}N = \int_0^\infty (a - G\alpha)^2\,\mu(\alpha)\left[\int_{-\alpha}^{\infty} f(\beta)\,d\beta\right]d\alpha + \int_0^\infty (-a - G\alpha)^2\,\mu(\alpha)\left[\int_{-\infty}^{-\alpha} f(\beta)\,d\beta\right]d\alpha = \frac{a^2}{2} + \frac{G^2\langle\alpha^2\rangle}{2} - 2\,G\,a\int_0^\infty \alpha\,\mu(\alpha)\left[\int_{-\alpha}^{\alpha} f(\beta)\,d\beta\right]d\alpha$$

The above variance is minimised for $G = 2a\,\frac{\int_0^\infty \alpha\,\mu(\alpha)\left[\int_{-\alpha}^{\alpha} (\beta)\,d\beta\right]d\alpha}{\langle\alpha^2\rangle}$.

When $\langle\langle\beta^2\rangle\rangle \gg \langle\langle\alpha^2\rangle\rangle$ we can proceed with the following justified approximations.

(9)

$$\int_0^\infty \alpha\,\mu(\alpha)\left[\int_{-\alpha}^{\alpha} f(\beta)\,d\beta\right]d\alpha = 2\int_0^\infty \alpha^2\,\mu(\alpha)\,f(0)\,d\alpha = \langle\alpha^2\rangle\,f(0)$$

$$N = a^2\left[1 - 4\,f(0)^2\,\langle\alpha^2\rangle\right],\quad G = 2a\,f(0),\quad \frac{G^2\langle\alpha^2\rangle}{N} = \frac{4f(0)^2\,\langle\alpha^2\rangle}{1-4\,f(0)^2\,\langle\alpha^2\rangle}$$

For a Gaussian signal of variance $\langle\beta^2\rangle$, $f(0)^2 = \frac{1}{2\pi\langle\beta^2\rangle}$ so that $\frac{G^2\langle\alpha^2\rangle}{N} \approx \frac{2}{\pi}\frac{\langle\alpha^2\rangle}{\langle\beta^2\rangle}$, which indicates a reduction of 2 dB of the signal to noise ratio.

For a sinewave of amplitude $a = \sqrt{2\langle\beta^2\rangle}$, the amplitude density at $\beta = 0$ is $f(0) = \frac{1}{\pi\sqrt{2}\,\langle\beta^2\rangle}$ so that $\frac{G^2\langle\alpha^2\rangle}{N} \approx \frac{2}{\pi}\frac{\langle\alpha^2\rangle}{\langle\beta^2\rangle}$, which indicates a 5 dB degradation of the signal over noise ratio.

It should be noted, however, that this significant degradation is obtained only for a sinusoidal strong jamming and digitizing with a single bit. The more common situation is a linear amplification with a limitation at a fixed level ± L higher than the variance mean square root.

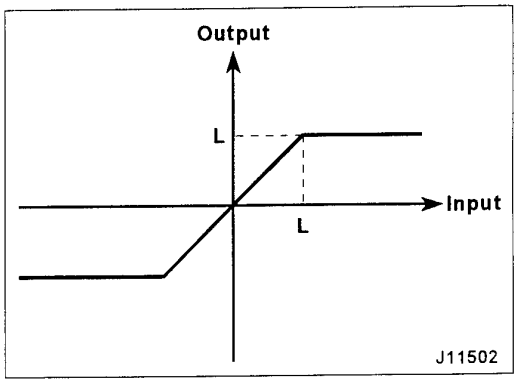


Figure 3 - Gain 1 with hard limiting

In this case, with gain 1 and a Gaussian input signal , one gets an upper bound of the variance $v_L = \langle\Delta\alpha^2\rangle$ of the subtracted signal by relation (10).

$$v_L \triangleq 2\int_L^\infty \frac{(x-L)^2}{\sqrt{2\pi\langle\alpha^2\rangle}}\left[\exp-\frac{x^2}{2\langle\alpha^2\rangle}\right]dx < 2\int_L^\infty \frac{x^2}{\sqrt{2\pi\langle\alpha^2\rangle}}\exp-\frac{x^2}{2\langle\alpha^2\rangle}\,dx$$

$$v_L < -\sqrt{\frac{2}{\pi}}\int_L^\infty x\sqrt{\langle\alpha^2\rangle}\;d\left[\exp-\frac{x^2}{2\langle\alpha^2\rangle}\right]$$

$$< \sqrt{\frac{2}{\pi}}\,L\sqrt{\langle\alpha^2\rangle}\exp-\frac{L^2}{2\langle\alpha^2\rangle} + \sqrt{\frac{2}{\pi}}\int_L^\infty \sqrt{\langle\alpha^2\rangle}\left[\exp-\frac{x^2}{2\langle\alpha^2\rangle}\right]dx$$

$$\int_L^\infty \left[\exp-\frac{x^2}{2\langle\alpha^2\rangle}\right]dx = \int_L^\infty -\frac{\langle\alpha^2\rangle}{x}\,d\left[\exp-\frac{x^2}{2\langle\alpha^2\rangle}\right] < \frac{\langle\alpha^2\rangle}{L}\exp-\frac{L^2}{2\langle\alpha^2\rangle}$$

Collecting partial results we get the upper bound (10) for the limiting noise variance.

(10)

$$v_L = \left\langle (\Delta\alpha)^2 \right\rangle < \sqrt{\frac{2}{\pi}} \left[L \sqrt{\left\langle \alpha^2 \right\rangle} + \frac{\left\langle \alpha^2 \right\rangle \sqrt{\left\langle \alpha^2 \right\rangle}}{L} \right] \exp - \frac{L^2}{2\left\langle \alpha^2 \right\rangle}$$

The signal over limiting noise ratio is thus :

$$\frac{\left\langle \alpha^2 \right\rangle}{\left\langle (\Delta\alpha)^2 \right\rangle} > \sqrt{\frac{\pi}{2}} \left(\frac{L}{\sqrt{\left\langle \alpha^2 \right\rangle}} + \frac{\sqrt{\left\langle \alpha^2 \right\rangle}}{L} \right)^{-1} \exp + \frac{L^2}{2\left\langle \alpha^2 \right\rangle}$$

Soft limitation can be studied using the same approach, but actual power amplifiers introduce a small delay proportional to the « **instantaneous power** » of the signal. This delay is comparable to a dephasing for small relative bandwidth signals. The combination of both phase and amplitude effects of soft limitation in this case speaks of analytical limitation (*Figure 4*). The real signal is considered as the real part of its analytical representation z(t) function of the time t. $z(t) = r(t) \exp j\theta(t)$, r(t) and $\theta(t)$ are real functions of time t. The modulus r(t) of z(t) , which charaterizes the instantaneous power , is given by $r(t) = \sqrt{z(t)\, z^*(t)}$ and determines the complex gain representing the analytical limiting process.

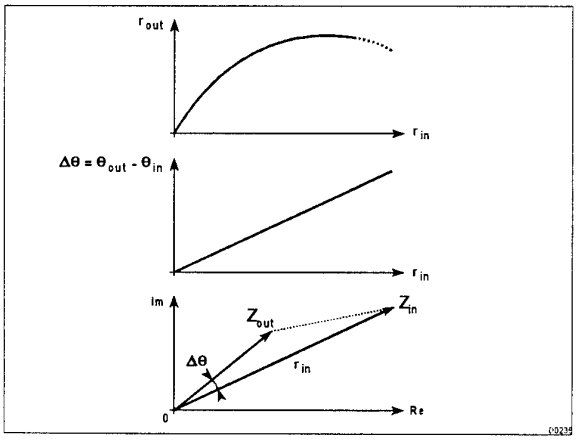


Figure 4 - Analytical soft limitation modeling the effects of actual amplifiers on the phase and amplitude of an input signal

This limiting process transforms z(t) by z(t) g(r(t)) ; to simplify notations we note r instead of r(t) and consider the simplest possible polynomial model for g(r).

$$g(r) = \alpha(1 - \gamma r) \exp j\,\eta\, r \qquad \text{with } 0 \le \gamma r < 1$$

No limiting corresponds to $\gamma = \eta = 0$

The amplifier is characterized by the values of α, γ, and η , and the input signal is characterized by the (*discrete*) distribution p(r) of its complex modulus.

As above we are looking for a constant gain of the general form $G = R \exp j\,\theta$ where R and θ are real constants. The quadratic form to minimise with respect to the unknown R and θ is the following :

$$N \overset{\Delta}{=} \sum_r r^2\, p(r) \left[g(r) - G \right] \left[g^*(r) - G^* \right]$$

$$\frac{\partial N}{\partial G^*} = \left[G \sum_r r^2\, p(r) - \sum_r r^2\, p(r)\, g(r) \right] = 0$$

$$G = \frac{\sum_r r^2\, p(r)\, g(r)}{\left\langle r^2 \right\rangle} \qquad \text{with } g(r) = \alpha(1 - \gamma r) \exp j\eta r$$

The (*analytic*) limiting noise is $N = \sum_r r^2\, P(r)\, g(r)\, g^*(r) \; \dfrac{\left| \sum_r r^2\, p(r)\, g(r) \right|^2}{\left\langle r^2 \right\rangle}$. The signal to noise ratio ρ corresponding to this limiting process is given by (11).

$$\textbf{(11)} \qquad \frac{1}{\rho} = 1 - \frac{\left[\sum_r r^2\, p(r)\, g(r) \right] \left[\sum_r r^2\, p(r)\, g^*(r) \right]}{\left\langle r^2 \right\rangle \left[\sum_r r^2\, p(r)\, g(r)\, g^*(r) \right]}$$

6) Carrier parameters estimation

Man made signals are considered as a succession of time limited reversible transformations of a continuous sinewave of designed amplitude a . This sinewave is by definition the **carrier** . Demodulating consists in finding this sequence of transformations or trying to choose at any time the most likely signal among a finite set of complex patterns up to « **some** » complex scaling (*rotation and real gain*). Each of these patterns is represented by its complex time function $z_h(t)$. Each received signal is of the form $z(t) = G\, z_h(t) + n(t)$ where n(t) represents the additive analytic noise. For band limited signal the complex argument of G can accomodate any small time delay. The maximum likelihood principle states that the preferred hypothesis is such that the error function $Q_h(G)$ is minimum.

$$Q_h \overset{\Delta}{=} \min_G \sum_t \left[z(t) - G\, z_h(t) \right] \left[z^*(t) - G^*\, z_h^*(t) \right]$$

$$\frac{\partial\, Q_h(G)}{\partial\, G^*} = 0 \;\Rightarrow\; G_h = \frac{\sum_t z(t)\, z_h^*(t)}{\sum_t z_h(t)\, z_h^*(t)}$$

$$\Rightarrow\; Q_h = \sum_t z(t)\, z^*(t) - \frac{\left| \sum_t z(t)\, z_h^*(t) \right|^2}{\sum_t z_h(t)\, z_h^*(t)}$$

This last expression proves that the preferred hypothesis has the greatest reconstructed carrier $\dfrac{\left|\sum_t z(t)\, z_h^*(t)\right|^2}{\sum_t z_h(t)\, z_h^*(t)}$.

The process of hypothesis identification leads to a (*regular*) sequence of complex samples of a single sinewave characterised by 3 parameters : phase reference, frequency and amplitude.

The presentation of the corresponding estimators is easier when the complex samples are represented in polar coordinates as below, where h is an integer and Δt is the sampling time step .

(12)

$$z(t) = r(t) \exp j\,\theta(t) \qquad \text{with} \begin{cases} x(t) = r(t)\cos\theta(t) \\ y(t) = r(t)\sin\theta(t) \end{cases}$$

$$= A \exp j(\omega t + \varphi) + n(t)$$

$$\langle n(t) \rangle = 0 \qquad \text{and} \qquad \langle n(t)\, n^*(t + h\,\Delta t) \rangle = \sigma^2\, \delta_{h0}$$

Dropping time t for simplicity, we obtain by differentiation, the expressions of the polar variances from the Cartesian variances.

$$\begin{cases} x = r\cos\theta \\ y = r\sin\theta \end{cases} \Rightarrow \begin{cases} dr = (\cos\theta)\,dx + (\sin\theta)\,dy \\ r\,d\theta = -(\sin\theta)\,dx + (\sin\theta)\,dy \end{cases}$$

(13)

$$\begin{cases} \langle dx \rangle = \langle dy \rangle = \langle dx\,dy \rangle = 0 \quad , \quad \langle (dx)^2 \rangle = \langle (dy)^2 \rangle = \sigma^2 \\ \langle (dr)^2 \rangle = \sigma^2 \ , \ \langle (d\theta)^2 \rangle = \dfrac{\sigma^2}{r^2} \ , \ \langle dr\,d\theta \rangle = 0 \ , \ \rho \overset{\Delta}{=} \dfrac{r^2}{2\sigma^2} = \dfrac{1}{2\langle (d\theta)^2 \rangle} \end{cases}$$

The carrier is considered as a the sum of a complex sinewave and a complex noise (cf. (12)) . In order to estimate ω and φ , we begin by developing the argument $\theta(t)$ defined in the interval $(-\pi, +\pi)$. The developed phase φ_t associated with $\theta(t)$ is such that :

$$\begin{aligned} \varphi_{t+1} - \varphi_t &= \theta(t+1) - \theta(t) && \text{if } |\theta(t+1) - \theta(t)| < \pi \\ &= \theta(t+1) - \theta(t) + 2\pi && \text{if } \theta(t+1) - \theta(t) \le -\pi \\ &= \theta(t+1) - \theta(t) - 2\pi && \text{if } \theta(t+1) - \theta(t) \ge \pi \end{aligned}$$

when the expected frequency is near zero. Removing the expected frequency rotation from φ_t fulfills this condition .

The most likely values of ω and φ minimises the quadratic form $Q = \sum_t \left[\varphi_t - \omega t - \varphi\right]^2$. From now on t runs over the relative integer domain plus $\frac{1}{2}$.

The method of calculation of the developed phase shows that φ_t is unbiased and has the same variance as $\theta(t)$.

The derivation of the estimators expressions is simplified by choosing the time origin such a way that $\sum_t t = 0$.

The expressions of the estimators of ω and φ are given by (14) obtained as the solution of the linear system $\dfrac{\partial Q}{\partial \varphi} = \dfrac{\partial Q}{\partial \omega} = 0$.

$$\textbf{(14)} \sum_t 1 = \ell \ , \ \sum_t t = 0 \ , \ \hat{\varphi} \overset{\Delta}{=} \frac{1}{\ell} \sum_t \varphi_t \ , \ \hat{\omega} \overset{\Delta}{=} \frac{\sum_t t\,\varphi_t}{\sum_t t^2}$$

where ℓ is the (*odd*) number of complex samples used for the estimation. It should be noted that the evaluation of $\sum_t t\,\varphi_t$ requires only 1.5 addition per sample : one calculates $s_i \leftarrow \varphi_i - \varphi_{-i}$, then $S_k \leftarrow S_{k-1} + s_k$ and finally $S = \sum_1^{[\ell/2]} S_k$.

In order to get the statistical behaviour of the estimators, we consider that the exact developed phase $\bar{\varphi}_t = \omega t + \varphi$ is related to its « **measured** » value φ_t by

$$\varphi_t = \bar{\varphi}_t + v(t) \text{ with } \langle v(t) \rangle = 0 \text{ and } \langle v(t)\, v(t+h) \rangle = \frac{\sigma^2}{r^2}\, \delta_{0h}$$

$$\langle \hat{\varphi} \rangle = \frac{1}{\ell} \left\langle \sum_t \left(\bar{\varphi}_t + v(t)\right) \right\rangle = \frac{1}{\ell} \left\langle \sum_t \left(\omega t + \varphi + v(t)\right) \right\rangle = \omega$$

$$\langle \hat{\omega} \rangle = \frac{1}{\sum_t t^2} \left\langle \sum_t t\,(\omega t + \varphi + v(t)) \right\rangle = \omega$$

These estimators are unbiased, the evaluation of their variance is obtained in the same way. For operational purposes it is useful to introduce the integration time T and the noise bandwidth B so that we obtain the following relations.

(15)

$$\begin{cases} \ell = BT \quad , \quad \sigma^2 = N_0 B \quad , \quad \sum_t t^2 = \dfrac{T^3 B}{12} \\ \left\langle (\hat{\varphi} - \varphi)^2 \right\rangle = \dfrac{\sigma^2}{\ell\, r^2} = \dfrac{N_0}{r^2\, T} \ , \ \left\langle (\hat{\omega} - \omega)^2 \right\rangle = \dfrac{\sigma^2}{r^2} \dfrac{1}{\sum_t t^2} = \dfrac{12\, N_0}{r^2\, T^3} \ , \ \left\langle (\hat{\varphi} - \varphi)(\hat{\omega} - \omega) \right\rangle = 0 \end{cases}$$

The amplitude r is obtained by minimizing the quadratic form $Q \overset{\Delta}{=} \sum_t \left(z(t) - r \exp j\,\varphi_t\right)\left(z^*(t) - r \exp - j\,\varphi_t\right)$.

Differentiation with respect to r gives (16).

(16)

$$\begin{cases} \hat{r} \overset{\Delta}{=} \dfrac{1}{\ell}\, \mathrm{Re} \sum_t \left[z(t) \exp - j\,\varphi_t\right] \approx \dfrac{1}{\ell}\, \mathrm{Re}\left[\sum_t z(t) \exp - j\,\hat{\varphi}_t\right] \\ \text{with } \hat{\varphi}_t \overset{\Delta}{=} \hat{\omega} t + \hat{\varphi} \end{cases}$$

As above we verify that $\hat{r}$ is not biased and its centred is variance obtained by taking the mathematical expectancy of the differential of the expression of $\hat{r} - r$.

(17) $$\left\langle (\hat{r} - r)^2 \right\rangle = \frac{\sigma^2}{\ell} = \frac{N_0}{T}$$

Remark : The critical step in the carrier evaluation is the phase development because of the risk of determination error also called « **click probability** ».

By suitable local integration and frequency hypothesis, this failure probability can be reduced to an acceptable level, the quantitative « **tuning** » of the corresponding parameters can be done by the suitable bounding methods given above.

Going back to the hypothesis testing using comparison of the modules of the carrier, it should be recalled that an error $\Delta\omega$ of the pulsation ω induces a reduction by a factor $\frac{T\Delta\omega}{2 \sin \frac{T\Delta\omega}{2}}$ of the reconstructed modulus as shown on the figure below.

To make things quantitive , $T\Delta\omega = \frac{\pi}{2}$ brings about an amplitude reduction of 1 dB.

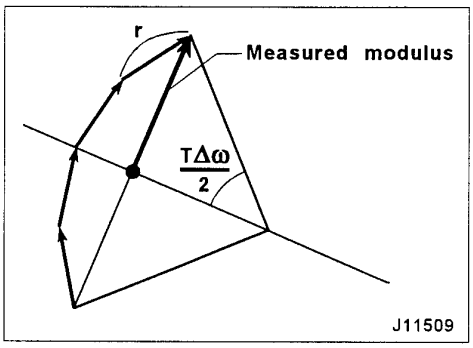


Figure 5 - Reduction factor resulting from a pulsation error

7) Phase noise characterisation

In any transmitter or receiver, sinewave sources are used for frequency translation and for carrier synthesis or identification. These sources are subject to 2 kinds of defects : amplitude distortion and phase noise. The first phenomenon can be eliminated in practice by hard amplitude limitation and band pass filtering. The remaining phenomenon can be reduced to the statistical description of the zero crossing time expressed as a phase angle of the carrier , it is thus possible to define a real discrete function s(t) which represents the signed difference between the exact value and the actual (*measured*) value. It seems reasonable to assume $\langle s(t) \rangle = 0$, but the statistical behaviour of s(t) is best described by its spectral function or Fourier transform of its autocorrelation function S(f). The Fourier transform factor is such that $\left\langle [s(t)]^2 \right\rangle = \int_0^\infty S(f)\, df$.

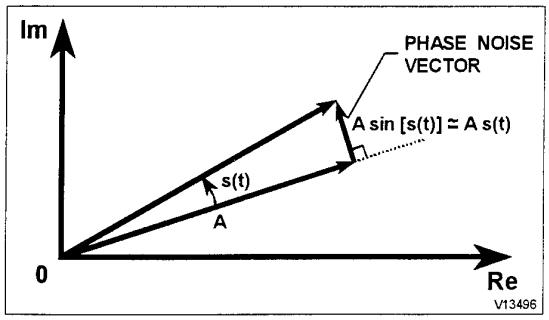


Figure 6 - Phase noise representation in the complex plane

It is thus possible to view s(t) as the result of the summation of random sinewaves $a(f)\cos(2\pi\, ft + \varphi)$ where φ is uniform within [0, 2π[and a(f) is defined by $\frac{a^2(f)}{2} = S(f)\ \Delta f$ for a bandwidth Δf small, or $a(f) = \sqrt{2\, S(f)\, \Delta f}$. Sinewaves at different frequencies are decorrelated. For a single sinewave $a\cos(2\pi\, ft + \varphi)$, the result of the integration (*carrier construction in the complex plane*) during time T is $\int_0^T a(f)\cos(2\pi\, ft + \varphi)\, dt = \frac{a(f)}{2\pi f}\left[-\sin(2\pi\, ft + \varphi)\right]_0^T$. In the mean, this corresponds to a power

$$\frac{1}{2}\,\frac{a^2(f)}{(2\,\pi\, fT)^2} = \frac{S(f)\ \Delta f}{4\pi^2\ f^2\ T^2}$$, so that the variance of the phase carrier is given by the integral below.

(18) $$\sigma_0^2 = \frac{1}{4\pi^2\ T^2} \int_0^\infty \frac{S(f)}{f^2}\, df$$

8) Quantification noise characterisation

The expression of the variance of additive quantification noise resulting from a real signal is well known when its amplitude histogram is uniform in a finite interval , we shall generalise this result and deduce optimum quantizer performance in the case of a finite number b of bits.

Given a real amplitude a with a local uniform distribution, any amplitude a in the interval $\left(a - \frac{\Delta a}{2}\ ,\ a + \frac{\Delta a}{2}\right)$ will be quantized by the same integer $\bar{a}$, so that this number $\bar{a}$ can be considered as a non biased estimator with variance

$$v_a = \int_{+a\,-\frac{\Delta a}{2}}^{a\,+\frac{\Delta a}{2}} (x - \bar{a})^2\, \frac{dx}{\Delta a} = \frac{(\Delta a)^2}{12}\ .$$

In the case of a (*smooth*) non uniform distribution , the probability density to observe the amplitude a is p(a) , so that p(a) Δa is the probability that a is in the interval $\left(a - \frac{\Delta a}{2}\ ,\ a + \frac{\Delta a}{2}\right)$. The global quantification noise variance will be $v = \sum_{\bar{a}} \frac{(\Delta \bar{a})^3}{12}\, p(\bar{a})$. In order to determine the minimum of v for an optimum digitizing law , we consider that

the sequence of intervals $\Delta\bar{a}$ is a partition of the variation domain of a .

When this partition is optimum, a small displacement ε of any boundary between two adjacent $\Delta\bar{a}_i$ leaves v unchanged.

$\Delta V = (\Delta\bar{a}_i + \varepsilon)^3\ p(\bar{a}_i) + (\Delta\bar{a}_{i+1} - \varepsilon)^3\ p(a_{i+1})$ must be stationary $\frac{d(\Delta V)}{d\varepsilon} = 0$, so that $(\Delta\bar{a}_i)^2\ p(\bar{a}_i)$ is a constant, independent of $\bar{a}_i$. The problem is completely defined when the number b of bits is given , knowing the amplitude extent 2A and the amplitude probability density p(a).

(19)

$$\begin{cases} v = \sum_{\bar{a}} \frac{(\Delta\bar{a})^3}{12}\ p(\bar{a}) \le \frac{(2^{-b+1}\ A)^3}{12} = V \text{ for uniform quantification} \\ \text{card}\ \{\Delta\bar{a}\} = 2^b\ ,\ \sum_{\bar{a}} \Delta\bar{a} = 2A\ ,\ (\Delta\bar{a})^2\ p(\bar{a}) = \text{constant} \end{cases}$$

The lowest value of v can be iteratively determined in the following manner :

Choose the first $\bar{a}$ (*lowest*) in the definition domain of a (*extent 2A*) , the value of $\Delta\bar{a}$ is then determined by (19) as well as the next $\bar{a}$. Repeat until the whole domain is covered by β intervals $\Delta\bar{a}$. By comparing the number of $\bar{a}$ to 2^b , it is possible to improve the choice of another first $\bar{a}$ to get closer to 2^b .

3. STANDARD MODULATIONS

1) FSK (frequency shift keying)

The elementary transmitted signal (*symbol*) is constituted by a sinewave of real amplitude a and of frequency belonging to the m-set f_1 , f_2 , ... , f_m during a constant time interval T . Usually $m = 2^b$, b bits are transmitted with each symbol.

The analytic signal associated to the received real signal is of the form (20) .

(20)

$$\begin{cases} z(t) = a \exp j\ (2\pi f_i t + \varphi) + n(t) \\ \text{with}\ i \in \{1,2,.....,m\}\ ,\ \langle n(t)\rangle = 0\ \text{and}\ \langle n(t)\ n^*(t+h\ \Delta t)\rangle = 2\sigma_0^2\ \delta_{0h} \end{cases}$$

Demodulation according to the maximum likelihood principle (cf II, 6) consists in looking for the G_i (*see below*) with the greatest modulus $|G_i|$.

$$G_i \triangleq \sum_{h=-N/2}^{N/2} z(t+h\ \Delta t)\ \left[\exp - 2\pi j f_i\ (t+h\ \Delta t)\right] \text{ with } T = N\ \Delta t$$

For the wrong hypothesis $|G_i|$ follows a Raleigh law with a variance σ^2 on each axis of the complex plane.

$$\text{Prob}\ (r \ge R) = \exp - \frac{R^2}{2\sigma^2}$$

For the correct hypothesis the square of $|G_i|$ is $R^2 = (A+x)^2 + y^2$ where $A = Na$ represents the reconstructed carrier modulus without noise , x and y are the Cartesian coordinates of n(t) oriented along the carrier.

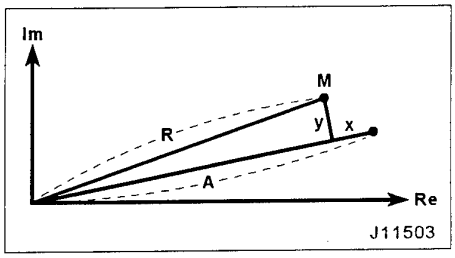


Figure 7 - Noise Components on the carrier reconstructed vector

For the correct hypothesis , the probability that the received analytic signal falls in the neighbourhood dxdy of M is

$$dP = \frac{1}{2\pi\sigma^2}\left(\exp - \frac{x^2+y^2}{2\sigma^2}\right) dx\,dy\ .$$

When comparing the correct hypothesis to any one wrong hypothesis , the result is incorrect if and only if the R corresponding to the wrong hypothesis is bigger than the R corresponding to the correct hypothesis. This occurs with probability P_2 :

$$P_2 \triangleq \iint \left(\exp - \frac{(A+x)^2+y^2}{2\sigma^2}\right) dP$$

$$= \left(\exp - \frac{A^2}{4\sigma^2}\right) \iint \frac{1}{2\pi\sigma^2} \left(\exp - \frac{\left(x+\frac{A}{2}\right)^2 + y^2}{\sigma^2}\right) dx\,dy$$

$$= \frac{1}{2}\exp - \frac{A^2}{4\sigma^2} = \frac{1}{2}\exp - \frac{Na^2}{4\sigma_0^2}$$

The demodulation of one symbol is correct if and only if the m-1 comparisons with the correct hypothesis are correct , the symbol error probability P_m is thus given by (21).

(21)

$$\begin{cases} P_2 = \frac{1}{2}\exp - \frac{E_b}{2N_0}\ \text{since}\ \sigma_0^2 = N_0 B\ ,\ B\Delta t = 1\ ,\ T = N\Delta t\ ,\ \frac{a^2\ T}{2} = E_b = b\ E_1\ ,\ m = 2^b \\ P_m = 1 - (1-P_2)^{m-1} \le \frac{m-1}{2}\ \exp - \frac{E_b}{2N_0} \le \frac{1}{2}\exp - \frac{b}{2}\left(\frac{E_1}{N_0} - 2\ell n\ 2\right) \end{cases}$$

2) PSK (Phase shift keying)

The elementary transmitted signal (*symbol*) is constituted by a sinewave of duration T of known frequency f and of phase φ_i taken among m equally spaced values. The analytic signal associated with the received signal is of the form (22).

(22)

$$\begin{cases} z(t) = a\exp j\left(2\pi f t + \varphi_i + \varphi_0\right) + n(t) & \varphi_0 \text{ is the phase reference} \\ i \in \{1,2,.....,m\}\ ,\ \varphi_{i+1} - \varphi_i = \dfrac{2\pi}{m}\ ,\ \left\langle n(t)\, n^*\left(t + h\,\Delta t\right)\right\rangle = \sigma_0\, \delta_{0h} \end{cases}$$

The optimal demodulation procedure consists , according to the maximum likelihood principle , in calculating the phase of z(t) and then choosing the closest phase hypothesis.

According to paragraph 2.6. the estimator of the above phase is unbiased.

Let θ be angle of error. In general the distribution of θ does not reduce to a simple form, so that to get a closed form, the shortest method consists in considering that θ follows a centered gaussian law (cf.(13)) with variance $\sigma_{\theta^2}^2 = \dfrac{\sigma_0^2}{a^2} = \dfrac{N_0}{2E_b}$ with $m = 2^b$, $\dfrac{a^2}{2} = BE_b$ as for FSK.

Demodulation is erroneous if and only if $|\theta| \geq \dfrac{\pi}{m}$, so that by (7) we get the approximation below.

(23)

$$\begin{cases} P_m \approx \dfrac{2}{\sqrt{2\pi}} \displaystyle\int_{\frac{\pi}{m}\sqrt{\frac{2E_b}{N_0}}}^{\infty} \left(\exp - \frac{\beta^2}{2}\right) d\beta < \frac{m}{\pi\sqrt{\dfrac{\pi E_b}{N_0}}} \exp - \frac{\pi^2}{m^2}\frac{E_b}{N_0} \\ P_2 = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{\sqrt{\frac{2E_b}{N_0}}}^{\infty} \left(\exp - \frac{\beta^2}{2}\right) d\beta \triangleq Q\left(\sqrt{\frac{2E_1}{N_0}}\right) \end{cases}$$

cf[3] : Case of BPSK (m = 2)

(23) shows that P_m is a little optimistic for m = 2 since $\dfrac{\pi}{2} > 1$, but the expression for P_m has been derived for high values of m. A more precise expression would be $P_m = Q\left(\sqrt{\dfrac{2E_b}{N_0}} \sin\dfrac{\pi}{m}\right)$.

3) Band limited channel capacity

Contrary to the situations just studied, information theory gives an upper bound to the maximum bit rate which can be transmitted without any error through a channel defined by its bandwidth B, the « useful » power $S = \dfrac{a^2}{2}$ and the variance v $= \sigma^2 = N_0 B$ of the additive noise.

S and σ^2 are defined at the same place, generally at the input of the receiver.

The binary bit rate is noted D, the useful received energy per bit is noted E_1 and the capacity C is the lowest upperbound of D. C is given by the Shannon formula (24).

(24) $C\ (B) = B\log_2\left(1 + \dfrac{S}{\sigma^2}\right) \Rightarrow \dfrac{D}{B} < \log_2\left(1 + \dfrac{E_1}{N_0}\dfrac{D}{B}\right)$

since $S = \dfrac{a^2}{2} = DE_1$, $D < C$, $\sigma^2 = N_0 B$

The inequality (24) shows that the energy efficiency $\dfrac{E_1}{N_0}$ is limited by the spectral efficiency $\dfrac{D}{B}$ as it is visible on the series development (25).

(25)

$$\frac{E_1}{N_0} > \frac{e^{\frac{D}{B}\ell n2} - 1}{\dfrac{D}{B}} = (\ell n2)\left[1 + \frac{1}{2!}\left(\frac{D}{B}\ell n2\right) + \frac{1}{3!}\left(\frac{D}{B}\ell n2\right)^2 + ...\right]$$

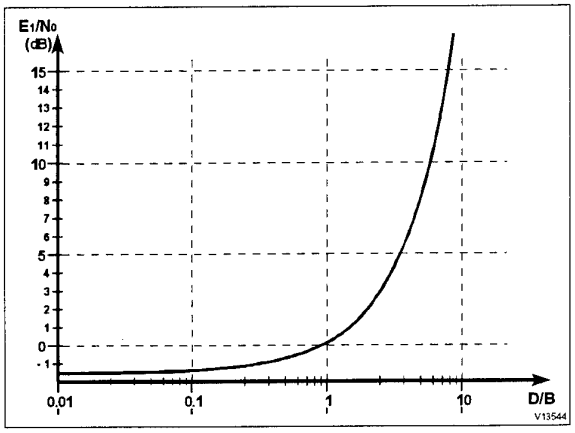


Figure 8 - Minimum bit energy over noise power density ratio E1/No necessary to transmit binary rate D through a bandwidth B channel

The proof of (24) is constructive but unpracticable because of the enormous number of operations needed. The engineering problem is to approach C with an acceptable implementation complexity, the organisation with an inner code and outer code is a promising way to achieve this goal.

4. INNER CODING

1) Introduction

In the decoding process, information theory incites us to take into account the amplitude of the received signal. The first coding example of this principle is the soft decoding of convolutional codes, and the most popular algorithm is due to Viterbi [2]. However the lack of powerful algebraic structure prevents the use of « long » codes necessary to reach low bit error rate. To bypass this difficulty, the objective of the innercode is to eliminate the most frequent errors so that the remaining errors are distributed in packets and are dealt with by the outer code, which is generally a character block code.

2) Algebraic structure

Since the pioneering publication of FORNEY in 1970 [1] the algebraic description of linear convolutional coding is quite satisfactory. Functionally, when k bits enter the encoder, n bits come out ; these n bits depend on all previous input bits.

The input sequence $x(\lambda)$ is represented by entire series in λ, whose coefficients are vectors with k binary components, where the exponents of λ increase with time. k represents the number of input multiplexed binary flows.

$$x(\lambda) = x_0 + \lambda\, x_1 + \lambda^2\, x_2 + \ldots\ldots$$

$$x_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ik} \end{pmatrix} \qquad \text{with} \qquad x_{ih} \in \{0, 1\}$$

The output sequence $y(\lambda)$ is represented in the same way by an entire series in λ.

$$y(\lambda) = y_0 + \lambda y_1 + \lambda^2\, y_2 + \ldots\ldots$$

$$y_i = \begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{ik} \end{pmatrix} \qquad \text{with} \qquad y_{ih} \in \{0, 1\}$$

Sometimes, it will be more convenient to consider that the input signal (*or the output signal*) is a vector whose components are (*binary*) entire series. We keep the same notations since the context can remove any interpretation ambiguity :

$$x(\lambda) = \begin{pmatrix} x_{01} + \lambda x_{11} + \lambda^2 x_{21} + \ldots\ldots \\ x_{02} + \lambda x_{12} + \lambda^2 x_{22} + \ldots\ldots \\ \vdots \\ x_{0k} + \lambda x_{1k} + \lambda^2 x_{2k} + \ldots\ldots \end{pmatrix}$$

In order to eliminate irrevelant calculation due to initial conditions , we consider that 2 entire series are equivalent if they differ by a factor λ^n where n is finite, or by a term of finite index (*or finite power of* λ). Let $P(\lambda)$ be a polynomial in λ , the series $x(\lambda)$ and $P(\lambda) + \lambda^n\, x(\lambda)$ are equivalent.

Since the encoding process is linear and causal (*finite time calculation needed to obtain the output step by step*) the output $y(\lambda)$ is related to the input $x(\lambda)$ by $y(\lambda) = G(\lambda)\, x(\lambda)$ where $G(\lambda)$ is matrix of n lines and k columns, the elements of which are polynomials in λ.

For instance the usual convolutional code 7, $\frac{1}{2}$ is defined by the matrix below.

$$G(\lambda) = \begin{pmatrix} \lambda^6+\lambda^5+\lambda^3+\lambda^2+1 \\ \lambda^6+\lambda^3+\lambda^2+\lambda+1 \end{pmatrix} = \begin{pmatrix} g_1(\lambda) \\ g_2(\lambda) \end{pmatrix} \qquad \text{with } k = 1 \text{ , } n = 2$$

$$y(\lambda) = \begin{pmatrix} y_1(\lambda) \\ y_2(\lambda) \end{pmatrix} = \begin{pmatrix} g_1(\lambda) \\ g_2(\lambda) \end{pmatrix} x(\lambda) \Leftrightarrow \begin{cases} y_1(\lambda) = g_1(\lambda)\, x(\lambda) \\ y_2(\lambda) = g_2(\lambda)\, x(\lambda) \end{cases}$$

Separate encoding of data and redundancy has practical advantages and it seems judicious to select generating matrix $G(\lambda)$ already partially diagonalised : $y_i(\lambda) = x_i(\lambda)$ for $i \le k$. Unfortunately it has been proven since a long time that the performance of the generated code is inferior to code generated by a more general encoding matrix $G(\lambda)$. Since then, many people reject the idea of separate encoding for convolutional coding.

In fact by exploiting results published by Forney [1], it is possible to prove that separate encoding, and existence of a syndrome and minimum encoding delay are compatible. This fundamental property results from the fact that encoding should be an injective mapping, i.e. distinct messages give distinct coded messages. We give now a sketch of the proof.

First of all, let us recall that any $x(\lambda)$ is possible and that the code efficiency is completely defined by the set of the $y(\lambda)$ generated.

The set of input messages $\{x(\lambda)\}$ is not modified if we multiply the entire series a $x_i(\lambda)$ by a non zero polynomial since the ring of polynomial has no zero divisor. $\{x(\lambda)\}$ is not modified if we interchange 2 entire series $x_i(\lambda)$ and $x_j(\lambda)$ or if we replace $x_i(\lambda)$ by $x_i(\lambda) + x_j(\lambda)$, provided that $i \neq j$. These operations on the $x_i(\lambda)$ series can be represented by $k \times k$ involutive matrices (*own inverse*) like :

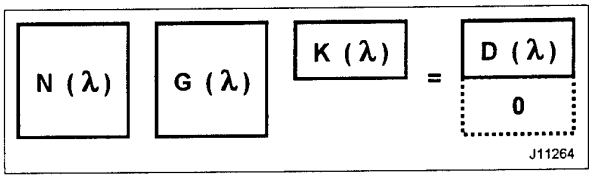
$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ and}$$

$$\begin{pmatrix} 1 & q(\lambda) \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & q(\lambda) \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ in GF(2) for k = 2}$$

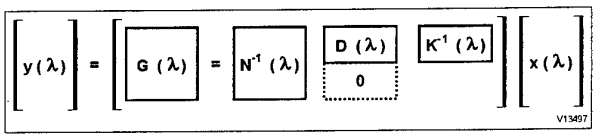
By successive applications of the above involutive matrix to the right of $G(\lambda)$ (*reversible linear combination of the columns of* $G(\lambda)$) we can diagonalise exactly k lines of $G(\lambda)$ since this generation matrix defines an involutive transformation of the $x(\lambda)$'s (*eliminate first the terms of highest degree*).

Let $D(\lambda)$ designate the $k \times k$ diagonalized part resulting from the above, and $K(\lambda)$ designate the $k \times k$ matrix which partly diagonalize $G(\lambda)$: $G(\lambda)\ K(\lambda) = \begin{bmatrix} D(\lambda) \\ \bullet \end{bmatrix}$.

The n-k lines of $G(\lambda)\ K(\lambda)$ which are outside of $D(\lambda)$ can be made equal to zero by linear line combination of $G(\lambda)\ K(\lambda)$. This last transformation is represented by a product of $n \times n$ involutive matrices noted $N(\lambda)$.



The inverse of $N(\lambda)$ and $K(\lambda)$ resulting from finite product of involutive matrices have polynomials in λ for elements and not rational fractions as might be feared.



By multiplying this equation on the left by $N(\lambda)$ we see that the last n-k lines of $N(\lambda)$ constitute an $(n\text{-}k) \times n$ matrix noted $H(\lambda)$ and called control matrix, which is such that any vector $y(\lambda)$

obeying the equation $H(\lambda)\, y(\lambda) = 0$ belongs to the code in the sense that one can find the $x(\lambda)$ such that $G(\lambda)\, x(\lambda) = y(\lambda)$.

$$\begin{bmatrix} N(\lambda) \\ H(\lambda) \end{bmatrix} \begin{bmatrix} y(\lambda) \end{bmatrix} = \begin{bmatrix} D(\lambda)\, K^{-1}(\lambda)\, X(\lambda) \\ 0 \end{bmatrix}$$

J11266

If $\hat{y}(\lambda)$ is the received (*and corrupted*) sequence, $y(\lambda)$ being transmitted, we can define a syndrome $S(\lambda)$ which is function only of the transmitted error sequence $\hat{y}(\lambda) - y(\lambda)$.

$$\textbf{(26)} \qquad S(\lambda) \triangleq H(\lambda)\, \hat{y}(\lambda) = H(\lambda)\, [\hat{y}(\lambda) - y(\lambda)]$$

In the case of convolutional code of rate $\frac{1}{2}$ the expression of $H(\lambda)$ is obvious.

$$y(\lambda) = \begin{bmatrix} y_1(\lambda) \\ y_2(\lambda) \end{bmatrix} = \begin{pmatrix} g_1(\lambda) \\ g_2(\lambda) \end{pmatrix} x(\lambda) \Rightarrow$$

$$H(\lambda)\, y(\lambda) = \left[g_2(\lambda)\, ,\, g_1(\lambda) \right] \begin{bmatrix} y_1(\lambda) \\ y_2(\lambda) \end{bmatrix} = 0$$

The code is completely defined by $H(\lambda)$, but $H(\lambda)$ is not unique, and we obtain the same code if we multiply or divide exactly any line of $H(\lambda)$ by a non zero polynomial in λ . It is also clear that we get an equivalent $H(\lambda)$ by adding to any line of $H(\lambda)$ a different line of $H(\lambda)$, eventually multiplied by a polynomial in λ . Permutation of lines of $H(\lambda)$ gives also an equivalent control matrix.

The equivalence relation that has just been defined allows us to build a control matrix of minimum degree in λ , and we can interpret the degree of $H(\lambda)$ as a measure of complexity.

Considering a control matrix in its minimal form $H(\lambda)$, we can write it as a polynomial of degree m with coefficients being $(n-k) \times n$ binary matrices.

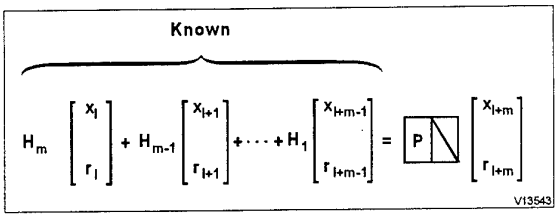
$$\textbf{(27)} \qquad H(\lambda) = \lambda^m H_m + \lambda^{m-1} H_{m-1} + \ldots + H_0$$
$$\Rightarrow \; H_m\, y_\ell + H_{m-1}\, y_{\ell+1} + \ldots + H_0\, y_{\ell+m} = 0$$

for any allowed $(\ell + m)$ power of λ

Since $H(\lambda)$ is in a minimal form, the lines of H_m must be linearly independent otherwise it would be possible to make $H_m = 0$ and decrease the degree of $H(\lambda)$. In a comparable manner the lines of H_0 must be linearly independent otherwise it would be possible to annihilate a line of H_0 and then divide the line of corresponding $H(\lambda)$ by λ . As a direct consequence we can derive an equivalent form of $H(\lambda)$ in which H_0 is partially diagonalized.

$$\begin{bmatrix} H_0 \end{bmatrix} = \begin{bmatrix} P \end{bmatrix} \begin{bmatrix} I_{n-k} \end{bmatrix}$$

A separate encoding of data and redundancy is then possible as can be checked on the control equation given in (27) : if the first k components of $y_{\ell+m}$ are taken identical to $x_{\ell+m}$, the n-k remaining components $r_{\ell+m}$ are then determined by the (n-k) control equations.



In the case of convolutional coding of rate $\frac{1}{2}$: $y(\lambda) = \begin{bmatrix} x(\lambda) \\ r(\lambda) \end{bmatrix}$

and the last redundancy $r_{\ell+m}$ is calculated from the control equation since the previous redundancies have already been calculated.

$$\left[g_{1,m}\, ,\, g_{2,m} \right] \begin{bmatrix} x_\ell \\ r_\ell \end{bmatrix} + \left[g_{1,m-1}\, ,\, g_{2,m-1} \right] \begin{bmatrix} x_{\ell+1} \\ r_{\ell+1} \end{bmatrix} + \ldots + \left[g_{10}\, ,\, g_{20} \right] \begin{bmatrix} x_{\ell+m} \\ r_{\ell+m} \end{bmatrix} = 0$$

3) Decoding efficiency

Given an input binary message represented by the entire series $x(\lambda)$, we define 3 entire series : the encoded sequence $y(\lambda)$, the received sequence $\hat{y}(\lambda)$ and the corrected sequence $\bar{y}(\lambda)$. In the absence of decoding error, we have $\bar{y}(\lambda) = y(\lambda)$, but in practice these 2 sequences are different and the decoding error sequence $e(\lambda) \triangleq \bar{y}(\lambda) - y(\lambda)$ belongs to the code and satisfies the control equation, since by construction $H(\lambda)\, \bar{y}(\lambda) = H(\lambda)\, y(\lambda) = 0$.

In order to evaluate the decoding efficiency of the maximum likelihood procedure, we introduce an adapted algebraic structure which preserves the weight distribution . Any sequence $y(\lambda)$ belonging to the code can be obtained by the concatenation of prefixes of finite length. A prefix is represented by a polynomial $Q(\lambda)$ whose coefficients are vectors with n binary components. Theses prefixes indexed by i satisfy the control equation $H(\lambda)\, Q_i(\lambda) = 0$ and no prefix can be built by concatenation of sequences corresponding to shorter prefixes. The weight w_i of the i^{th} prefix is equal to the number of its binary components equal to 1 and its length l_i is equal to the number of its n dimensional vectors. The first prefix is the zero vector having n zero components. To derive the other prefixes the easiest way is to use the impulse response of the transformation defined by the encoding (*or generating*) matrix $G(\lambda)$: $G(\lambda)\, x(\lambda)$ with $x(\lambda) = 1$.

For the convolutional code 7, $^1/_2$ defined at the beginning of the paragraph, the impulse response is

1 0 1 1 0 1 1 $\quad \ell = 7 \quad$ w = 10

1 1 1 1 0 0 1

By exhaustive search 5 prefixes of weight 10 are found corresponding to data 1, 11, 111, 1001 and 1011.

The decoding error sequence e(λ), since it belongs to the code, is also made of a sequence of prefixes. For each , the decoding efficiency is evaluated by the probability of being different from the zero prefix. The probability P(w) that a wrong prefix appears in the decoding error sequence e(λ) is a function only of its weight : the binary components equal to 1 characterizing a difference of likelihood , P(w) is given by the demodulation (*carrier reconstructing*) performance. In the case of optimum soft decoding of 2PSK, $P(w) = Q\left(\sqrt{\frac{2wE_1}{No}}\right)$ (cf. § III.2).

A wrong prefix (w, ℓ) induces a bit error rate $\frac{W}{n\ell}$ so that we can define a bit error rate p by summing the contributions of all prefixes :

(28) $$p = \frac{1}{n} \sum_{i \geq 1} \frac{w_i}{\ell_i} P(w_i)$$

5. OUTER CODING

1) Introduction

The aim of this coding is to deal with errors in packets as it is the case with the output of a convolutional code as we have seen just before. In general the outer code is a nonbinary block code with strong algebraic properties for hard decoding simplicity. The outer code is nonbinary in the sense that the elementary symbols are constituted by b bits so that each elementary symbol or character can take $q = 2^b$ distinct configurations. It can be considered that the Reed-Solomon Code family constitutes an optimum solution for outer coding.

This claim is justified by the fact that each member of this family is Maximal Distance Separable (MDS) and contains a large variety of configurations.

A MDS code is a code which meets the Singleton bound (1964) [5] with equality (cf. (29) below). This bound can be derived for any separable block code, even non linear. Indeed let us consider a block code of n characters which can take q distinct forms (*or values*) but where only q^k distinct blocks are allowed. Since the code is separable there is a k-tuple of components of the code which takes the q^k distinct possible values. Thus the (*reached*) minimum Hamming distance d between 2 k-tuples is 1 , and the remaining n-k components can only raise this distance by (n-k), this completing the proof of (29).

(29) $$d \leq n - k + 1$$

For a MDS code k = n-d+1 : the set of (*distinct*) k-tuple values must contain q^k elements, otherwise 2 codewords (*at least*) would be identical within these k places and their distance would be at most n-k which is less than the required value n-k+1. An important aspect of MDS code theory is that their operational performance can be evaluated by straightforward argumentation : essentially the inclusion-exclusion principle.

2) Topology of a MDS code

We are interested in deriving the number A_W of codewords which are exactly at the distance w from a codeword taken as reference.

By relabeling each of the n components of the code we can suppose that reference codeword is the n zero components.

To list the codewords of weight w ($\geq$ d) , we consider all codewords which have zero on (n-w) specified places, then we fix a component of the w-tuple of the non zero components (which gives (q-1) choices). We now evaluate the number E_0 of (w-1)-tuples composed of non zero elements and belonging to a codeword of weight w and having n-w zero at the above specified places.

In order to calculate E_0 we define W_j as follows : assuming that the (w-1) components defined above can take all values among the q choices , we fix j specified components among the (w-1) at the zero value and count the number of possible code-words having j or more zero components among the (w-1) and zero at the (n-w) first specifications . We iterate that count with another combination of j zero among (w-1) , until the total number of combinations is reached . W_j is defined as the sum of the counts. We can choose the zero components in $\binom{w-1}{j}$ manners. k components values can be chosen (*among q*) but n-w , 1 , j have been already fixed so that we get the relations below.

(30)

$$\begin{cases} W_j = \binom{w-1}{j} q^{k-[n-w+1+j]} = \binom{w-1}{j} q^{w-d-j} & \text{with } j \in \{0, 1, \ldots, w-d\} \\ A_w = \binom{n}{w}(q-1)\ E_0 & \text{with } w \in \{d, d+1, \ldots, n\} \end{cases}$$

To calculate E_0 from the W_j's , we introduce several quantities E_j which is the number of codewords having exactly j zero components among the w-1 considered for the definition of the W_j .

These 2 sets of quantities are related by relation (30).

(31)

$$\begin{cases} W_j = E_j + \binom{j+1}{j} E_{j+1} + \ldots + \binom{w-1}{j} E_{w-1} \\ W_1 = E_1 + 2E_2 + 3E_3 + \ldots + (w-1)\ E_{w-1} \\ W_0 = E_0 + E_1 + E_2 + \ldots + E_{w-1} \end{cases}$$

The expression of E_j for $j \neq 0$ is rather complex , but we need only the expression of E_0 given by (32). The easiest proof consists in replacing the W_j by their expression (31) and using the fact that for $j > 0$ the factor of E_j equals $\sum_{i=0}^{j} (-1)^i \binom{j}{i} = (1-1)^j = 0$.

(32) $$E_0 = W_0 - W_1 + W_2 \ .... + (-1)^{w-d} \ W_{w-d}$$

Collecting (30) and (32) we get the expression of A_W.

(33)

$$\begin{cases} A_0 = 1 \quad A_w = 0 & \text{for } 0 < w < d \\ A_d = \binom{n}{d}(q-1) \quad A_{d+1} = \binom{n}{d+1}(q-1)(q-d) \geq 0 \Rightarrow q \geq d \\ A_w = \binom{n}{w}(q-1)\, q^{w-d} \sum_{j=0}^{w-d} (-1)^j \binom{w-1}{j} q^{-j} & \text{for } w \geq d \end{cases}$$

From R-S code theory , we can construct a MDS linear code for any set of values satisfying (34).

(34) $$\begin{cases} d = n - k + 1 \\ q > n \end{cases}$$

$\sum_{j=0}^{w-d} (-1) \binom{w-1}{j} q^{-j}$ appears as the beginning of the expansion of $\left(1 - \frac{1}{q}\right)^{w-1}$. In the usual situation where (34) is satisfied , the absolute value of the terms of this expansion is small and decreasing since $\frac{\binom{w-1}{j} q^{j+1}}{q^j \binom{w-1}{j+1}} = \frac{q\,(j+1)}{w-1-j} \geq \frac{q\,(j+1)}{n-(j+1)} > 1$.

As a direct consequence we get a good approximation of A_W by replacing the above sum by $\left(1 - \frac{1}{q}\right)^{w-1}$

(35)

$$A_w \approx B_w \triangleq q^k \binom{n}{w} \left(1-\frac{1}{q}\right)^w \left(\frac{1}{q}\right)^{n-w} \qquad \begin{cases} B_d = \binom{n}{d} q \left(1-\frac{1}{q}\right)^d \\ A_d = \binom{n}{d}(q-1) \end{cases}$$

The expression of B_W shows that the distribution of the weights of the q^k codewords is nearly the same as if each of its n components were selected at random, uniformly among the q possibilities.

The global density of codewords in the whole set of n-tuples is $\frac{q^k}{q^n} = \left(\frac{1}{q}\right)^{d-1}$. The approximation B_W shows that the local density of codewords of a R-S code is nearly constant and equals the global density since $\binom{n}{w}(q-1)^w$ is the number of n-tuples of weight w :

(36) $$\frac{B_w}{\binom{n}{w}(q-1)^w} = \left(\frac{1}{q}\right)^{d-1}$$

In the case of R-S code , the error correction is correct if the number of errors at most equals $t \leq \frac{d-1}{2}$ so that the number of distinct syndromes leading to a correction is $\sum_{w=0}^{t} \binom{n}{w}(q-1)^w$.

When the number of errors is larger than t , the decoder may behave in two different ways :

1) One of the syndromes leads to a correction : this gives the false correction output.
2) None of the syndromes lead to any correction and the n-tuple is output unchanged : this gives the non correction output.

The probability P_f of false correction is equal to the number of syndromes leading to a correction multiplied by the codeword density.

(37)

$$P_f \approx \frac{\sum_{w=0}^{t} \binom{n}{w}(q-1)^w}{q^{2t}} < \frac{\sum_{w=0}^{t} \binom{n}{w}\left(\frac{1}{q}\right)^w}{q^t} < \left(\frac{1+\frac{1}{q}}{q}\right)^t$$

In all practical situations the risk of false decoding is negligeable , and the operational problem is what to do when the decoder refuses to give a correction.

3) Non-correction probability evaluation

Let p_{in} be the symbol error probability observed at the output of the inner decoder , the non-correction probability P_{out} is given by the expression (38) .

(38)

$$P_{out} = \sum_{i=t+1}^{n} \binom{n}{i} (p_{in})^i (1-p_{in})^{n-i} \qquad 2t+1 \leq d = n-k+1$$

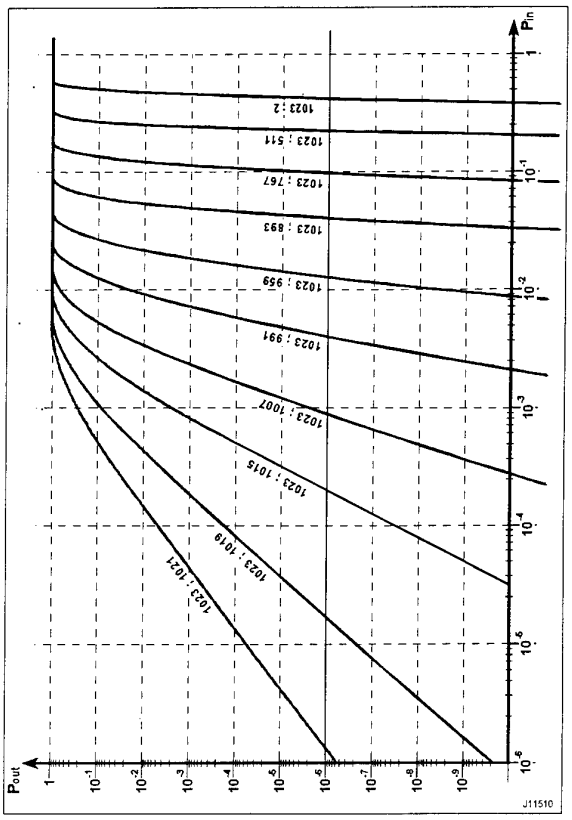


Figure 9 - Reed Solomon Code efficiency for n=1023 and various k

By using the bouding method of paragraph 2.4. it is possible to derive a closed form expression which can be used as a starting point for dimensioning.

$$P_{out} < \exp - n \frac{\left(\frac{t+1}{n} - p_{in}\right)^2}{2p_{in}\left(1 - p_{in}\right)} \qquad (39)$$

6. CONCLUSION

The results presented in this paper rely on two conceptual strategies : consider only optimum implementations and look for bounds rather than exact formulae. Optimal estimators reduce the number of hypothesis to test and as a consequence decrease the overall complexity. In addition simple bounds are more suggestive of physical interpretation. These are exemples of efficient methods to cope with combinatorial diversity, which increases faster than computing technology.

BIBLIOGRAPHY

[1] FORNEY (G.D) (1970) « Convolutional Codes I : Algebraic Structure » IEEE Trans. Inform. Theor. , vol. IT-16 - pp. 720 - 738.

[2] VITERBI (A.J) and OMURA (J.K) (1979) « Principle of Digital Communication Theory » Mc GRAW-HILL KOGAKUSHA , LMD.

[3] PROAKIS (J.G) (1989) « Digital Communications » Mc GRAW-HILL , Inc.

[4] MARGUINAUD (A.) and GIENGER (S.) (1995) « Discrete Hilbert Transform » AGARD.

[5] MAC WILLIAMS (F.J) and SLOANE (M.J.A) « The theory of error coding codes » (1978) North Holland Publishing Company.

Paper 26

P. Salmon (BE)

Ne pensez-vous pas qu'une application de votre simulateur de communications pourrait être celle de l'analyse de la qualité résiduelle d'une image brute codée transmise d'une station de réception à une station d'interprétation?

[Do you not think that a possible application for your communications simulator would be for analysing the residual quality of a rough coded image transmitted from a receiver to an interpretation station?]

Author's reply:

Je suis tout à fait d'accord, un codage en 2 couches (interne et externe) permet d'améliorer l'efficacité energétique et fournit un test de l'exactitude de la transmission. Le code externe préféré est un R-S de paramètres adaptés aux exigences opérationnelles, le code convolutif 7½ n'est pas la seule solution intéressante pour le codage interne.

[I entirely agree. Two layer coding (internal and external) gives an improvement in energy performance and tests the accuracy of the transmission. The preferred external code is a parameter R-S adapted to operational needs, the convolutive code 7½ is not the only valid solution for internal coding].

A Gradient Based Line Detector*

V. Lacroix M. Acheroy

Ecole Royale Militaire
Av. de la Renaissance, 30
1050 Bruxelles, Belgium
vinc@elec.rma.ac.be, http://www.elec.rma.ac.be

Abstract

As the resolution of satellite images progresses, the quantity of data to be treated becomes so large that much effort has to be done to design automatic procedures to perform routine tasks and to develop algorithms extracting specific information.
In this framework, we present a line extractor that, combined with an edge detector, will provide a useful tool for obtaining a vector-like description of satellite images.
The line extractor output is based on the dot product of the gradient vectors computed in two pixels taken symmetrically around the current pixel. If the latter lies on the middle of the line, and not the others, the dot product will be negative, as gradient vectors point in opposite directions. In practice, the investigation of the four nearest neighbour pairs is sufficient to determine the presence of a line. Exploring a larger neighbourhood enables to get an approximation of the local line width. Moreover, the operator can be set to detect selectively dark or bright lines, or both.
As for the edge detection process, a non-maximum suppression and a line following algorithm are needed to generate one pixel wide line elements.
The GLD is compared to the Duda Road Operator on a test image and on satellite images.

1 Introduction

Many computer vision tasks start from a simple image description usually obtained by a contour extraction process. The latter involves three steps [5]: first, the local edge strength and direction are computed, then local maxima are selected, and finally edgels are linked in order to form one pixel wide contours. The Canny operator [2] is a very efficient tool for the first step. However, as this operator is optimised for detecting a step edge in white noise, it hardly responds to thin lines and it does not respond *at the middle* but *at each side* of larger lines.
Most line detectors designed so far implicitly use an intensity model [6], [3]. Masks are used as templates: the best fit provides the operator output and the characteristics of the template (the local direction of the bar, for example) are assigned to the pixel. The results thus depend on how well the templates can represent a given line. Other detectors use the gradient; in [1] for example, a segmentation based on the gradient orientation guides the road detection. Finally, some road finders use the edges as in [7] where the author defines road-centre-hypotheses as anti-parallel intensity edges.
In this paper, we propose a new method based on the gradient vector field to detect lines *at their middle.*

*This work is subsidized by the Walloon Region of Belgium and THOMSON-CSF ELECTRONICS BELGIUM

2 The Gradient-based Line Detector (GLD)

An ideal one dimensional (1D) line is shown on Figure 1. If the line is lighter than the background, the derivative of the intensity, $f'(x)$, will be positive at the left side of the bar and negative at the right side. As data are discrete, the derivative is computed by convolving the signal with a kernel; for example, the kernels [-1 0 1] and [-0.04 -0.28 -0.56 0 0.56 0.28 0.04] respectively provide the discrete derivative and the derivative of the intensity regularized by a Gaussian.

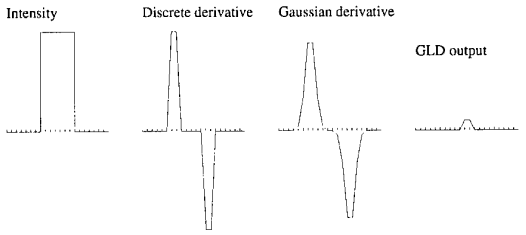


Figure 1: A 1D line, its derivative and detection

Let $f'(i)$ denote the derivative at pixel i. Then $-f'(i-1)\times f'(i+1)$ will be null outside the bar, negative around the intensity jump, and positive nearby the middle of the bar. Moreover, $-f'(i-1)\times f'(i+1)$ will be maximum *at the middle of the bar.* This property will remain as long as the convolving kernel is larger than the line. Thus a simple 1D line detector is defined as $b(i) = \max\{0, -f'(i-1)\times f'(i+1)\}$. In Figure 1, the bar output has been scaled for display

Paper presented at the AGARD SPP Symposium on "Remote Sensing: A Valuable Source of Information", held in Toulouse, France from 22-25 April 1996, and published in CP-582.

purpose.
Only bright lines will be detected if $b(i)$ is set to zero for $f'(i-1) < 0$. Inverting the condition provides dark lines.
The 2D generalisation is straightforward. At each pixel, the local neighbourhood is investigated. Let $\overline{G(p)}$ denote the gradient at p, then the bar operator b is defined as the maximum of the negative dot product of the gradient vectors taken symmetrically around the current pixel:
$b(p) = \max\{0, \max_k\{-\overline{G(a_k)} \cdot \overline{G(b_k)}\}\}$ where $(a_k\ b_k)$ denotes the k^{th} pair of nearest neighbours (see Figure 2 (a)).

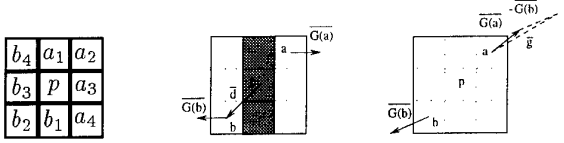


Figure 2: (a) The $(a_k\ b_k)$ pairs; (b) Condition for a dark line at p; (c) Direction perpendicular to the bar

Let $\overline{d}$ denote the displacement vector from p to a nearest neighbour q, then the condition for extracting only dark lines is given by (see Figure 2 (b)) $\overline{d} \cdot \overline{G(q)} \geq 0$. Bright lines are obtained by inversion of this condition.
The output of the operator thus depends on the size of the line *and* on the contrast of the line. Indeed, a large contrasted line may produce the same output as a thin poorly contrasted line.
Let m denote the pair providing the maximum of the negative dot product. The local direction perpendicular to the bar is taken as the direction of the vector $\overline{g_m} = \overline{G(a_m)} - \overline{G(b_m)}$ (see Figure 2 (c)).

3 Extended GLD

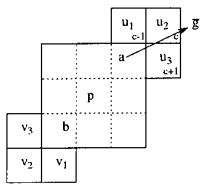


Figure 3: Extended GLD

The operator has been extended in order to produce an output (EGLD) proportional to the contrast of the line. A first approximation of the direction perpendicular to the line is given by $\overline{g_m}$. Let c be the chain code associated to this direction. Three other pairs are then investigated; they are made of the points lying in the direction c plus and minus one (modulo 8) and their symmetrical to the point p. Let $(u_k\ v_k)$ denote these pairs (see Figure 3). If the line is larger than one or two pixels, the gradient will be more important at u_k and at v_k than at a_m or b_m. The GLD computation will thus be larger for the pairs $(u_k\ v_k)$ than for the $(a_m\ b_m)$ pair. A new approximation of the bar direction may then be computed and further pairs may be investigated while the GLD output is larger. The number of times a further neighbourhood is investigated provides an approximation of the line width. Three informations are thus available at pixel p: an approximation of the local bar width, an approximation of the bar direction, and a line strength.

4 The line detection scheme

As for the edge detection process, the design of a filter is not sufficient; the filter should be integrated in a procedure transforming its output into a vectorial description, that is, one pixel wide line elements.
The complete line detection scheme involves four modules: a gradient module, the (E)GLD module, a local maxima selection module, and a line following module. Any algorithm may be used in modules preceeding or following the (E)GLD filter. Nevertheless, we present here the algorithms that we recommend.
For the first module, the Hummel and Lowe implementation [4] of the Gaussian Gradient [2] is the most attractive as it is based on the gradient of the intensity regularized by a Gaussian, the only regularizing function that is separable and isotropic.
The x component of the gradient is computed by first convolving the columns of the image with a mask $c[j] = \overline{R(j)}$, then convolving the lines of the resulting image with a mask $d[i] = -\overline{\partial R(i)}$, where $\overline{R(i)}$ and $\overline{\partial R(i)}$ denote the mean value of a 1D gaussian over the interval $[i-0.5, i+0.5]$ and of its derivative respectively. As masks are finite, the last coefficient at r is computed from the mean over the interval $[r-0.5, \infty]$.
More precisely,
$c[j] = 0.5\ (\text{erf}(\frac{j+0.5}{\sqrt{2}\sigma}) - \text{erf}(\frac{j-0.5}{\sqrt{2}\sigma}))$ for $0 \leq j < r_j$,
$c[r_j] = 1 - \sum_{j=0}^{j=r_j-1} c[j]$,
$c[j] = c[-j]$ for $j < 0$,

$d[i] = e^{-(\frac{i-0.5}{\sqrt{2}\sigma})^2} - e^{-(\frac{i+0.5}{\sqrt{2}\sigma})^2}$ for $0 \leq i < r_i$,
$d[r_i] = 1 - (1 - e^{-(\frac{0.5}{\sqrt{2}\sigma})^2}) - \sum_{i=1}^{i=r_i-1} d[i]$,
$d[i] = -d[-i]$ for $i < 0$.
Inverting the role of lines and columns provides the y component. The first module thus outputs two real rasters $G_x(i,j)$ and $G_y(i,j)$.
The (E)GLD computation performed in the second module is described in the previous section. The (E)GLD representing the bar norm is put in $B(i,j)$, the code associated to the bar direction is stored in $D(i,j)$, and the size in $S(i,j)$. The latter is not used further.
The local maxima are extracted in the third module by a non-maximum deletion algorithm: at each pixel, the bar norm of the current pixel $b(p)$ is compared to the bar value of the two pixels lying in the direction perpendicular to the line. If $b(p)$ is not

maximum, the point is deleted otherwise $\sqrt{b(p)}$ is stored in a raster $M(i,j)$. The square root computation makes the output directly proportionnal to the contrast of the line, thus keeping the values in a more practical range.
Finally, in the last module, a line following algorithm is performed. The image is scanned. A line starts at a pixel of $M(i,j)$ which has a bar norm larger than a threshold t_s. Then, the 8-neighbourhood is investigated to find the maximum of the bar norm $B(i,j)$. The algorithm then moves to that maximum. Let c_i be the code associated to this move. A "forward" line is initiated while always considering the maximum of the three next pixels characterised by the last move code c, $c+1$, and $c-1$. The maxima are kept in a list and marked. The "forward" line is stopped when the maximum of the bar output is null or is an already visited point. Similarly, a "backward" line is then followed using the opposite of the the initial direction c_i. The forward and backward lines are then appended and the list forms a "line" if its size is large enough.

5 Experiments

The GLD, the EGLD, and the enhanced Duda Road Operator [8] (DRO) have been applied to a test image and to several satellite images. The different outputs are difficult to evaluate; therefore, they served as input to a non-maximum suppression and a line following as described in section 4.
The DRO has been used to evaluate the GLD as it has been considered by Roux [8] as the best low level operator to extract the "thin line network" from satellite images. Here is a brief description of the operator. The intensity over a 5×5 window is analysed in order to assign a score to the centre pixel. The computation is made using four masks; two of them are shown in Figure 4. The masks measure uniformity of the intensity along a potential road track (a_1, a_2, a_3) and contrast of this potential track with adjacent terrain $(b_1, b_2, b_3; c_1, c_2, c_3)$.

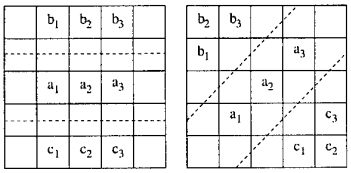


Figure 4: DRO Masks: (a) for horizontal and (b) for right diagonal road segments; similar masks are used for vertical and left diagonal road segments.

The measures of uniformity and contrast are performed thanks to the scoring functions $G(u)$ and $F(u)$ shown in Figure 5; $G(u)$ is designed to ignore slight variations in intensity along the road and to penalise all significant variations equally; similarly, $F(u)$ rewards all large contrasts equally. F and G are then combined in a scoring function S which represents the likelihood of the presence of a road. Roux modified the scoring function so that the operator does not respond to step edge.

$$F(u) = \begin{cases} M & u < 0 \\ M - \frac{(M-1/6)u}{\theta} & 0 \le u < \theta \\ 1/6 & u > \theta \end{cases} \quad \text{(a)}$$

$$G(u) = \begin{cases} G(-u) & u < 0 \\ 1 & 0 \le u < \theta_1 \\ 1 - (1-\epsilon)\frac{u-\theta_1}{\theta_2-\theta_1} & \theta_1 \le u < \theta_2 \\ \epsilon & u > \theta_2 \end{cases}$$

$$S = \frac{G(|a_1 - a_2|) \times G(|a_2 - a_3|)}{\max(\sum_{i=1}^{3} F(a_i - b_i), \sum_{i=1}^{3} F(a_i - c_i))}$$

with $M = 1.5, \theta = 15, \theta_1 = 5, \theta_2 = 20, \epsilon = 0.1$.

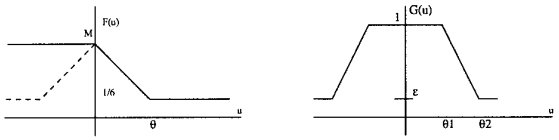


Figure 5: The F and G scoring functions involved in the DRO computation

The scores of the four masks are computed for each pixel. The maximum value provides the DRO output, and the mask that outputs this maximum value determines the local road direction. The score of a uniform area is $S_u = 1/3M$; therefore we enhance the operator output setting all scores $\le S_u + \delta$ to 0.

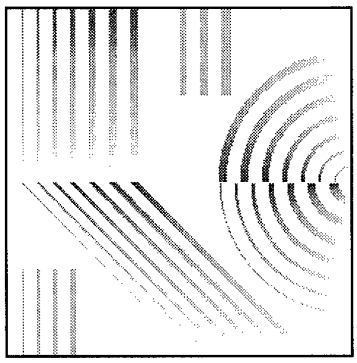
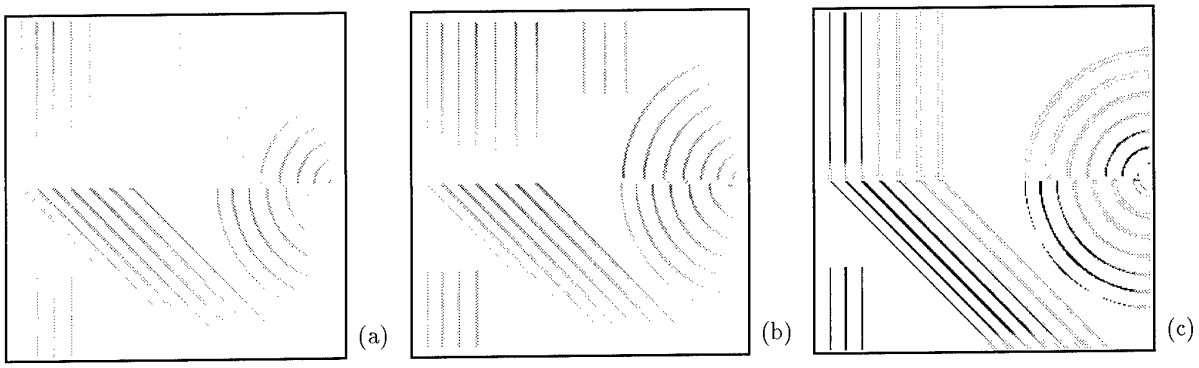
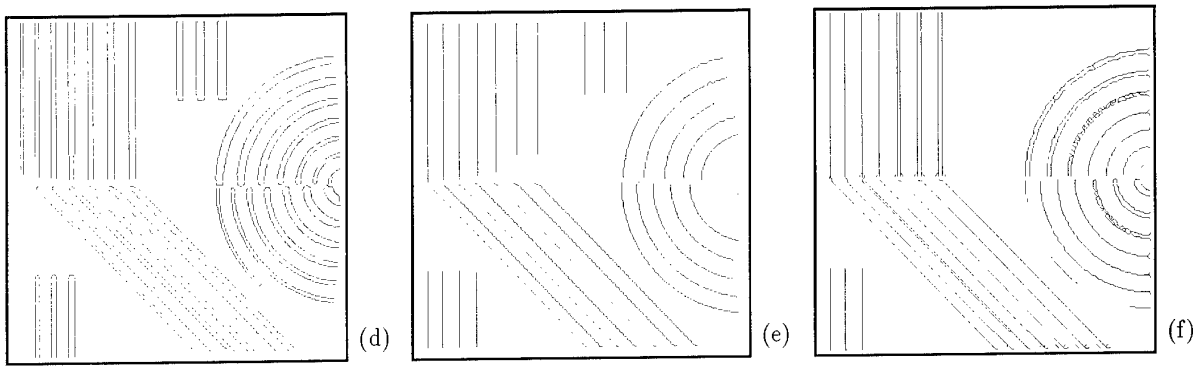


Figure 6: The test image

The test image is made of various lines and half rings of different width and varying contrast and of various lines of different width and of constant intensity (see Figure 6).
The scaled operator outputs are displayed in Figure 7 (a) – (c) and the lines extracted by the complete scheme are displayed in (e) – (f). The edges



Operator Output: (a) GLD; (b) EGLD; (c) DRO



(d) contours; (e) line following (GLD-EGLD); (f) line following DRO

Figure 7: Experimental results on the test image displayed in Figure 6

are displayed in (d) for comparison purpose (white corresponds to 0 and black to 255).

Let us first analyse the GLD output. In this experiment, letting $\sigma = 1$ in the gradient computation enables to detect all lines (from one to seven pixel wide). Note that the one pixel wide line could be detected by the GLD while not by the contour following algorithm; the gradient norm was too low to start a line there.

The even oblique lines in (e) are jingling because of the constraining line following algorithm. The problem arises because, the line having an even width, two pixels in the middle of the line have the same maximum GLD outputs, as one can see on a zoomed area shown in Figure 8. A clever line following algorithm should avoid such a behaviour.

109	109	0	0	0
108	108	108	0	0
107	107	107	107	0
0	106	106	106	106
0	0	105	105	105
0	0	0	104	104
0	0	0	0	103

97	49	0	0	0
96	96	49	0	0
48	96	95	48	0
0	47	95	95	48
0	0	47	94	94
0	0	0	46	93
0	0	0	0	46

Figure 8: Zoomed area (a) Intensity; (b) GLD

Note that the large oblique lines of decreasing intensity could be completely detected by the GLD, while only part of them could be detected on the large vertical ones. This anisotropy is probably due to discretization aspects (discrete gradient, relative positions of the neighbours considered in the GLD computation, etc.).

Let us now examine the DRO output. Thanks to Roux, the operator does not indeed respond to the perfect step edges of large vertical bars; however, it responds to the edges of the wide vertical lines and large rings of decreasing intensity.

The thin vertical lines (one to four pixel wide) whether being be of constant intensity or not, are all detected. Large diagonal lines are detected at their middle while vertical are detected at their edges. The thinnest ring could only be partly detected. This detector thus shows an important anisotropy.

The third diagonal line is partly double because the DRO output was maximum for the pixel lying along the axis of the line and for both lines lying at each side as one can see on the zoomed area in Figure 9. A non-maximum suppression algorithm considering a larger window might cope with this situation.

Finally, this filter has a strange effect on the extremities of all lines.

108	108	108	0	0
0	107	107	107	0
0	0	106	106	106
0	0	0	105	105
0	0	0	0	104

254	254	254	0	0
0	254	254	254	0
0	0	254	254	254
0	0	0	254	254
0	0	0	0	254

Figure 9: Zoomed area (a) Intensity; (b) DRO



(a) (b) (c)

Figure 10: Experiences on a spot image; (a) original image; (b) DRO output; (c) GLD output ($\sigma = 1$).

The operators are then applied to a SPOT and an ERS image. Results are shown in Figures 10 and 11. One problem of the DRO is its sensitivity to the δ parameter; a too small value has the effect of keeping pixels that are in almost constant area while a too large value removes much of the candidates. In Figure 10 we have set $\delta = 0.02$. The second problem of the DRO is its sensitivity to a rescaling of the intensity. None of these problems is encountered with the GLD.

3	0	0	0	0	0	0
3	0	0	0	0	0	0
0	6	6	6	0	0	0
0	0	0	6	8	0	0
0	0	0	0	8	0	0
0	0	0	3	0	0	0
0	0	0	0	4	0	0
0	0	0	3	0	0	0
0	0	0	6	0	0	0
0	0	0	0	3	0	0

1	0	0	0	0	0	0
1	1	0	0	0	0	0
0	1	1	0	0	0	0
0	0	1	1	0	0	0
0	0	0	1	1	0	0
0	0	0	0	1	1	0
0	0	0	0	0	1	0
0	0	0	0	0	2	1
1	0	0	0	0	0	2
0	0	0	0	0	0	0

Figure 12: ((a) Size at maxima around the white field; (b) Size at maxima of the road.

0	0	2	0	0	0
0	0	2	0	0	0
0	0	2	0	0	0
0	0	3	0	0	0
0	0	3	0	0	0
0	0	4	0	0	0
0	0	4	0	0	0
0	0	3	0	0	0
0	0	2	3	0	0
0	0	0	2	0	0

3	0	0	0	0	0
0	3	0	0	0	0
0	2	0	0	0	0
0	3	0	0	0	0
0	0	3	0	0	0
0	0	3	0	0	0
0	0	0	3	0	0
0	0	0	3	0	0
0	0	0	3	0	0
0	0	0	0	3	0

Figure 13: ((a) Size at maxima around the grey field; (b) Size at maxima of the airport track.

For the spot image, the GLD provides an output that is a little more better than the DRO; indeed, larger parts of the roads and of the airport tracks appear in the output. Note that the large white and grey fields located between the road and the airport are also extracted; this may be seen as a drawback of the procedure. In such a case, the size information might be useful: large size numbers and a high variation of these numbers are found in at the middle of the field compared to size numbers obtained at the middle of the road (see Figure 12) or at the middle of the airport track (see Figure 13).

For the ERS image, the DRO output is almost empty because the algorithm is too noise sensitive. Note that despite this noise, the GLD is able to detect parts of the roads and of the airport (here, a $\sigma = 2$ has been used for the gradient computation). This ability to work at any resolution is another strength of the GLD.

Another way to extract the lines at a coarser resolution is to shrink the image and extract the lines on this filtered image. In this case, only part of the airport appears on the DRO output while almost the whole road is extracted with the GLD ($\sigma = 1$).

In summary, the first advantage of the GLD over the DRO is its ability to detect lines of any size[1] and to respond at their middle. Indeed, though the DRO correctly responds to line from one to four pixels wide, it generates two responses nearby the edges of larger lines. Moreover, the direction of the line is not correct for pixels belonging to large lines.

The second advantage of the GLD is to provide, for a line of given width, an output that depends on the line contrast. As regard the DRO, it provides the same output for all pixels of thin lines, except for the very few where the contrast is weak. The output of the EGLD is even more useful as it only depends on the line contrast. Nevertheless, the GLD is able to detect the largest line as it provides a weak but non-null response at its middle. Therefore, if the contrast information is useless, the line detection scheme using the GLD is sufficient.

The third advantage of the GLD is to cope with noise, thanks to its ability to work at any resolution, the resolution being set in the gradient computation module.

Finally, in this framework, no threshold is necessary to get the GLD lines while the DRO output should

[1] as long as the line width is narrower than the derivative kernel

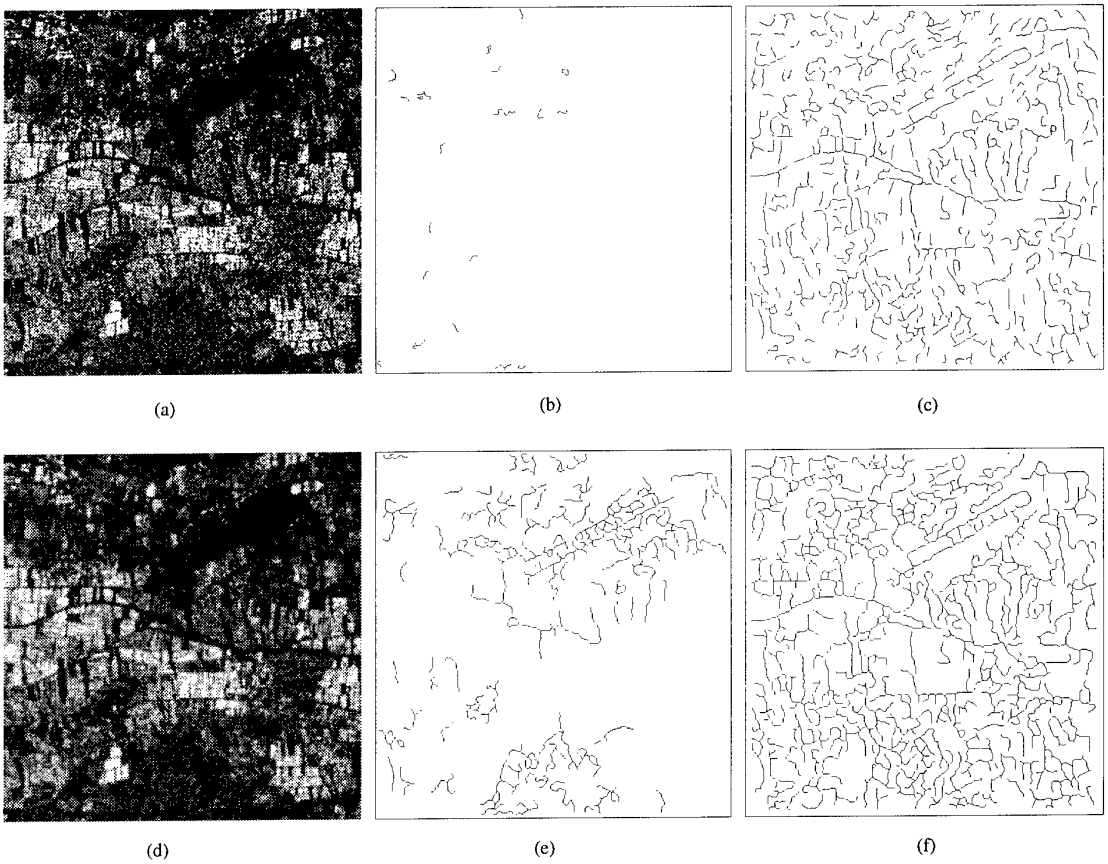


Figure 11: Experiences on an ERS image; (a) original image; (b) DRO output; (c) GLD output ($\sigma = 2$); (d) shrinked image; (e) DRO over shrinked image; (f) GLD over shrinked image ($\sigma = 1$).

be thresholded in order to remove the pixels lying in an almost constant area. Moreover, in the 3-module strategy proposed here, the DRO output is different if an intensity rescaling of the image is performed.

6 Conclusions

We proposed the GLD, a line detector filter. The operator makes use of the dot product of neighbouring gradient vectors. One pixel wide lines may be obtained after a line following algorithm that started on local maxima of the GLD and continue in the direction of the next maximum.

The method presents several interesting aspects: it produces a clean output, it can cope with noise in the intensity, it detects lines of various widths, it may generate an output dependant on the line contrast, and finally it provides an approximation of the local line width.

Applying such a scheme for detecting roads on satellite images provides very good results.

References

[1] Burns, Hanson, and Riseman, "Extracting straight lines", PAMI-8:425–455, 1986.

[2] Canny, "A computational approach to edge detection," PAMI-8:679–697, 1986.

[3] Fischler, Tenenbaum and Wolf "Detection of roads and linear structures in low-resolution aerial imagery using a multisource knowledge integration technique," CGIP:15:201-223, 1981.

[4] Hummel and Lowe, From the Pixels to the Features, Elsevier Science Publishers, The Nederlands, 1988.

[5] Lacroix, "A three module strategy for edge detection," PAMI-10(6):803–810, 1988.

[6] Nevatia and Babu "Linear feature extraction and description," CGIP-13:257–269,1980.

[7] McKeown "Towards automatic cartographic feature extraction," Mapping and Spatial Modelling for Navigation, NATO ASI Series, Vol. F 65, Springer-Verlag 1990.

[8] Roux, "Recalage d'images multi-sources. Application au recalage d'une image SPOT et d'une carte," Thesis Report, p.54, ENST, Paris, 1992.

Paper 27

M. McKean (CA)

How long does your operator take to work on a complete scene?

Author's reply:

I think it is not worth discussing CPU time as it depends on the machine and on some implementation details. I believe it is more interesting to look at the computations involved. Thus, if the gradient has already been computed, 4 dot products should be considered for the GLD norm computation at each pixel, and 4 more for some pixels, then 4 more for a subset of these (if one is interested in the line's width or if one wants an output proportional to the contrast of the line). Moreover in this last case, a sqrt should be computed.

Note that this scheme can be done in parallel.

M. Acheroy (BE)

Have you made a comparison with other methods?

Author's reply:

Yes, we have chosen the Duda Road Operator (DRO) as a basis of comparison as it has been considered recently as the best "fine line" detector. This operator provides good results if the parameters are correctly chosen. A problem is thus the automatic choice of these parameters. Moreover, as this operator is based on the intensity - and not on its derivative as our operator - a rescaling of the intensity will change the output of the detector; the output of the DRO does not provide an output proportional to the line contrast. Also, the output of the DRO is not nil in constant areas, so that this case should be treated carefully.

Paper 27

E. Schweicher (BE)

Pour vous débarasser des petites lignes de bruit, n'est-il pas imaginable d'utiliser un filtrage de Kalman, (ou un filtre α,β) et si oui avez-vous l'intention d'améliorer ainsi votre méthode?

[In order to remove the small noise lines, have you thought of using Kalman filtering (or an α,β filter) and if yes, do you intend to improve your method?]

Author's reply:

On pourrait en effet l'imaginer. Cependant, si un tel filtre existait au niveau du signal, on serait obligé de considérer un ensemble de directions car le GLD ne donne pas la direction de la barre avec précision. Cette procédure allongerait significativement le temps de traitement.

De plus, les lignes qui nous interessent peuvent être courbes, et un tel filtrage les effacerait sans doute. Néanmoins, il est certain qu'une procédure supplémentaire serait nécessaire, surtout dans le cas des images SAR.

On imaginerait plus volontiers une procédure qui lie les éléments de ligne entre eux si localement on trouve en effet confirmation, ou les supprime dans le cas contraire. On pourrait également se servir des informations sur la taille des barres.

[It could be envisaged. However, if such a filter existed at the signal level, you would have to consider the overall directions as the GLD does not give an accurate direction of the bar. This method would lengthen significantly the processing time.

Morever, the lines which interest us could be curved, and a filter as suggested would no doubt remove them. Nevertheless, it is certain that an additional procedure would be necessary, particularly in the case of SAR images.

It is easier to envisage a procedure which links the line elements between them if one actually found confirmation locally, or eliminates them in the opposite case. One could also make use of the information on the size of the bars.]

Eco-Climatic Classifications by Use of NOAA-AVHRR Data A Case Study: Tuscany (Central Italy)

L. Petkov[1], F. Maselli[2], C. Conese[2], G.Tacconi[3]

1) Ce.S.I.A (Centro di Studi per l'Applicazione dell'Informatica all'Agricoltura)
- Accademia dei Georgofili, Logge Uffizi Corti, 50122 Firenze, Italy
2) I.A.T.A. (Istituto di Agrometeorologia e Analisi ambientale applicata all'Agricoltura)
- C.N.R., P.le delle Cascine 18, 50144 Firenze, Italy
3) I.A.N.-C.N.R., 16149 Genova, DIBE , Università di Genova, 16145 Genova, Italy.

1. SUMMARY

Eco-climatic classifications have proved to be of great utility for the planning and management of various agricultural and forestry activities. Since usual methods are often expensive and not easy to apply, remote sensing data processing has recently been proposed to identify ecologically homogeneous land units. NOAA AVHRR data in particular have been demonstrated to be suitable for this purpose thanks to their spatial resolution and frequent acquisition. While most works in this field have dealt only with NDVI imagery, it has been suggested that thermal and ancillary data could also bring significant information. These considerations have been kept in mind during the present study, which concerns the eco-climatic classification of a complex Italian Region by the use of NOAA NDVI and thermal images from two years in addition to ancillary data. A complete methodology was developed for the processing of this integrated data set with particular reference to the identification of the suitable numbers of main and sub classes. The results obtained testify to the eco-climatic value of the product obtained.

2. INTRODUCTION

Climatic classifications were introduced several decades ago as a means of subdividing the earth surface into climatically homogeneous regions (Ref 1, 2). Their utility is greatest for environmental applications such as agriculture and forestry planning and management. For these operations, however, other factors also need to be taken into consideration mainly related to the topography of the terrain, the fertility of the substrate and the prevalent land use. The term eco-climatic classification has therefore been coined including all the environmental parameters which can affect land primary productivity (Ref 3).

These classifications can nowadays be essentially seen as a tool for extrapolating point measurements to wide areas, in such a way that the agro-meteorological parameters measured in a few stations (temperature, rainfall, radiation, etc.) can be extended over a Region. The importance of this extrapolation is especially evident for planning agricultural activities which are linked to meteorological events (water management, application of fertilizers and pesticides, etc.). The need therefore exists for objective methodologies for the agro-ecological characterization of the land surface with a spatial scale suitable for each application. Unfortunately, this is often not simple, since primary productivity is affected by numerous environmental factors the spatial distribution of which is generally unknown (Ref 4).

Modern techniques are generally based on the use of Geographic Information Systems (GISs) to process several information layers regarding morphology, climatology, geology and land-use. Serious problems persist, however, related to the inaccurate definition of boundaries in the layers and to the problematic development of standard methodologies for their composition. A different approach could be based on the use of vegetation as an indicator of existing ecological conditions, since vegetation integrates all the ecological effects which are relevant to eco-climatic classifications. Such an approach was originally proposed by classical authors (Ref 5), but is now more feasible owing to the availability of remotely sensed data, which provide continuous and synoptic information on vegetated surfaces.

Among the remote sensing systems presently available, the NOAA-AVHRR sensor has a spatial resolution and a temporal frequency suited for this objective on a regional scale. The information derived from its first two channels through the Normalized Difference Vegetation Index (NDVI) and from the thermal infrared bands has been demonstrated to be related to vegetation composition and condition (Ref 6, 7). It is therefore expected that agro-ecological classifications on a regional scale can be derived by suitably processing AVHRR multitemporal data. In effect, this expectation has been fullfilled by preliminary investigations of several research groups (Ref 8, 9, 10).

In the present paper a Region in Central Italy is introduced (Tuscany) which is particularly suitable for the research thanks to its high variability in land features. An original methodology is then presented for the agro-ecological classification of the Region based on the integrated use of NOAA-AVHRR visible and thermal infrared data. This classification is compared to meteorological data from ground stations and existing cartography about the Region. Finally, some conclusions are drawn as to the possible extension of the methodology to other areas.

Paper presented at the AGARD SPP Symposium on "Remote Sensing: A Valuable Source of Information", held in Toulouse, France from 22-25 April 1996, and published in CP-582.



Figure 1 - Geographical position of the study Region (Tuscany)

3. STUDY REGION

The Tuscany Region, which is part of north-central Italy, was chosen as the test site. Geographically it is located between 9°-12° east longitude and 44°-42° north latitude (Figure 1). The whole Region is characterized by a complex physical geography (topography, climate and vegetation), with extreme heterogeneity in morphological and climatic features, in land cover and land use.

The topography varies from the plain areas near the coast-line and around the principal river valleys to the hilly and mountainous zones towards the Appennine chain. From the orographic point of view the Region is covered approximately 2/3 by hilly areas, 1/5 by mountains and only 1/10 by plains and valleys.

The climate ranges from typically Mediterranean to temperate warm and temperate cool following the altitudinal and latitudinal gradients and the distance from the sea (Ref 11).

The land use is prevalently agricultural in the plain zones and mixed agricultural and forestry in the hilly and mountainous areas. The main agricultural cover types are cereal crops in the plains and olive groves and vine-yards on the hills. The upper mountain zones are almost completely covered by pastures and forests.

4. MATERIALS AND METHODS

4.1. Selection of the data set

The data set for the cluster analysis consisted of monthly NDVI Maximum Value Composites (MVC) from two years (1989-1990), of four surface temperature (ST) images and of two ancillary images reporting South-North and East-West gradients. The MVCs reported information about vegetation dynamics in the Region, while the ST images were related to land temperatures in different seasons. Finally, the distance images were chosen to introduce a factor of spatial homogeneity into the cluster analysis process.

The NDVI data were taken from the Telespazio (Rome) archives already in the form of vegetation index (NDVI) data. Some pre-processing operations (radiometric calibration, geometric rectification, decadal Maximum Value Compositing) had already been performed on the data. Thus, only the monthly compositing and a change in geographical projection were necessary for these images. The former operation was carried out in order to eliminate possible residual atmospheric disturbances and cloud contaminations which could remain in the ten-days composites. Seven monthly MVCs for each year covering the growing season (March - September) were finally obtained for further processing.

The ST images were derived from calibrated bands 4 and 5 AVHRR data taken from ESA (Rome). Among the scenes available for the two study years (about 40), only six were completely cloud-free and practically unaffected by atmospheric variations throughout the entire Region. From these six dates, four were selected representative of different seasons (Table 1). The scenes were first georeferenced using a third degree polynomial equation trained on ground control points. Since the data were in Sharp 2A format (brightness temperature in channels 4 and 5), it was sufficient to apply the split-window algorithm to derive surface temperature data (Ref 12). No correction for variations in surface emissivity was made at this stage.

The distance images were synthetically generated through an apposite program. S-N and E-W gradient images were considered as additional channels in the data set. The purpose for this inclusion was to facilitate the clustering of spatially contiguous pixels, which was deemed useful in view of the generation of an eco-climatic zonation.

Table 1
NOAA-AVHRR imagery used in the research, divided into NDVI MVCs (A) and TI images (B)

(A)

1989: Decade 7, 9, 10, 12, 13,15, 16,17, 19, 20, 21, 22, 23, 24, 25, 26, 27
1990: Decade 7, 8, 9 11, 12, 13, 14, 15, 16, 17, 18, 21, 22, 23, 24, 25, 26, 27

(B)

Date	Time
09.05.1989	11.51
05.10.1989	12.58
05.03 1990	13.03
21.07.1990	13.13

4.2. Pre-Processing

The data processing was carried out using Fortran 77 and C programs written in house. A Principal Component Transformation (PCT) was first applied to each study year (seven monthly NDVI MVCs) in order to compress these data which were markedly redundant and could raise problems for the subsequent clustering.

No statistical transformation was necessary for the ST images, which, taken in different seasons, were generally not strongly correlated. These images were merely stretched to render their information comparable with that of the NDVI data. A similar stretching was applied to the distance images in order to make their dynamic range consistent with that of the other images.

4.3. Cluster Analysis

The cluster analysis adopted for the present research is a conventional non-hierarchical process based on the iterative stabilization of the class centroids (Ref 13). These centroids are selected randomly in the data set during the initial phase of the process differently for each run. Only the number of categories must therefore be chosen together with the stabilization threshold (usually kept near 0).

4.4. Selection of the Optimum Number of Clusters

A critical point for all non-hierachical cluster analysis procedures is the selection of the optimum number of classes. A specific methodology was therefore developed for the current research based on a statistical non-parametric approach.

The rationale for this approach is that, given that the initial centroids are selected randomly, the final configuration derived from different runs will be more stable the more suitable it is to describe the intrinsic division of the data set examined. In other words, the final division into classes from several repetitions will remain stable if it is natural for the study data while it will vary in different runs if this is not the case.

Stability for different runs of the cluster analysis can be evaluated by the Tau (τ) coefficient (Ref 14). This is a non-parametric measure of agreement between nominal scale variables which does not require correspondence between the levels of the variables and is therefore suited to evaluating the stability of different clusterings. In practice, the τ statistic is a normalized χ^2 coefficient which ranges from 0 (no agreement) to 1 (complete agreement). Mathematically, it can be calculated from a contingency table comparing two classifications by means of the formula:

$$\tau = \sqrt{\frac{\chi^2}{(N_c - 1)N_t}} \qquad (1)$$

where:
χ^2 = Chi square from the contingency table;
N_c = number of classes;
N_t = number of points (pixels).

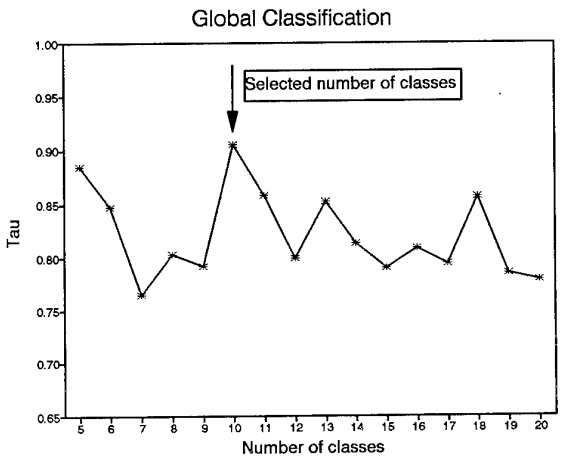


Figure 2 - Tau coefficients found from the five repetitions of the cluster analysis with varying number of classes (from 5 to 20).

The average τ index for five repetitions of the cluster analysis was computed with various numbers of classes (from 5 to 20). As can be seen from Figure 2, the highest stability was found with ten classes, which were therefore selected as the optimum configuration for the study data set.

4.5. Reclassification of the Main Categories

The same clustering process and evaluation of optimum number of classes were applied to each of the main categories in order to check their internal homogeneity. The average τ index from five repetitions was computed as above for the ten main categories, indicating that each of them could be further divided into a number of sub-categories ranging from two to four.

4.6. Comparison of the Classifications with Existing References

As demonstrated by previous works, no single validation method can be used when dealing with spatial resolutions such as that of NOAA-AVHRR pixels, since landscape characteristics are extremely difficult to define on these scales (Ref 15). The eco-climatic value of the classifications obtained was therefore evaluated indirectly by various comparisons. Land use information derived from the data-base of the Tuscany Region was first considered. Next, our classification was compared to a climatic classification independently produced by Rapetti and Vittorini (Ref 11).

The actual land use of the Region is mapped in the periodic inventory carried out by the Regional administration, which reports this information on 1:25000 cartography.

The climatic classification considered as reference is a very recent product of a large scale research project aiming at

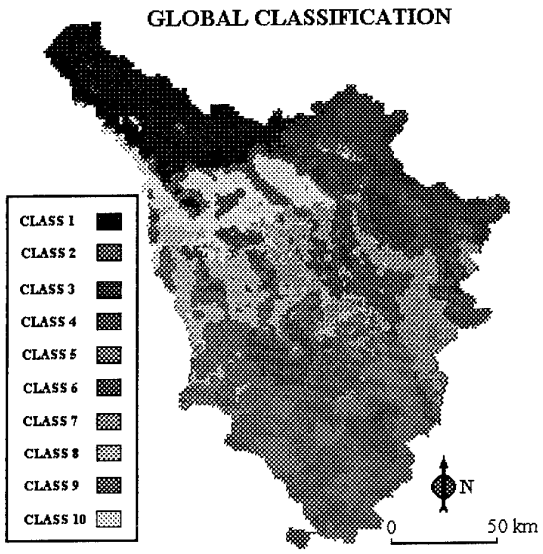
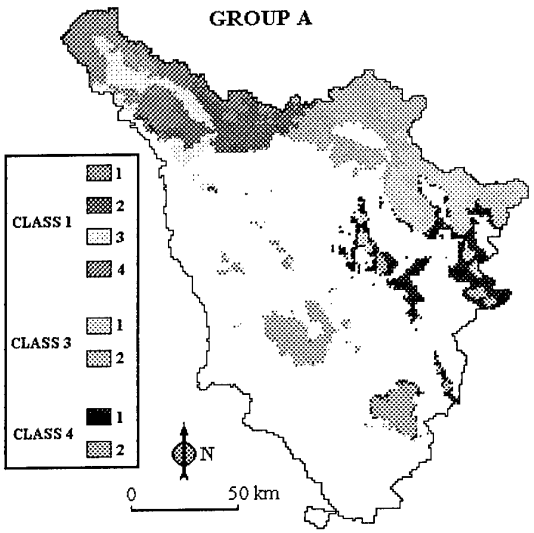
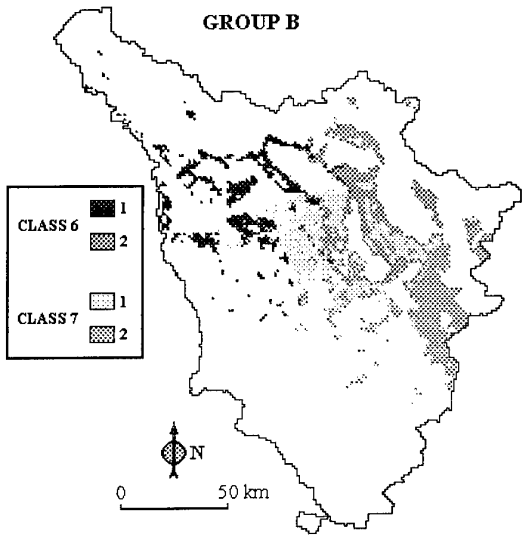


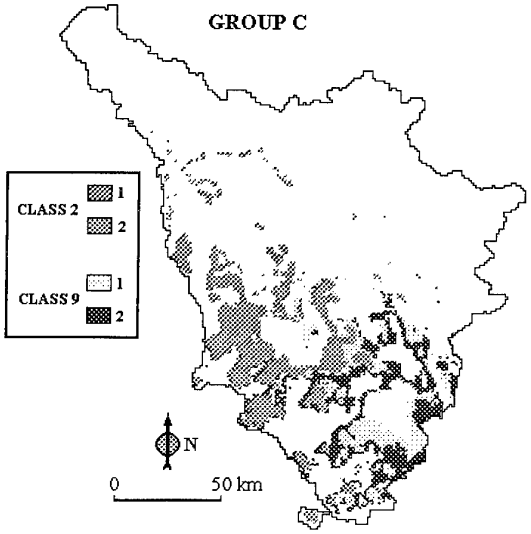
Figure 3 - Distribution of the ten main classes in the study Region.

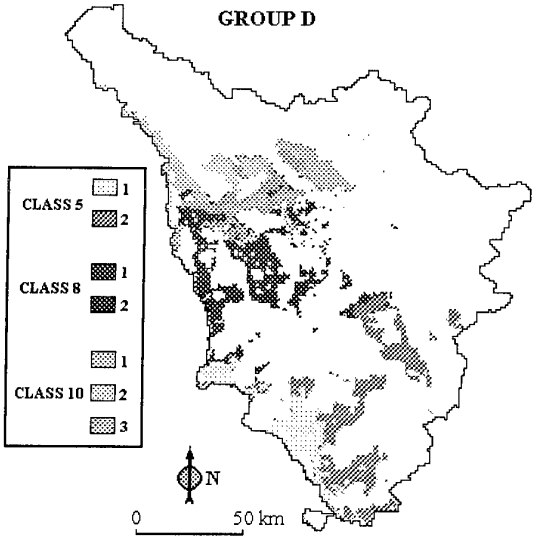
characterizing Tuscany using meteorological time series from almost 60 stations (Ref 11). Several factors influencing the climate have also been considered in this work, such as topography, latitude, distance from the sea, etc. In practice, the classification is based on the definition of annual average temperature and two main limiting factors for plant growth (number of dry months and number of cold months). A complete description of the methodology used is beyond the scope of the present paper and can be found in Rapetti and Vittorini, (Ref 11).

Since the subjects of all these data were not completely consistent with the classification produced by the use of NOAA-AVHRR data, no quantitative comparison was attempted. Common and differing features of the various information sources were only evaluated visually.









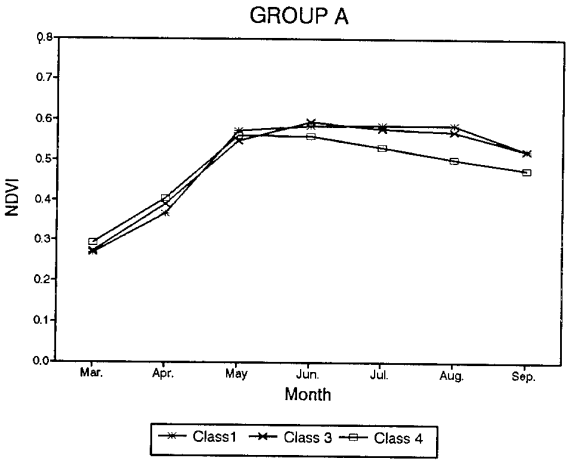
Figures 4A-D - Distribution of the sub-classes in the four ecological zones identified.

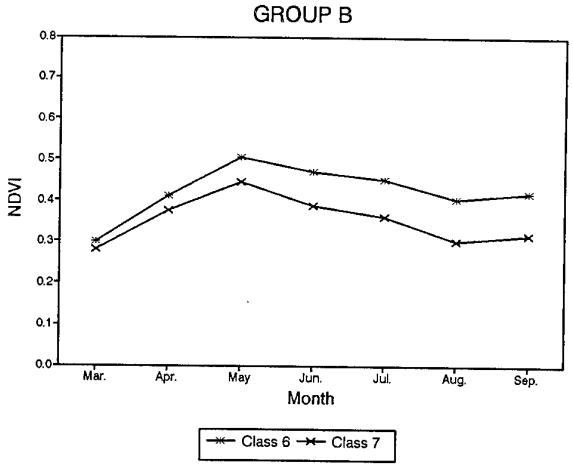
5. RESULTS AND DISCUSSION

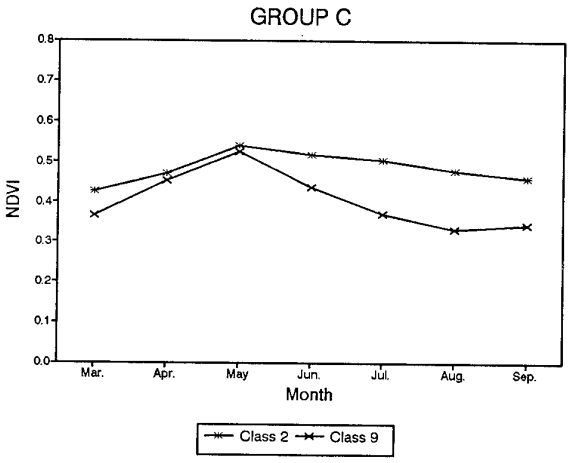
From a first visual evaluation, the classification obtained by the first clustering processes appears globally informative and substantially descriptive of the main environmental zonation of the Region (Figure 3). Mountain zones covered by forests are represented by classes 1, 3, 4. Hilly inner areas fall in classes 6 and 7, while maritime hilly areas are in classes 2 and 9. Finally, agricultural plain areas are represented by classes 5, 8 and 10.

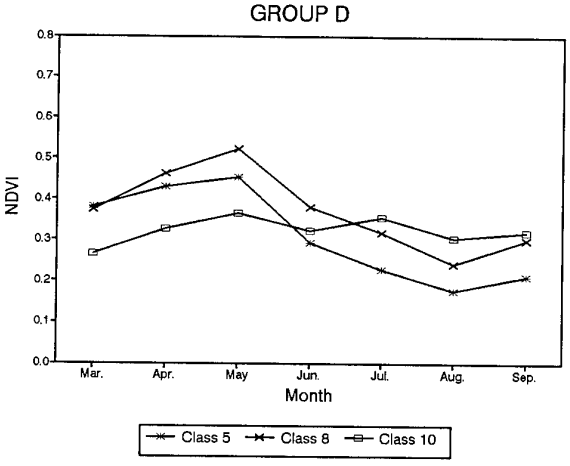
As previously described, all these classes were further subdivided into 2 or more sub-classes. This grouping was then compared to and to existing cartography. The final output will be therefore discussed considering the four main ecological zones mentioned above (forest mountain land, inner hilly areas, maritime hilly areas, agricultural plain land) in comparison with these reference data. These four zones, divided into classes and sub-classes, are shown in Figures 4A-D.

In the first group (A) classes 1, 3 and 4 are included, which represent the mountain zones belonging to the Alps and Appennine chains chiefly located in the inner part of the Region. They are prevalently covered by forests and, subordinately, by pastures. The relevant NDVI profiles, shown in Figure 5A, are typical for the areas where temperate forests dominate. The maximum NDVI values are in June and July, when green vegetation reaches its full development, but these values remain high for the whole growing season until September, which indicates that the classes are not markedly affected by the summer aridity. The geographical distribution of these classes strongly resembles that of areas with low annual temperature, 0-1 dry month and more than 3 cold months in the climatic classification of Rapetti and Vittorini (Ref 11).









Figures 5A-D - Average NDVI profiles of the ten main categories for the two study years grouped into the four ecological zones.

The second group (B) corresponds to hilly inner areas which belong to classes 6 and 7. The land use is in this case mixed agricultural and forestry. The NDVI profiles show their maximum in May for both classes; after that the values diminish through the rest of the season (Figure 5B). This feature, particularly accentuated for the southern class (9), is clearly related to the influence of the long Mediterranean arid season. These areas have a more arid climate, with rainfall which has two maxima in spring and in autumn and is very low for the rest of the year. The climatic classification considers these as zones with medium temperature, 1-2 dry months and 2-3 cold months.

The third group (C) is represented by classes 2 and 9, which are hilly areas near the coastline. The land use is very similar to that of the previous group. The main difference with respect to this is connected with a more maritime climate, which leads to more moderate variations in NDVI during the season (Figure 5C). The climate in these areas is typically Mediterranean with two rainfall maxima and a long arid season in summer. The areas are climatically classified as having medium temperature, 2-3 dry months and less than 2 cold months.

In the fourth group (D), classes 5 and 8 belong to the predominantly plain areas of the Region mainly located near the coastline. Vegetation activity is at a maximum in May, after which the NDVI values decrease reaching a minimum in August (Figure 5D). Class 5 has typically low NDVI values probably because of the scarce rainfall in these zones during the whole growing season. Class 10, which corresponds to plain zones along the course of the Arno river, has also been included in this group. In this case the NDVI profiles maintain very low nearly constant values for all the growing season due to the predominance of urban areas and the presence of mixed winter and summer crops. This is another case in which the present classification has a notable similarity with that of Rapetti and Vittorini (Ref 11). These are areas with medium-high temperature, 2 or more dry months and no cold month.

The reclassification of the main classes further divided the areas chiefly according to their climatic and land use patterns. As examples, the sub-classes of three representative classes (1, 5 and 10) are here described. The four sub-classes of class 1 shows particularly interesting features (Figure 4A). Sub-class 3 corresponds to the valleys around the principal rives of Northern Tuscany and is well differentiated from the other sub-classes which cover the surrounding reliefs. The highest mountains (Apuane and Appennine) corresponds to sub-class 4, while sub-classes 1 and 2 cover the medium-high mountain zones towards North and South, respectively. The land use of the lower areas is mixed agricultural and forestry, while forests and pastures dominate in the upper zones. In the case of class 5, the separation of plain areas near the coast (sub-class 1) and inner plain-hilly areas (sub-class 2) is evident (Figure 4D). In a similar way, the reclassification of class 10 separated the plains near the coast from the inner areas mainly according to the different land use. The land use of sub-class 1 is mixed urban and agricultural, that of sub-class 2 is prevalently agricultural and sub-class 3 represents the hilly areas around the river valleys.

6. CONCLUSIONS

Eco-climatic classifications have conventionally been produced by extrapolating local measurements through usual methodologies. This was generally the only means of obtaining environmental information for long periods on extensive areas before the advent of operational remote sensing techniques, which can now take synoptic information on vegetation conditions at various spatial scales and temporal frequency. NOAA-AVHRR data, in particular, can be used to monitor vegetation type and status on a regional scale. This property has been used in the present research to yield an eco-climatic classification of a complex Region in Central Italy.

From a methodological viewpoint, this process was fraught with problems. A first effort was devoted to the selection of a suitable data set. To this end, not only vegetation index data were considered, as in many similar works, but also thermal infrared images from different seasons to supply significant information on land surfaces. Pre-processing methods were then utilized to uncorrelate the vegetation index data and introduce some factors of spatial homogeneity into the subsequent clustering process. A major difficulty derived from the identification of the most suitable number of clusters with respect to the study data set. An original methodology was therefore developed based on the intuition that a natural clustering should remain stable across several random repetitions.

Through these operations, a visually appealing classification was achieved. The evaluation of its agro-ecological value was necessarily performed indirectly, since the scale of the obtained map prevented other types of checking. A comparison of the map with the available climatic information indicated that the former substantially identified different homogeneous eco-climatic zones. Land use patterns were also influential, but to a degree that is difficult to assess due to the strong influence of eco-climatic factors on land use. Of course, the practical utility of the map obtained has yet to be fully assessed. This will be possible in the near future by direct comparison with regional managers operating on the territory.

From a technical viewpoint, the potential of supervised classifications could be also evaluated when applied to the same problem. In this case agro-meteorological parameters logically related to the AVHRR signal (beginning of growing season, beginning and end of arid season, etc.) could be estimated in ground stations and extrapolated by an approach similar to that proposed by Maselli et al. (Ref 16) for forest parameters. The agro-meteorological information obtained could then be further processed relying on well known concepts (Ref 3), in order to derive a final classification more directly applicable. Some studies are presently underway in this research direction.

Acknowledgements

The NDVI MVCs for the present work were kindly provided by Nuova Telespazio S.p.A. (Rome) within the framework of an existing collaboration; this contribution is gratefully acknowledged.

References

1. Koppen, W., 1923, Die Klimate der Erde; Grundriss der Klimakunde. Walter de Gruyter Co., Berlin, 369 pp.

2. Thornthwaite, C., 1933, The climates of the earth. Geographical Review, **23**, 433-440.

3. Conese, C., Maracchi, G., Maselli, F., Miglietta, F. and Romani, M., 1989, ISAP: Integrated System for the Evaluation of Agriculture Productivity: A case study in the Mugello Region. Proceedings of the Conference on The Application of Remote Sensing to Agricultural Statistics, Villa Ponti (VA, Italy), 10-11 October 1989, pp.269-278.

4. Strahler, A.N., 1984, Geografia Fisica, Piccin Ed., Padova (Italy), 303 pp..

5. Susmel, L., 1990, Principi di Ecologia, Fattori Ecologici, Ecosistema, Applicazioni. CLEUP Editore Padova (Italy).

6. Tucker, C.J., Justice, C.O., Prince, S.D., 1986, Monitoring the grasslands of the Sahel, 1984-1985. International Journal of Remote Sensing, **7**, 1271-1318.

7. Prince, S.D., 1991, A model of regional primary production for use with coarse resolution satellite data. International Journal of Remote Sensing, **12**, 1313-1330.

8. Derrien, M., Farki, B., Legle'au, H., and Sairouni, A., 1992, Vegetation cover mapping over France using NOAA-11/AVHRR. International Journal of Remote Sensing, **13**, 1787-1795.

9. Gaston, G.G., Jackson, P.L., Vinson, T.S., Kolchugina, T.P., Botch, M., and Kobak, K., 1994, Identification of carbon quantifiable regions in the former Soviet Union using unsupervised classification of AVHRR global vegetation index images. International Journal of Remote Sensing, **15**, 3199-3221.

10. Benedetti, R., Rossini, P., and Taddei, R., 1994, Vegetation classification in the middle Mediterranean area by satellite data. International Journal of Remote Sensing, **15**, 583-596

11. Rapetti, F., and Vittorini, S., 1994, Carta climatica della Toscana Centro-Meridionale e Insulare. Pacini Editore, Pisa (Italy), 1994.

12. ESA User Guide, 1992, Sharp Level 2A, Earthnet Programme Office, Rome.

13. Genstat, 1993, Genstat 5 Reference Manual, Oxford Science Publications, Oxford, 644 pp.

14. Foody, G.M., 1994, Ordinal-Level Classification of Sub-Pixel Tropical Forest Cover, Photogrammetric Engineering and Remote Sensing, **60**, 61-65.

15. Defries, R.S., and Townshend, R.G., 1994, NDVI-derived land cover classification at a global scale. International Journal of Remote Sensing, **15**, 3567-3586.

16. Maselli, F., Conese, C., De Filippis, T., Norcini, S., 1995, Estimation of Forest Parameters Through Fuzzy Classification of TM Data, IEEE Transactions on Geoscience and Remote Sensing, **33**, 77-84.

Paper 28

P. Boekaerts (BE)

Why did you use NOAA rather than SPOT for vegetation studies?

Author's reply:

The use of NOAA DATA is due to the assessed methodology of data acquisition, which guaranty, for this applicaton, the requested a-priori knowledge of the experimentally measured data, systematic and random fluctuations of different parameters. This fact is, even more requested because the final performances evaluation is based on the comparison with the ground-truth model (assessed mapping) which has a high degree of stability. The eventual estimation of the quality of the conceptual approach, can be considered robust with respect to the experimental input data.

TARGET DETECTION VIA MEASUREMENTS TAKEN BY A TRANSMITTER - INDEPENDENT RECEIVER NETWORK

Dipl. El. Eng. Nikos J. Farsaris

Prof. Peter P. Stavroulakis

Telecommunication Systems Institute

Technical University of Crete.

73133 Chania Crete GREECE

TEL:+30821-46564N FAX:+30821-41920

ABSTRACT:

In this paper the technique of target detection via measurements of target range and Doppler shift taken by a set of cooperating radars in an area is extended for potential use on a Transmitter Independent Receiver Network (T.I.R.N).

The T.I.R.N. is obtained by the exchange of target data between Independent Bistatic Radar Receivers of existing radars or other R.F. transmitters in a specific area. Combined data of a number of receivers unveil the target's location and velocity vector. Efforts have been made to minimize the complexity requirements of individual receivers and thus making cost-effectiveness the basic feature of a T.I.R.N.

This technique is explored for obtaining a detection system that comprises these parameters that are substantial for military use: The silent operation (undetectability) of an Independent Bistatic Radar Receiver, the accuracy of a monostatic radar and the invulnerability of a distributed computer network.

1) Introduction.

1.1) Radio detection issues.

The evolution of radar just before WW2 changed the battlefield conditions in a way that no other weapon or communication system had changed them before.

Although the first radars were bistatic (the receiver and the transmitter were located in different places) the evolution of microwave electronic elements made the construction of the today's monostatic radars possible.

Despite its ease of use and its operational flexibility the monostatic microwave high-gain E.W. Radar suffers some serious shortcomings listed below:

1. It has a very powerful transmitter that can be detected and destroyed by the enemy.

2. It is vulnerable to enemy ECM because its location is known, and even if not attacked by A.R.M.s its performance might be seriously degraded by means of intensive interference.

3. It can extract only the radial component of the target's velocity (Its projection on the r-t direction as described in the appendix B)

4. The cost of a microwave receiver is a fraction of that of a microwave high

Paper presented at the AGARD SPP Symposium on "Remote Sensing: A Valuable Source of Information", held in Toulouse, France from 22-25 April 1996, and published in CP-582.

power transmitter and it seems plausible to exploit the transmitted signal using only one receiver.

2) The Transmitter - Independent Receiver Network. (T.I.R.N.)

2.1) Definition and consideration of a T.I.R.N.

Let's consider a radar in a given position transmitting in a known frequency band. In fact we can consider everything about the transmitted signal as known to us even if the transmitter is non-cooperative. (There are numerous methods of detecting a transmitter and identifying its signal.)

Thus it is easy to construct a receiver suitable for exploiting the characteristics of the transmitted signal. This receiver can receive and detect the signal directly transmitted by the radar as well as its reflections scattered by any target. This is actually the principle of how the independent bistatic radar receiver works (or the hitch-hiker as described by Dr. Shoenenberger in [2] and [3]).

The Transmitter - Indipendent Receiver Network can be defined as a set of Independent Bistatic Receivers in an area that exchange data in order to achieve better target detection in both terms: Target location and velocity measurement.

The location of the target can be done either by the bistatic receiver methods or using T.D.O.A. between the receivers if more than three are networked and using the same transmitter as a signal source. The target's position can be identified as the intersection point of three hypervoloids (or the gravity center of the volume they include in the case of some error in T.D.O.A measurement.) Here is the equations in brief:

$$|\vec{r}_1|-|\vec{r}_i| = c * \Delta t_{i1}, \quad i=2,3,\ldots,n$$

or if extended

$$\sqrt{(x-x_1)^2+(y-y_1)^2+(z-z_1)^2} - \sqrt{(x-x_i)^2+(y-y_i)^2+(z-z_i)^2} = c * \Delta t_{1i} \qquad (1)$$

where the 1st receiver is used as a reference receiver to the others. Of course any of the receivers can be used as a reference or it might be no reference at all. The reference receiver is implied for mathematical simplicity only. More issues on Bistatic and Multistatic target location can be found on Appendix A and at the References.

If the radar signal has all these characteristics needed for measuring accurate Doppler frequency shift, then the projection of the velocity of a target can on a receiver-target direction can be estimated by this as follows:

$$f_D = \frac{(v_R + v_i)}{\lambda} \qquad (2)$$

In this case v_R and v_i are the projection magnitudes on the transmitter (radar) - target and receiver - target (t-t or r-t) directions respectively, signed positively if the target is approaching or negatively if the target is retreating. That means that from an independent Doppler measurement we can estimate the $\delta=v_i+v_R$ (algebraic) sum.

But we have noted in the appendix B that all the projections **vi** form a sphere with diameter v and center $\Delta(\frac{\vec{v}}{2})$ if considered as bound vectors on a target-centered coordinate system.

Thus supposing that measurements can be taken from three receivers (let them be f_{D1}, f_{D2}, f_{D3} respectively) they can give estimates about $\delta_1, \delta_2, \delta_3$ where

$$\delta_i = v_R + v_i \Rightarrow \delta_i - v_R = v_i \qquad (3)$$

For the receivers then the equations (B6) give:

$$\begin{aligned} &u*\sin\vartheta_1*\cos\varphi_1+w*\sin\vartheta_1*\sin\varphi_1+\zeta*\cos\varphi_1=\delta_1-v_R\\ &u*\sin\vartheta_2*\cos\varphi_2+w*\sin\vartheta_2*\sin\varphi_2+\zeta*\cos\varphi_2=\delta_2-v_R\\ &u*\sin\vartheta_3*\cos\varphi_3+w*\sin\vartheta_3*\sin\varphi_3+\zeta*\cos\varphi_3=\delta_3-v_R \end{aligned} \qquad (4)$$

$$u*\sin\vartheta_R*\cos\varphi_R+w*\sin\vartheta_R*\sin\varphi_R+\zeta*\cos\varphi_R=v_R$$

The last equation being for the Radar (transmitter) site.

A similar equation system can be extracted using four receivers and ignoring the transmitter. (Then the fourth equation would be similar to the first three.)

Knowing the ϑ_i,φ_i and ϑ_R,φ_R from relations previously stated, (actually measuring ϑ_i,φ_i and estimating the corresponding ϑ_R,φ_R), the necessary condition for the system (15) to have a solution is:

$$D_S=\begin{vmatrix} \sin\vartheta_1*\cos\varphi_1 & \sin\vartheta_1*\sin\varphi_1 & \cos\vartheta_1 & \delta_1-v_R\\ \sin\vartheta_2*\cos\varphi_2 & \sin\vartheta_2*\sin\varphi_2 & \cos\vartheta_2 & \delta_2-v_R\\ \sin\vartheta_3*\cos\varphi_3 & \sin\vartheta_3*\sin\varphi_3 & \cos\vartheta_3 & \delta_3-v_R\\ \sin\vartheta_R*\cos\varphi_R & \sin\vartheta*\sin\varphi_R & \cos\vartheta_R & v_R \end{vmatrix}=0$$

(5)

Which is equivalent to:

$$D_S=v_R*D_R-(\delta_1-v_R)*D_1+(\delta_2-v_R)*D_2-(\delta_3-v_R)*D_3=0\Leftrightarrow$$

$$v_R*(D_R+D_1-D_2+D_3)=D_1*\delta_1-D_2*\delta_2+D_3*\delta_3\Leftrightarrow$$

$$v_R=\frac{D_1*\delta_1-D_2*\delta_2+D_3*\delta_3}{(D_R+D_1-D_2+D_3)} \qquad (6)$$

Di in these equations is the minor determinant extracted from D_S by removing the fourth column and the i-th line.

Then solving (4) and replacing v_R we can calculate the real **v**.

In conclusion, we observe that using three receivers matched to the radar signal or three adaptive receivers and knowing the radars position or ignoring it and using four receivers we can obtain full detection of the target in both terms of location and velocity even the transmitter is non cooperative. In fact the transmitter might be an enemy radar!

This is why we can refer the term Transmitter-Independent Receiver Network (T.I.R.N.) as regards to the described detection model. A brief block diagram is shown at Fig.1

Comparing a T.I.R.N. with other means of target detection its advantages become obvious:

1. Being passive it is difficult to be detected.

2. Even if some receivers are detected and destroyed the system may continue its operation.

3. Being multistatic, it may detect Low Radar Cross Section targets because the cross section appeared at the receivers comes from different angles of scattering. In fact this characteristic might be more useful than the other ones.

4. It can be used for monitoring enemy radar operation silently over a large period of time.

2.2) Example of a Two-Receiver T.I.R.N. Model of the Operational System

The 2-receiver model presented here is by no means a full scale operational model but it is presented for its simplicity and avoidance of utilizing extensive mathematical expressions.

If there is an enemy radar at the location $\Pi(x_T,y_T)$ and two of our receivers at $\Delta_1(x_1,y_1)$, $\Delta_2(x_2,y_2)$ respectively (Fig.2), then it is easy to locate the transmitter of the radar using radiogoniometry.

Using the same technique or by measuring the Range Sum of the direct and the

reflected signal we can locate the target. (The Radar and one receiver can always form a bistatic radar with independent receiver as described in [1],[2],[3]and[5]). The technique used for the location depends on the characteristics of the transmitted signal and the directivities of the receiver antennas. For the following analysis, let's suppose that the signal has the characteristics needed for Doppler measurements.

On any 2-dimensional coordinate system for the $\varphi_1, \varphi_2, \varphi_R$ angles (bearings from the receivers and the radar), it is obvious that:

$$\varphi_i = \arctan \frac{y - y_i}{x - x_i} \quad i = 1,2,(R)$$

From if we extract the velocity sums δ_1, δ_2 from the Doppler shifts f_{D1}, f_{D2} from the relations (13),(14) we have:

$$\delta_1 = v_1 + v_R \quad \delta_2 = v_2 + v_R$$

Calculating D_S and solving for $D_S=0$ we get:

$$\begin{vmatrix} \cos\varphi_1 & \sin\varphi_1 & \delta_1 - v_R \\ \cos\varphi_2 & \sin\varphi_2 & \delta_2 - v_R \\ \cos\varphi_R & \sin\varphi_R & v_R \end{vmatrix} = 0 \Leftrightarrow$$

$$v_R * D_R + (\delta_1 - v_R) * D_1 - (\delta_1 - v_R) * D_2 = 0 \Leftrightarrow$$

$$(D_R - D_1 + D_2) * v_R = \delta_2 * D_2 - \delta_1 * D_1 \Leftrightarrow$$

$$v_R = \frac{\delta_2 * D_2 - \delta_1 * D_1}{(D_R - D_1 + D_2)} \qquad (7)$$

where

$$D_R = \cos\varphi_1 * \sin\varphi_2 - \cos\varphi_2 * \sin\varphi_1$$
$$D_1 = \cos\varphi_2 * \sin\varphi_R - \cos\varphi_R * \sin\varphi_2$$
$$D_2 = \cos\varphi_1 * \sin\varphi_R - \cos\varphi_R * \sin\varphi_1$$

and if

$$u * \cos\varphi 1 + w * \sin\varphi 1 = \delta 1 - vR$$
$$u * \cos\varphi 2 + w * \sin\varphi 2 = \delta 2 - vR$$

is the velocity equation system then:

$$u = \frac{(\delta_1 - v_R) * \sin\varphi_1 - (\delta_2 - v_R) * \sin\varphi_2}{D_R} \qquad (8)$$

$$w = \frac{(\delta_2 - v_R) * \cos\varphi_2 - (\delta_1 - v_R) * \cos\varphi_1}{D_R}$$

are the components of the velocity on the r-t level plane.

It is clear that the correction of the angles described in the relations (B-12) is not needed because of the omission of he third dimension (z or ϑ).

For more accurate measurements the ranges between the receivers, the transmitter and the target must be of the same order of magnitude. This is shown by the error analysis that follows.

2.3) Error analysis.

The solution of the velocity components equations led to the relations (19) or their equivalents:

$$u = \frac{v_1 * \sin\varphi_2 - v_2 * \sin\varphi_1}{\cos\varphi_1 * \sin\varphi_2 - \sin\varphi_1 * \cos\varphi_2} \qquad (9)$$

$$w = \frac{v_2 * \cos\varphi_1 - v_1 * \cos\varphi_2}{\cos\varphi_1 * \sin\varphi_2 - \sin\varphi_1 * \cos\varphi_2}$$

Changing the coordinate system in that way so $\varphi_1=0, \varphi_2=\varphi$ we obtain the much simpler equations:

$$u = v_1 \quad w = \frac{v_2 - v_1 \cos\varphi}{\sin\varphi} \qquad (10)$$

Now supposing an error in one velocity measurement we can obtain the overall error for a Doppler ambiguity. For example supposing that we measure $v_1+\varepsilon$ instead of v_1 then:

$$u' = v_1 + \varepsilon$$
$$w' = \frac{u_2 - (v_1 + \varepsilon) * \cos\varphi}{\sin\varphi} = \qquad (11)$$
$$w + \varepsilon * \cot\varphi$$

and the error of the velocity is:

$$\vec{\varepsilon}(\varepsilon, \varepsilon * \cot\varphi),$$
$$|\vec{\varepsilon}| = \varepsilon * \sqrt{1 + \cot^2\varphi} \Leftrightarrow |\vec{\varepsilon}| = \varepsilon * \left|\frac{1}{\sin\varphi}\right| \qquad (12)$$

If instead of v_2 we had measured $v_2+\varepsilon$ then the conclusion would be the same.

In the angular error case things come slightly more complicated. If we measure $\psi'=\psi+\gamma$ instead of ψ (γ being the angular error) then:

$$u = v_1 \quad w' = \frac{v_2 - v_1 * \cos(\varphi + \gamma)}{\sin(\varphi + \gamma)} \qquad (13)$$

If δ is the difference of the magnitudes of the projections then :

$$v_1 = v_2 + \delta \qquad (14)$$

and:

$$w' = v_1 * \left(\frac{1 - \cos(\varphi + \gamma)}{\sin(\varphi + \gamma)}\right) + \frac{\delta}{\sin(\varphi + \gamma)} \qquad (15)$$

and the velocity error is divided in two terms: the inherent system error term ε_i and the direction-dependent term ε_d:

$$\varepsilon_i = v_i * \left(\frac{1 - \cos(\varphi + \gamma)}{\sin(\varphi + \gamma)} - \frac{1 - \cos\varphi}{\sin\varphi}\right) (16)$$

$$\varepsilon_d = \delta * \left(\frac{1}{\sin(\varphi + \gamma)} - \frac{1}{\sin\varphi}\right) (17)$$

(It is obvious that $\delta=v_2-v_1$ defines the direction of the target's movement. The fact that this error is getting bigger when δ is large and the target moves near-tangentially is a pleasant coincidence!).

Both of these terms tend to infinity when ψ tends to zero or π as well as the velocity error does. That means that the error becomes bigger at far or very close distances or if the target approaches the baseline between the receivers.

A simple analysis has shown that for an angular error of 2mrad and a δ/v_1 ratio of 0.1 (10%), the relative error magnitude did not exceed 5% at angles ψ between 0.3 to 2.2 rads (30 to 120 degrees. The coverage area is shown in Fig. 3)

Consideration of both errors combined is given below. Supposing that the measured v_1', v_2', ψ' are different from the real ones :

$$\varepsilon_t = \sqrt{\varepsilon_1^2 + \left(\varepsilon_1 * \frac{1 - \cos(\varphi + \gamma)}{\sin(\varphi + \gamma)} + (\varepsilon_1 - \varepsilon_2) * \frac{\cos(\varphi + \gamma)}{\sin(\varphi + \gamma)} + \varepsilon_i + \varepsilon_d\right)^2}$$

(18)

where $\varepsilon_1=v_1-v_1'$ and $\varepsilon_2=v_2-v_2'$. Again a computer analysis has shown that the error magnitude is of the same order with the angular error magnitude.

3) Conclusions:

Having considered the 2-receiver model the errors seem mathematically significant but in a tactical case become of little importance because even a 50% error will give the information that the target will be directed to a well defined 30 degree sector. This can be determined by a two-receiver only T.I.R.N. from a large distance. If more receivers are combined in a wider area (using perhaps different transmitters in the area as illuminators) then the error will be significantly decreased. In fact there can be an area covered by relatively simple but many networked MTI-only receivers: The only thing to know then is the relative movement of the target (approaching or retreating) but significant estimation about its accurate direction can be made by comparing signs only of the (corrected) Doppler shift. Then only one Doppler measurement can give the actual velocity because if we know the direction and the magnitude of one projection of a vector we know in fact the entire vector.

A combination of Doppler measuring and MTI-only receivers appear as an attractive solution giving the most cost-effective solution: 3 or 4 Doppler measuring receivers if supported by a number of MTI-only direction finders will give outstanding results making T.I.R.N.s an attractive passive alternative to large expensive and vulnerable radars.

Now simulation work is under way in order to verify the theoretical results of this research and eforts have been made to minimize the complexity (and subsequently the cost) of the T.I.R.N. making it a useful system for military and civil use.

References:

1. N. Willis "Bistatic Radars" Artech House 1990

2. Shoenenberger J.G. and Forrest J.R.: "Principles of Independent Receivers for Use with Cooperative Radar Transmitters". Radio Electron. Eng. vol. 52 pp. 93-101 February 1982.

3. Forrest J.R. and Shoenenberger J.G.:Totally Independent Bistatic Radar Receiver with Real Time Microprocessor Scan Correction". IEEE Int. Radar Conf. pp. 380-386 1980

4. Joshua Wurman et. al. "Design of a Bistatic Dual Doppler Radar for Retrieving Vector Winds Using One Transmitter and a Remote Low-Gain Passive Receiver" IEEE Porch. vol. 82 pp. 1861- 1871 1994

5. Mavrokoukoulakis , J. Forrest et. al."Bistatic Radar Studies: A Collaborative Research Programme between KETA (Greece) and UCL (UK)". AGARD Cotract AVP/82/571 Final Report 1984.

6. Yaakov Bar - Shalom (Editor) "Multitarget Multisensor Tracking" vol. 1 "Advanced Applications", vol. 2 " Applications and Advances" Artech House 1989

7. Doviac R.J. and Weil C.M. "Bistatic radar detection of the melting layer" Journal of Appl. Met.11 1012-1016, 1972

Appendix A

Multistatic Location Issues

In a three - dimensional space the location of a point can be determined in many cases as the intersections of three surfaces.

$$S_1 \cap S_2 \cap S_3 \qquad \text{(A1)}$$

or the intersection of a curve and a surface

$$S \cap L \qquad \text{(A2)}$$

In case of the Independent Bistatic Receiver the target detection model can be approached by the relation (A2) in case that the transmitting radar has a pencil - beam transmitting section.

Then the target can be located at the surface of an ellipsoid determined by the time t needed for the wave front to travel from the transmitter to the target ad then to the receiver. (Range Sum model). The equation of this ellipsoid is:

$$r_R + r_r = c * t \qquad \text{(A3)}$$

or

$$\sqrt{(x-x_R)^2+(y-y_R)^2+(z-z_R)^2} + \sqrt{(x-x_r)^2+(y-y_r)^2+(z-z_r)^2} = c * t \qquad \text{(A4)}$$

where $R(x_R, y_R, z_R)$, $r(x_r y_r z_r)$ are the Radar and receiver locations (and subsequently the foci of the ellipsoid). The line L needed for (A2) can be represented by the main beam traveling path.

In the simple case of straight line propagation and supposing that the transmitter and receiver locations are R(0,0,0) and r(D,0,0) the following solution has been determined by Doviac and Weil. ([7], [4])

$$\begin{aligned} x &= \frac{(ct)^2 - D^2}{2(ct - \cos p)} * \cos\phi * \sin\vartheta \\ y &= \frac{(ct)^2 - D^2}{2(ct - \cos p)} * \sin\phi * \sin\vartheta \\ z &= \frac{(ct)^2 - D^2}{2(ct - \cos p)} * \cos\vartheta \end{aligned} \qquad \text{(A5)}$$

where φ and ϑ are the angular transmitter - centered spherical coordinates of the beam axis and p is the angle between the transmitter - target and transmitter - receiver (baseline) directions. It is easily figured that

$$\cos p = \cos\varphi * \cos\vartheta \qquad \text{(A6)}$$

as shown in. Fig A-1

In case of an Independent Bistatic Receiver using the illumination of a non - cooperative radar, the D is known and φ and ϑ can be algorithmically estimated after a prolonged transmission observation, or they can be replaced by their receiver - centered values if they are available.

In case of the proposed T.I.R.N. this is not the best target location model. If we take a closer glance then we can see that the errors in measurements are greater near the baselines. And using an enemy radar as illuminator these directions are operationally critical. An expanded T.I.R.N. of course can process the target location data ((x,y,z) vectors) taken by all its receivers and give greater accuracy: If V_1, V_2, ..., V_n are the ambiguity volumes of the received target location vectors then their intersection

$$V_1 \cap V_2 \cap \ldots \cap V_n \qquad \text{(A7)}$$

represents a smaler ambiguity volume.

Another location model is the T.D.O.A. model. Supposing that we have deployed more than four operational receivers in an area and a signal (a pulse or an electromagnetic transient) is received.

Then using one receiver as a "reference receiver" (lets say the first receiver randomly chosen) we can form the equations (as (A1) suggests):

$$\sqrt{(x-x_1)^2+(y-y_1)^2+(z-z_1)^2}- \qquad \text{(A8)}$$
$$\sqrt{(x-x_i)^2+(y-y_i)^2+(z-z_i)^2}=c*t_{1i}$$

where t_{1i} is the time difference of arrival of the transient at the receiver r_1 and r_i respectively and r_i are the receiver locations. These equations represent hypervoloid sectors,and the target is expcted to be at their intersection. This method has the advantage of not using the transmiter location. I fact the tracsmitter can be the taget itslelf.

Care must be taken that at least four receivers are not located at the same planar surface. In that case there exist, at least, three hypervoloid sectors that have only one intersection point. If no more than three receivers are on any planar surface then there are $\binom{n-1}{3}$ combinations of receivers that can give adequate detection data.

Examining the T.D.O.A. method, we must note that the distances between the receivers must be several orders of magnitudes greater than the received distance traveled by the transient signal at the time it happens e.g. if the illuminator radar transmits pulses of duration τ then the distances between the receivers must be much greater than $c*\tau$. This will give adequate resolution at close distances (comparable to the distances between the receivers) but the location shells will always be finite. Targets being far away can not be located precisely but only directionwise by the asymptotes of the hyperboloids. The T.I.R.N. is then acting as a symbolometric antenna system.

All the above have led to the conclusion that Range Sum method and T.D.O.A. can be used simultaneously in order to correct each other in any case, close or distant location.

Summarizing the above consideration we can conclude that:

- In close distances (in comparison to he receivers' baselines) T.D.O.A. can adequately locate possible targets.
- In large distances Range Sum is more precise especially if we use the T.D.O.A. Symbolometry Technique to detect the direction of the target from the T.I.R.N. receivers.

Appendix B

Velocity vector synthesis by adequate number of linearly independent projections

B.1) Definition of the problem

It is known that a monostatic radar can measure the range and estimate the radial velocity of a target via a Doppler shift measurement. The radial velocity is the projection of the velocity vector **v** on the line connecting the radar and the target.

It is also know that we can compose a vector knowing two (on a plane) or three (in space) linearly independent projections of it. So if we know the elevations, azimuths and the Doppler shifts of a target measured by three radars we can locate it and estimate its velocity.

Now let's define this problem on one plane firstly and then expand our analysis in a three-dimensional space. It is obvious that if we know the directions (or elevations and azimuths) of the target from two radar sites then we know exactly its position. This is the intersection point of the radar-target directions (r-t directions). (Fig. B-1)

If we know the projection of the target velocity on two directions then we know the velocity on the plane the two radars ad the target defines using simple vector analysis.

We can define a coordinate system on the radars-target plane (r-t plane). Then the projections of the velocity can be analyzed to components parallel to the axes u_1, w_1 for the $\mathbf{v_1}$ measured by the radar R_1 and u_2, w_2 for the $\mathbf{v_2}$ measured by the R_2 respectively. Then it is obvious that:

$$\begin{array}{ll} u_1 = v_1 * \cos\alpha_1 & u_2 = v_2 * \cos\alpha_2 \\ w_1 = v_1 * \sin\alpha_1 & w_2 = v_2 * \sin\alpha_2 \end{array} \qquad \text{(B1)}$$

Where α_1, α_2 are the bearings on the r-t plane according to the defined coordinate system.

Knowing that $\mathbf{v_1}, \mathbf{v_2}$ are the projections of the vector **v** on the r-t directions then it's easy to recompose it:It is the vector from the coordinates' center to the intersection point of two lines' perpendiculars to $\mathbf{v_1}, \mathbf{v_2}$ that conclude the points with coordinates (u_1, w_1) and (u_2, w_2) respectively. (Fig. B-2) These two lines can be represented by the linear equation system:

$$\begin{array}{l} u * u_1 + w * w_1 = u_1^2 + w_1^2 = v_1^2 \\ u * u_2 + w * w_2 = u_2^2 + w_2^2 = v_2^2 \end{array} \qquad \text{(B2)}$$

A solution of the problem may be obtained by solving this system and replacing u_1, w_1, u_2, w_2 from the relations (B1).

In the 3-D space the location is defined again by the intersection point of two of the three r-t directions while the velocity is known by its projections on three r-t directions. (Fig.B-3)

Choosing a target-centered coordinate system any of these projections (let them be v_1, v_2, v_3 or v_i, i=1,2,3)can be represented by its components (u_i, v_i, ζ_i), i=1,2,3 through the relations:

$$\begin{array}{l} u_i = v_i * \sin\vartheta_i * \cos\varphi_i \\ w_i = v_i * \sin\vartheta_i * \cos\varphi_i \\ \zeta_i = u_i * \cos\vartheta_i \qquad i = 1,2,3 \end{array} \qquad \text{(B3)}$$

Now perpendicular level planes to $\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}$ can be composed and in a manner similar to this already considered, the system to be solved is:

$$u * u_i + w * w_i + \zeta * \zeta_i = u_i^2 + w_i^2 + \zeta_i^2 = v_i^2 \quad i = 1,2,3 \qquad \text{(B4)}$$

B.2) Solution of the defined problem.

The linear system (B4) with the obvious simplification turns out to its equivalent:

$$\begin{aligned} u * u_1 + w * w_1 + \zeta * \zeta_1 &= v_1^2 \\ u * u_2 + w * w_2 + \zeta * \zeta_2 &= v_2^2 \\ u * u_3 + w * w_3 + \zeta * \zeta_3 &= v_3^2 \end{aligned} \qquad \text{(B5)}$$

The above system may have a unique or more than one solutions; it cannot be impossible because it is the analysis of a real problem. Replacing the components of the projections in system (B5) by their equivalents in the relations (B3) we get the following linear system:

$$\begin{aligned} u*\sin\vartheta_1*\cos\varphi_1 + w*\sin\vartheta_1*\sin\varphi_1 + \zeta*\cos\varphi_1 &= v_1 \\ u*\sin\vartheta_2*\cos\varphi_2 + w*\sin\vartheta_2*\sin\varphi_2 + \zeta*\cos\varphi_2 &= v_2 \\ u*\sin\vartheta_3*\cos\varphi_3 + w*\sin\vartheta_3*\sin\varphi_3 + \zeta*\cos\varphi_3 &= v_3 \end{aligned} \qquad \text{(B6)}$$

with its associated determinant D:

$$D = \begin{vmatrix} \sin\vartheta_1*\cos\varphi_1 & \sin\vartheta_1*\cos\varphi_1 & \cos\vartheta_1 \\ \sin\vartheta_2*\cos\varphi_2 & \sin\vartheta_2*\cos\varphi_2 & \cos\vartheta_2 \\ \sin\vartheta_3*\cos\varphi_3 & \sin\vartheta_3*\cos\varphi_3 & \cos\vartheta_3 \end{vmatrix} \qquad \text{(B7)}$$

If $D \neq 0$ then the system has an unique solution $\mathbf{v}(u,w,\zeta)$ which is the velocity of the target.

If $D = 0$ that means that the lines or the columns of the determinant are linearly dependent. It is worth to examine the following cases:

a) If two lines are directly linearly dependent for example:

$$\begin{aligned} \sin\vartheta_1 * \cos\varphi_1 &= a * \sin\vartheta_2 * \cos\varphi_2 \\ \sin\vartheta_1 * \sin\varphi_1 &= a * \sin\vartheta_2 * \sin\varphi_2 \\ \cos\vartheta_1 &= a * \cos\vartheta_2 \end{aligned} \qquad \text{(B8)}$$

then

$$\vartheta 1 + \vartheta 2 = \pi \,,\ \varphi 1 \pm \varphi 2 = \pi \text{ or } \vartheta 1 = \vartheta 2 \,,\ \varphi 1 = \varphi 2 \qquad \text{(B9)}$$

That means that the target comes through or moves on the line connecting the two radars (R_1 and R_2 in this case).

b) Any other case of linear dependence between the lines of the determinant leads to the conclusion that the target comes through or moves on the level plane the three radars define.

Combining cases (a) and (b) it must be noted that (a) is a subcase of (b). Then the system (B6) has infinite solutions with one independent variable. In such a case we can always measure the projection of the velocity on the three radars plane combining two linearly independent equations of the system (B6) and the equation of the radars plane that is:

$$\begin{aligned} \vec{v}_1 &= \mathbf{a} * \vec{v}_1 + \mathbf{b} * \vec{v}_2 \\ \mathbf{a} * \mathbf{u} + \mathbf{b} * \mathbf{w} - \zeta &= 0 \end{aligned} \qquad \text{(B10)}$$

This must be satisfied by the measured $\mathbf{v_i}(u_i, w_i, \zeta_i)$ of the two radars chosen (an equation is adjusted to a radar) which leads to

$$\frac{\zeta_1 * w_2 + \zeta_2 * w_1}{u_1 * w_2 + u_2 * w_1} * u + \frac{\zeta_1 * u_2 + \zeta_2 * u_1}{u_1 * w_2 + u_2 * w_1} * w = \zeta \qquad \text{(B11)}$$

c) Finally the case that the system has infinite solutions with two independent variables leads to the conclusion that the three radars and the target are aligned. This not desirable in our case and the radar should not be placed in line because they must form a level plane of reference.

From the above analysis we must note :

A) The locus of the projection ends of a bound vector $\mathbf{v}(u,w,z)$ in any direction is a sphere with center: $\Delta\left(\frac{u}{2}, \frac{w}{2}, \frac{\zeta}{2}\right)$

and radius v/2 (or diameter equal to the magnitude of the vector).

Actually (Fig. 4) if $\overline{O}\vec{A}$ is a bound vector with A(u,w,z) and $\overline{OA}'$ its projection a random direction then by definition $O\hat{A}'A$=90deg hence A' belongs to a sphere with center $\Delta\left(\frac{u}{2},\frac{w}{2},\frac{\zeta}{2}\right)$ and diameter |OA|.

For any other point A" of the sphere it is obvious that $O\hat{A}''A$ =90deg , so $\overline{OA}''$ is the projection on its direction of $\overline{O}\vec{A}$.

B) There are limitations in the values of the angles of the projections: The target and any two of the three considered radars define a level plane in the 3-D space. That can be explicitly formalized as follows:

$$\vec{v}_1 = c_1 * \bar{R}_1\vec{T} \quad \vec{v}_2 = c_2 * \bar{R}_2\vec{T}$$

and if $\bar{R}_1\bar{R}_2 = \vec{b}_{12}$ there are real numbers p, q, r not all equal to zero so that

$$p * \vec{v}_1 + q * \vec{v}_2 + r * \vec{b}_{12} = 0$$

where $R_1(x_1,y_1,z_1)$, $R_2(x_2,y_2,z_2)$ are the locations of the radars and T(x,y,z) is the location of the target.

Then analyzing the vectors to their components we get:

$$p*v_1*\cos\varphi_1*\sin\vartheta_1 + q*v_2*\cos\varphi_2*\sin\vartheta_2 + r*b_{12}*\cos\varphi_{12}*\sin\vartheta_{12} = 0$$
$$p*v_1*\cos\varphi_1*\cos\vartheta_1 + q*v_2*\cos\varphi_2*\cos\vartheta_2 + r*b_{12}*\cos\varphi_{12}*\cos\vartheta_{12} = 0$$
$$p*v_1*\cos\vartheta_1 + q*v_2*\cos\vartheta_2 + r*b_{12}*\cos\vartheta_{12} = 0$$

which leads to the conclusion:

$$D_{12} = \begin{vmatrix} \cos\varphi_1*\sin\vartheta_1 & \cos\varphi_2*\sin\vartheta_2 & \cos\varphi_{12}*\sin\vartheta_{12} \\ \cos\varphi_1*\cos\vartheta_1 & \cos\varphi_2*\cos\vartheta_2 & \cos\varphi_{12}*\cos\vartheta_{12} \\ \cos\vartheta_1 & \cos\vartheta_2 & \cos\vartheta_{12} \end{vmatrix} = 0 \quad \text{(B12)}$$

Similar analysis for any other radar pair can lead to:

$$D_{i,j} = \begin{vmatrix} \cos\varphi_i*\sin\vartheta_i & \cos\varphi_j*\sin\vartheta_j & \cos\varphi_{ij}*\sin\vartheta_{ij} \\ \cos\varphi_i*\cos\vartheta_i & \cos\varphi_j*\cos\vartheta_j & \cos\varphi_{ij}*\cos\vartheta_{ij} \\ \cos\vartheta_i & \cos\vartheta_j & \cos\vartheta_{ij} \end{vmatrix} = 0 \quad \text{(B13)}$$

$\forall i,j \in \{1,2,3\}, i \neq j$

which is the limitation that can be used for correction in case of angular error.

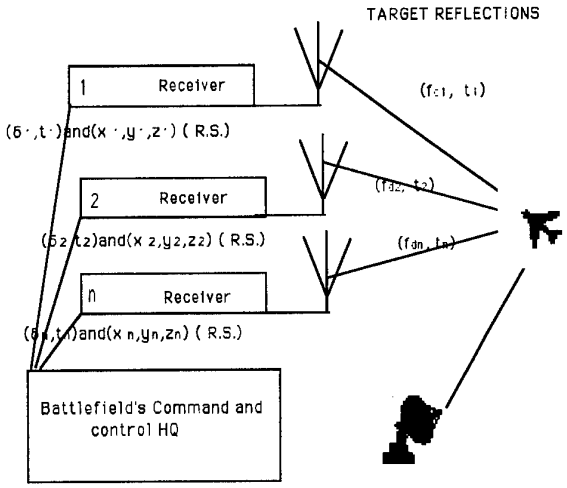


Fig. 1 Simplified block diagram of a T.I.R.N.

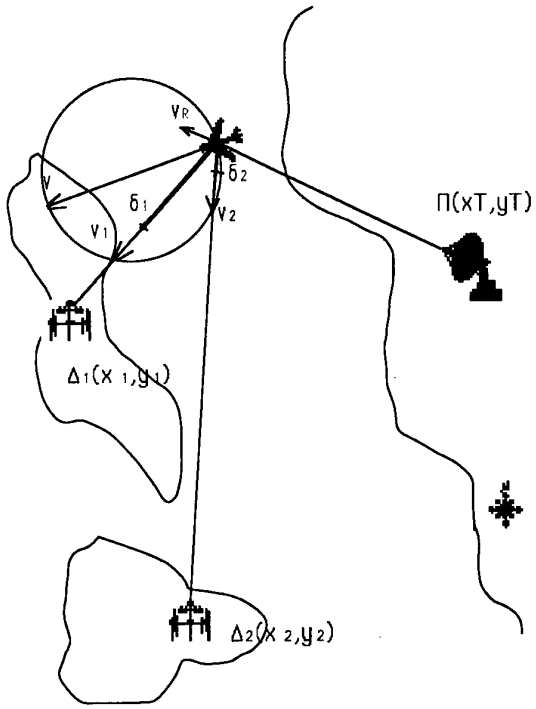


Fig. 2 Example of a 2-receiver T.I.R.N.

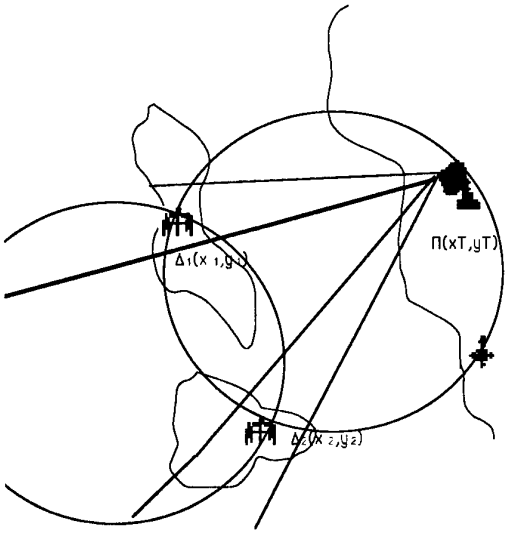


Fig. 3 The full detection region a T.I.R.N. can cover.

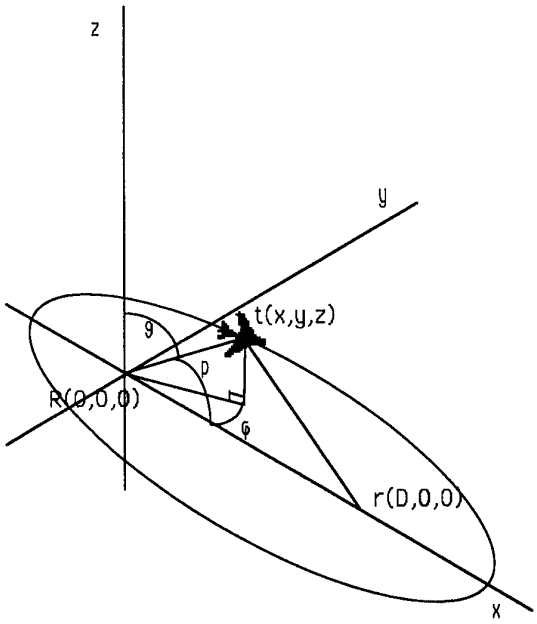


Fig. A-1: Target location by the Range Sum method.

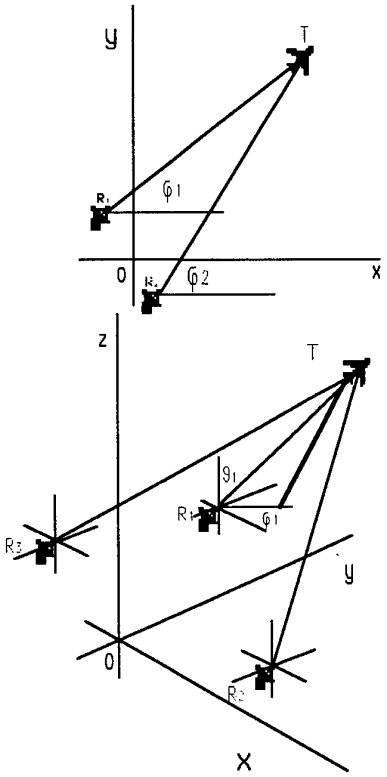


Fig. B-1: Example of target location using 2 or 3 direction finding radars.

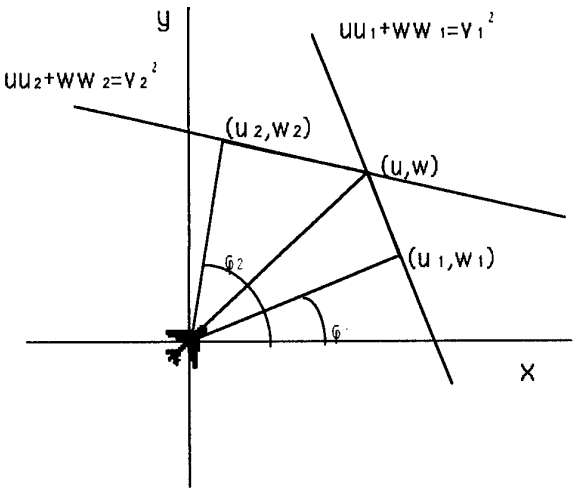


Fig. B-2 Velocity synthesis from 2 projections

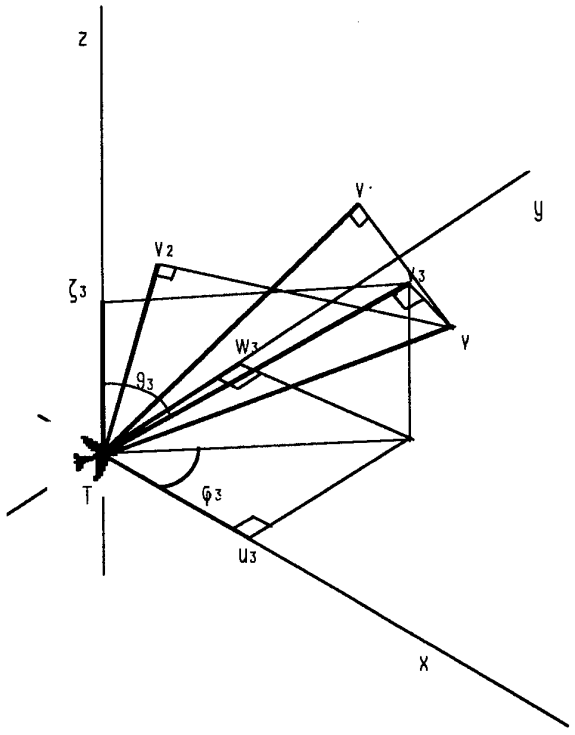


Fig.B-3 : Velocity synthesis uzing 3 doppler measuring radars

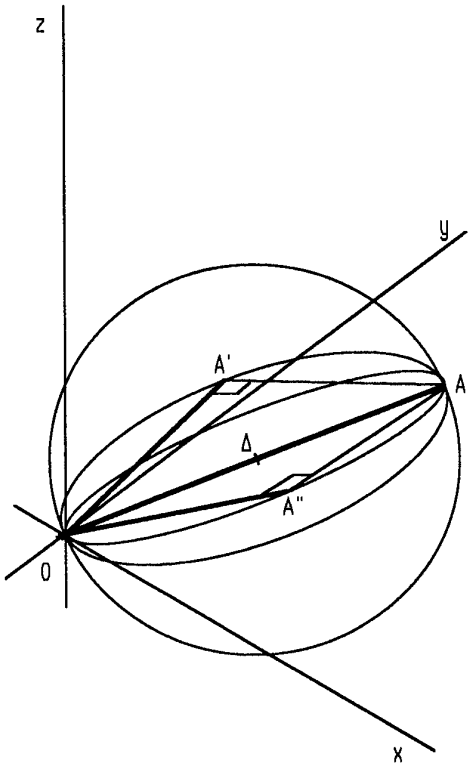


Fig. B-4 : Sphere of the projections of a bound vector or"velocity sphere" of a radar net.

Paper 29

Salmon (BE)

Your system seems very promising because it can be used both in the passive and active modes. Furthermore, you stated that the geometry is very important for a correct measurement. Can you explain how the network geometry of the receivers in actual use will be optimized for maximum accuracy of the target position estimate.

Author's reply:

Well the best geometry for Doppler is when the target is in a position such that the r-t ranges are in the same order of magnitude with the ranges between the receivers. There is no optimal geometry for the location because in long ranges the range-sum method is accurate while in close ranges T.P. O.A. is better (smaller ambiguity volume). The combination of both can povide an accurate location in all cases. (App. A-1).

Short-Term Persistence of Low Cloud Features

Roger A. Helvey
Geophysics Branch, Code 521420E
Naval Air Warfare Center, Weapons Division
Point Mugu, CA 93042-5001

1. SUMMARY

Considerable structure is present in low cloud features seen in weather satellite images of the region offshore southern California. From geostationary imagery, it is evident that translation of this structure is largely responsible for the variability in cloudiness experienced at any fixed point. This report describes use of an interactive computer technique for tracking cloud features, to determine the transport field. Statistics related to the persistence of cloud features are obtained, to assess the possible value of extrapolation of pre-existing patterns in forecasting conditions at any fixed point for a few hours.

2. INTRODUCTION

The northeastern Pacific Ocean is usually partially covered by "fair-weather" stratus and stratocumulus clouds imbedded within the marine layer, for example as seen in Figure 1, a GOES-7 4km visual image for 1130 PST 23 August 1993. Missile tests over the Sea Range west of the Naval Air Warfare Center, Point Mugu (NAWCWPNS) often require visual monitoring of flight operations or ship traffic from aircraft, and are hampered by the presence of the low-level cloudiness. Operations could benefit from successful short-term predictions of cloud distribution.

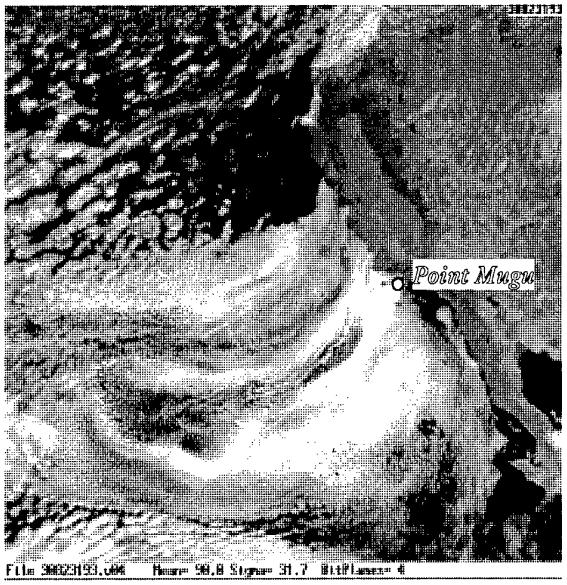


Figure 1. Low cloudiness over northeastern Pacific Ocean west of Point Mugu, California.

Observed from geostationary weather satellite, the low cloud patterns typically persist for many hours. These patterns move generally in a uniform manner indicating transport by the marine layer wind field, although gravity waves and other phenomena may also be present (especially in the vicinity of the coast). The slow-changing nature of the cloud patterns suggests the possibility of predicting cloud conditions for at least a few hours in advance, using mostly kinematic techniques. This paper describes some preliminary results from an attempt to extrapolate low cloud features offshore southern California for a period from the August 1993 Variability Of Coastal Atmospheric Refraction (VOCAR) experiment, using GOES-7 imagery. Statistics are obtained to estimate accuracy of the approach, as related to time and space scales involved in the extrapolation process.

3. METHOD

The wind field used to advect cloud features was obtained from observed motions of the clouds themselves. A program which enabled interactive assessment of cloud motion vectors was written for an Amiga 4000 computer, for use with sequential imagery from geostationary satellites. Small rectangular test areas around distinct cloud features of interest were manually selected using a computer "mouse". The field of view within these areas was animated to alternate between source and target images separated in time by one-half hour or more. Cursor (arrow) keys were used to move the corresponding area on the current source image until local alignment of cloud features was achieved. In this manner individual displacement vectors were determined and recorded for the center of each test area selected on a given source-to-target image pair. The irregularly located vectors were then interpolated to a grid point array using a Cressman successive approximation algorithm. As new vectors were added, the gridded analysis could be periodically recomputed, and updated first guess alignments applied to the current comparison region indicated by the mouse position, to improve the efficiency of the interactive process.

4. APPLICATION

The interactive technique was applied to full-resolution (1-km sub-satellite) GOES-7 visible images at half-hourly intervals from 0830 to 1430 Pacific Standard Time, 23 August 1993. Sectors consisting of 640 by 500

Paper presented at the AGARD SPP Symposium on "Remote Sensing: A Valuable Source of Information", held in Toulouse, France from 22-25 April 1996, and published in CP-582.

pixels were extracted from the full disk images, covering a portion of southern California and offshore waters roughly 500 by 400 km in extent, as depicted in Figure 2. Also shown are three comparison regions (designated A, B, and C) chosen for statistical analysis. The cloudiness apparent offshore is typical of this region at the eastern edge of the Pacific Ocean, which is dominated by high pressure and fair weather conditions. Subsiding warm, stable air caps a cooler marine layer usually 200 to 600 meters deep. Within the marine air mass widespread stratiform cloudiness is often present, with individual features slowly evolving while being transported by the low level winds, which generally prevail from the northwest. Wind speeds are commonly around 10-14 mps over the open ocean, but tend to weaken to 5-8 mps over coastal waters. Thus low cloud displacements of roughly 20 to 40 km per hour are characteristic, and individual features, if persistent, might possibly be tracked for up to 12 daylight hours in the visible imagery over the domain represented here. Cloud definition is poor due to low illumination during hours just after sunrise and before sunset, and the low clouds tend to diminish during the afternooon. Therefore the six-hour interval from 0830 to 1430 PST was chosen as the best period to follow cloud patterns during their transit across the domain.

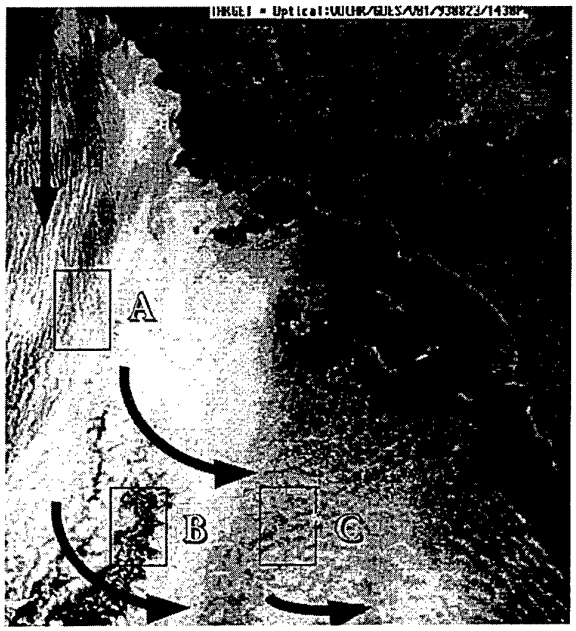


Figure 2. GOES-7 1km VIS image for 1430PST 23 August 1993 showing 64x64 pixel analysis regions A. B. C off southern California. Region A is in an area containing linear stratocumulus elements suggesting strong winds and shear. Region B contains cumuliform elements bordering an area of broken clouds where winds are turning cyclonically. In region C winds are light, with very slowly changing polygonal cloud elements.

Cloud displacements were obtained in a backwards direction from successive "destination" images to the previous half-hour image. The backward trajectory method assures that each pixel in the destination test regions can be identified with a pixel in the various prior source images, thus minimizing distortions which might otherwise arise due to the need to interpolate to fill in gaps resulting from areas of divergent flow. Several dozen vectors were obtained by matching features for the cloudy portions of each pair of images, then interpolated to a regular 36 by 28 element grid covering the image domain. After the entire set of images was processed, the gridded data were used to estimate cumulative displacements for each pixel in the three test regions, from successive source images starting at 0830 PST to a fixed destination image at 1430 PST. Figure 3 illustrates the translation and deformation associated with movement of the cloud features, showing the computed upwind boundaries of the test regions at 1130 PST subsequently arriving in the destination regions at 1430 PST as depicted in Figure 2.



Figure 3. Same as Figure 2, but for 1130PST 23 August 1993. Computed upwind boundaries of regions A, B and C are shown..

5. RESULTS

To assess the fidelity of the extrapolated imagery, in terms of capability to predict cloud coverage for specified regions and times, differences between coverage were calculated between the source and destination images. Cloud amount was estimated by relating the raw digital count value transmitted by the satellite for each pixel, to minimum and maximum count values which were established empirically for each image time to enable contrast stretching compensation

	REGION A							REGION B							REGION C						
	Computation Cell Sizes [pixels]							Computation Cell Sizes [pixels]							Computation Cell Sizes [pixels]						
Source-Dest. Lag	0001	0004	0016	0064	0256	1024	4096	0001	0004	0016	0064	0256	1024	4096	0001	0004	0016	0064	0256	1024	4096
1330-1430PST 1hr	10	10	9	8	8	7	3	11	9	6	4	3	2	1	6	5	4	4	4	3	3
1230-1430PST 2hr	15	15	14	14	13	13	13	14	12	9	6	4	3	1	11	10	10	10	10	10	10
1130-1430PST 3hr	20	20	19	19	18	18	18	17	16	13	9	7	4	4	17	17	17	17	17	17	17
1030-1430PST 4hr	24	23	23	23	22	22	22	20	18	15	12	9	3	2	21	21	20	20	20	20	20
0930-1430PST 5hr	27	26	26	26	25	25	25	18	17	14	12	10	5	5	29	29	29	29	29	29	29
0830-1430PST 6hr	32	32	31	31	31	31	31	19	18	16	13	11	8	8	31	31	31	31	31	31	31

Table I. Variation of extrapolated cloud cover for 6 time lags, 7 size scales, and 3 regions off southern California on 23 Aug 93. Each entry in the Table indicates the change in cloud amount after displacement along trajectories from source to destination image, for the given time lag, computation cell size, and region. The values are mean absolute deviations between all source and destination cells in the same category, expressed as percent of variation between clear and overcast. A value of 0% would signify complete agreement between source and destination conditions, while a value of 100% would indicate complete disagreement. Time lags denote the 1 to 6 hour differences between a source image (0830, 0930, 1030, 1130, 1230 & 1330PST) and the destination image (always 1430PST). Sizes are areas of computation cells (in pixels) subdividing comparision regions. Regions A, B, and C are located in Figure 2. Each region consists of 4096 pixels, divided into 1 to 64 cells of 4096 to 64 pixels each, respectively, corresponding to the sizes labeled.

for varying illumination with time of day. Mean absolute deviations between source and destination cloud amounts were computed, and are given in Table I. Each entry in the table measures the persistence in cloud amount after displacement along trajectories from source to destination images, for the indicated categories of time lag, scale, and region. The values are expressed as percent of variation between clear and overcast, thus ...100% would indicate complete agreement and 0%... complete disagreement between predicted and verified conditions. Time lags of 1 to 6 hours correspond to the time differences between source and destination images. The size category refers to areas of cells subdividing the test regions, for the purpose of measuring effects of spatial scale of cloud features and motion vector precision on cloud predictability.

The best performance was indicated for Region B, which included a large region of only scattered or broken cloudiness surrounded by overcast. The other regions examined had somewhat poorer results, probably because larger, more persistent features were less evident. The effect of scale shows in the trend for better scores for the larger subdivisions. Smaller cells are more vulnerable to imprecise motion vectors, and by resolving clouds at scales more likely to be affected by local dynamics within the marine layer. The degradation in the accuracy of the extrapolations with time shows clearly, as would be expected due to aspects of the cloud evolution not explained by simple low-level wind trajectories, and due to the difficulty in reliably recognizing features to properly determine motions over longer time periods. Large changes in cloud cover and brightness were sometimes evident due to gravity waves rapidly translating across the marine layer, with no relation to the motion of the individual clouds. Another factor, especially at lower illumination levels in morning and afternoon, was the effect of "striping" and other imperfections in the data from the satellite upon ability to recognize details in the imagery.

6. CONCLUDING REMARKS

There are practical limits to the size of features which can be predicted. Time changes in clouds are generally more rapid at smaller spatial scales, and individual clouds on the order of a few pixels across can appear or disappear completely within less than an hour. The ability to precisely describe cloud motions is especially difficult on the smallest scale because of the complexity of local wind fields and interaction between adjacent cloud circulations, and problems in correctly matching features between images. Screening by higher clouds when present of course precludes use of the motion extrapolation technique for lower clouds. Cloud changes from gravity waves can cause ambiguity and incorrect assignment of motion vectors, and the presence of complex eddying, sea breeze and terrain effects will necessitate special methods in coastal zones. The method used here accumulated inaccuracies when compiling the net displacements over several hours from the motion vectors between successive images. The method will be improved by adding a second pass to verify and refine displacements over the longer intervals, and by automating feature matching with a cross-correlation technique. Also, other factors contributing to changes in cloud cover will be investigated, including those associated with gravity waves, day-night variations in air temperature, sea breeze divergence near the coast, and terrain-induced waves and eddies.

Paper 30

K. Anderson (US)

Have you considered using large scale features?

Author's reply:

There have been some other attempts to make use of cloud persistence to predict cloud distributions by extrapolation in time, based on either cloud motion vectors or numerical model gridded winds to describe the transport field. Those of which I am aware have emphasised larger scale features and motion fields than in my study so far. One example is a cloud advection technique developed by Mr. Roland Nagle on contract to the Naval Research Laboratory in Monterey, California, and with whom we are working to demonstrate an operational capability for navy applications. In my study, I am first interested in establishing the predictability of the smallest, most perishable features. Of course it is evident that accuracy improves at larger scales, but for the best predictions at any given location, all scales should be incorporated.

C. R. Philbrick (US)

You have described the use of cloud persistance over periods of 1 to 4 hours. What is possible for longer time scales, such as the diurnal persistance on the next day?

Author's reply:

The smallest features seen in the visible imagery (dimensions in the order of kilometers across) evolve so rapidly that they cannot be followed overnight. During the night, infra-red imagery must be used, and because of its lower resolution and small thermal contrast between the low cloud tops and the ocean surface, such features are not resolvable in the first place. Nevertheless, sufficiently large features, say with horizontal dimensions 100's of km in extent, are often persistant enough to be identified from one day to the next, whether in visible or infra-red wavelengths.

ABSTRACT

of

Paper 31

"Electromagnetic Properties of Sea Ice"

Kenneth C. Jezek
Byrd Polar Research Center
The Ohio State University
Columbus OH 43210

This paper reviews research completed by a team of investigators as part of an Office of Naval Research Sponsored Accelerated Research Initiative on the Electromagnetic Properties of Sea Ice. The objectives of the program are to: understand the mechanisms and processes that link the morphological physical and the electromagnetic properties of sea ice: develop and verify predictive models for the interaction of visible, infrared and microwave radiation with sea ice: develop and verify inverse scattering techniques applicable to problems involving the interaction of EM radiation with sea ice. Guiding principles for the activity have been first that all EM data are supplemented by careful physical property data (salinity, density, roughness for example) and second that broad spectral data be acquired in as near simulaneous fashion as possible.

Research has been conducted by over 30 investigators who have participated in laboratory, field and modeling studies. Measurements on thin ice types (<30 cm) have been performed primarily under controlled laboratory conditions. Key results have been improved understanding of the temporal evolution of microwave backscatter as the ice ages (and the suggestion that such signatures can be used as proxy indicators of ice thickness), observations of the effect of thin layers of fresh water on the infrared signature of open sea water: measurements of optical albedo of new ice types: development of forward models that predict the contribution of microwave scattering from the volume and from the surface: and the initial development of inverse models for computing depth dependent dielectric properties.

Thick, first year ice types (50 cm - 2 m) have been studied at field locations in the Arctic. In situ observations have led to new ideas about the role of sea ice density, particle concentration, and layer effects on optical attenuation and absorption. The effect of snow cover on microwave backscatter has been documented and models to estimate snow thickness from SAR data have been proposed. The relationships between optical and microwave signatures and geophysical effects such as insolation and heat flux have been documented.

Along with describing results from experiments and modeling efforts, we will briefly discuss possible paradigms for using broad spectral data in developing algorithms for analyzing remote sensing data in terms of ice concentration, age, type and possibly thickness.

Paper presented at the AGARD SPP Symposium on "Remote Sensing: A Valuable Source of Information", held in Toulouse, France from 22-25 April 1996, and published in CP-582.

Paper 31

M. McKean (CA)

What impact does polarization have on the information available to you?

Author's reply:

In passive micro-wave sensors the polarization can be quite important. For active microwave information there does not appear to be a great deal of difference - although several researchers have argued that cross-polarization can provide additional information.

T.Wahl (NO)

There are many satellites up there now giving information about sea ice. If you were allowed to launch one more instrument in space for ice studies, what should that be?

Author's reply:

I would recommend the launch of a high resolution (resolution and coverage commensurate with current SAR and optical imaging systems), thermal infra-red system. The combination of surface temperature data with satellite-derived ice extent, concentration and, possibly, thickness data would be a major contribution to polar ocean studies.

A Global Positioning System (GPS) Tide Gauge

Kenneth D. Anderson
Propagation Division
NCCOSC RDTE DIV 883
49170 PROPAGATION PATH
SAN DIEGO CA 92152-7385
USA

SUMMARY

A non radiating, non intrusive, remote sensing method to measure changes in water height is examined. This *tide gauge* can be easily set up and be operating in a short amount of time -- the amount of time is controlled by the time it takes to set up an antenna and acquire Global Positioning System (GPS) satellite signals. The equipment package can be small, about the size of two briefcases, and consists of a handheld GPS receiver, a small patch antenna, and a notebook computer. There is no requirement to install equipment in the water to measure the changes in water height (even though modern tide gauges use non intrusive height sensors, such as ultrasonic acoustic transducers that send a pulse toward the surface and measure the time delay between the transmission and reception of the pulse, accurate measurements always require the use of a stilling well and that must intrude into the water surface.) While not amenable for all locations the size, ease of installation, and remote sensing aspects significantly reduces the logistics cost of a tide gauge. The US National Geodetic Survey (NGS) is interested in evaluating this tide gauge. Possible use is contemplated as a storm surge sensor during hurricanes and as a replacement for gauges installed in difficult areas. In addition, it is likely that US Navy could use the means during amphibious operations, particularly in areas where the tidal effects are not well known.

1. INTRODUCTION

The height of an earth based receiving antenna above a water surface is usually determined using both geodesy to establish the receiving antenna location (latitude, longitude, and height above the geoid) and either direct measurements or calculations of the height of the water surface in relation to the geoid. For land based surveying work the height at a specified location is referenced to an accurate vertical datum (in North America the vertical datum is the North American Vertical Datum of 1988). Tide gauges, which directly measure the height of a water surface through various means, are also referenced to the same vertical datum. So by simple addition or subtraction one is readily able to determine the height above the water surface at any location. The obvious problem here is that there must be a tide gauge installed on the body of water and both the tide gauge and the location of interest must be referenced to a common vertical datum. In addition, land masses are generally moving, albeit slowly, in the horizontal and vertical planes, so there is a need to continually resurvey the desired land location and the location of the tide gauge to maintain an accurate relation to the datum. If the tide gauge and the location of interest are nearby, the relative height offset between them could be established by standard survey techniques. This height offset could be used instead of relating measurements to a vertical datum but this would not eliminate the need for continually resurveying to maintain an accurate relative offset.

To estimate the height above water at a site where there is no tide gauge, one can use tide gauge measurements from several sites in the vicinity and mathematically estimate the tide height at the site of interest. For example, although there is no tide gauge on the ocean side of Point Loma, where the Naval Command, Control and Ocean Surveillance Center RDT&E Division (NRaD) is located, there are tide gauges at Scripps Pier in La Jolla and at Broadway Pier in San Diego Harbor. Applying hydrology and geodesy techniques to many years of data from these two sites, one can create a model of the Point Loma tides. However, the model is secular and therefore unable to adequately describe short term effects such as storm surges.

There are places where it is very difficult to position a tide gauge, the ocean side of Point Loma is one such example. Along the coast, because of the destructive and forceful nature of the waves, surf action generally precludes the installation of a tide gauge unless it can be installed out near or beyond the surf zone. On island sites, logistics may make it prohibitively expensive to install and maintain a tide gauge. A remote sensor would reduce the difficulties of measuring tides at isolated sites or along the coast.

Recent work on the Office of Naval Research sponsored GPS Sounder project [*Anderson,* 1994] strongly suggests that measurements of GPS signals as the satellite is rising or setting on horizon can be used to directly determine the height of the antenna above the water surface. Figure 1 is an illustration of the concept. Provided that there is only a water surface between the

Paper presented at the AGARD SPP Symposium on "Remote Sensing: A Valuable Source of Information", held in Toulouse, France from 22-25 April 1996, and published in CP-582.

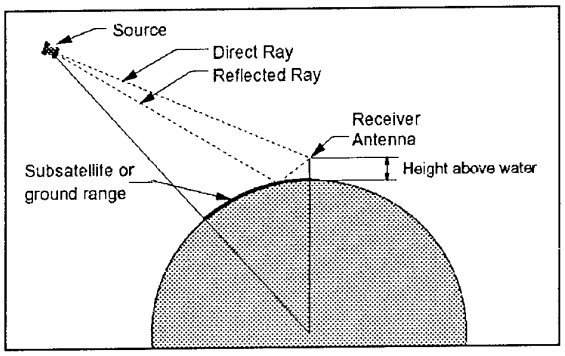


Figure 1. An overview of the GPS Tide Gauge concept.

receiver and the horizon where the satellite rises or sets, there will be two paths for the signal from the satellite; one path is the direct path, or direct ray, and the other is the reflected path, or the reflected ray. As the satellite moves in its orbit, the path length difference between the direct ray and the reflected ray will change many wavelengths so, at the receiver, the signal will appear as an interference pattern when plotted as a function of time or subsatellite range, which is the range from the receiver to the nadir point of the satellite on the earth. For very low-elevation angles (less than a degree or so) refractive effects, such as ducting, strongly influence the spacing between peaks (or nulls) in the interference pattern; this is the basis of the GPS Sounder project, which infers the vertical tropospheric refractivity profile from these low-elevation angle measurements. However, at higher elevation angles (above a degree or so), refractive effects are minimal and the spacing between peaks in the interference pattern is almost entirely due to the height of the antenna above the reflecting surface. Therefore, by both careful measurements and examination of the spacing between interference peaks one can deduce the height of the antenna above the surface, which implies a *remote sensing tide gauge*.

The following sections will review the fundamentals of ray optics and develop a model of the received signal interference pattern. Tide heights derived from GPS measurements will be compared to directly measured tide heights, and the errors, primarily from pattern recognition problems, will be examined. Results to date indicate that the technique of determining water level from GPS measurements has promise and, with some additional effort, is a workable solution as a remote sensing tide gauge.

2. RAY OPTICS AND MODELING

Snell's Law in a spherically symmetric medium can be expressed as

$$\frac{\rho n}{a}\cos(\alpha) = C \qquad (1)$$

where ρ is the distance from the earth center to a height Z above the earth surface, a is the earth's radius (6378.135 km), n is the refractive index at height Z, α is the local elevation angle of the ray, and C is a constant. From geometry

$$\frac{d\rho}{d\theta} = \rho\frac{\pm\sqrt{\xi^2 - C^2}}{C} \qquad (2)$$

where $d\theta$ is the earth interior incremental arc corresponding to the progression of the ray from ρ to $\rho+d\rho$ and ξ is shorthand notation for $\rho n/a$. Rearranging terms and integrating gives

$$\theta = C\int\frac{d\rho}{\rho\sqrt{\xi^2 - C^2}} \qquad (3)$$

The ground or subsatellite range is simply $a\theta$. Along the same lines of development, the optical path length can be written as

$$S = \int\xi\frac{d\rho}{\sqrt{\xi^2 - C^2}} \qquad (4)$$

If n is expressed as a linear function of $1/\rho$ then both Eqs 3 and 4 can be solved analytically (*cf Gradshteyn*, [1980]).

The refractivity model consists of three regions: A linear region that extends from the surface to a altitude of 4.85 km; an exponential region that extends from 4.85 km to 50 km, and, finally, a constant region that extends from 50 km to the satellite altitude (approximately 20,200 km) where the refractivity is assumed to be zero. For the tide gauge analysis, the linear region (surface to 4.85 km) is treated as a single layer such that the gradient of refractivity dN/dZ, where $N = (n - 1)10^6$, varies from -27 to -57 *N*/km. *Bean and Dutton* [1968] found that the refractivity N in the region from 9 to 50 km above the earth's surface behaves as

$$N = 105\exp\left(-0.1424(Z-9)\right) \qquad (5)$$

for Z expressed in km. If Eq. 5 is extended downwards to Z = 4.85 km, the gradient dN/dZ approaches the gradient expected in an adiabatic atmosphere, about -27 *N*/km. Experience (with refractive measurements at NRaD in Southern California) has shown that Bean & Dutton's expression is a reasonably good fit for heights greater than 4.85 km. The exponential region refractive profile is linearized into some 60 segments to facilitate the ray optics calculations.

Above 50 km, the atmosphere is treated as a vacuum. From geometry the interior earth angle θ_v for a ray

traveling from Z=50 km to the satellite altitude is given as

$$\theta_v = \frac{\pi}{2} - \alpha_v - \sin^{-1}\left(\frac{\rho_v}{\rho_s}\sin\left(\alpha_v + \frac{\pi}{2}\right)\right) \quad (6)$$

where α_v is the local elevation angle at Z=50 km, ρ_v is the vector from the earth center to Z=50, and ρ_s is the vector from the earth center to the satellite.

Calculations for the direct ray are straightforward applications of Eqs 1, 3, 4, and 6 for a specified ray elevation angle at the receiver. Calculations for the reflected ray are somewhat more difficult as the procedures involve an iteration scheme to find the reflected ray elevation angle at the receiver such that θ is identical (within a tolerance of 1 part in 10^{13}) on the two paths. The iteration normally uses a Newton technique, however, at the lowest elevation angles, it has been found very effective to use a binary-tree search. The total phase lag between the direct ray and the reflected ray includes the phase change from surface reflection (GPS signals are circularly polarized and the Fresnel reflection coefficient is computed using the circular polarized expressions described by *Hattan* [1990]).

An example of an interference pattern calculated by ray optic techniques for transmission from a GPS satellite to an earth based receiver 21.2 m above a water surface is shown in Figure 2. Instead of signal intensity the signal is plotted as signal-to-noise ratio (snr), which is a typical output from a high quality GPS receiver. This example uses the GPS L1 signal (1,575.42 MHz) and assumes that the refractivity gradient in the lowest portion of the atmosphere (surface to 4.85 km) is -27 *N*/km. The abscissa is subsatellite range, which is defined as the range along the earth's surface from the receiver to the nadir of the satellite, and corresponds to geometric elevation angles from about 10 degrees at 7400 km to about 1 degree at 8400 km. The interference pattern is clearly evident and the location of snr maxima or peaks are indicated by the vertical hashmarks. An arbitrary integer number can be assigned to the location of a peak at a selected range. Peak numbers are assumed to increase for increasing range and decrease for decreasing range. For example, in Figure 2, the peak at 7400 km is assigned the number 0. The next peak to the right is assigned the number 1, and so on such that the last peak, at about 8400 km, is assigned the number 35.

Figure 3 shows the relationship between the L1 peak number and subsatellite range calculated for different antenna heights (from 10 to 30 m in steps of 2 m) where the refractive gradient in the region from the surface to 4.85 km is -27 *N*/km. For geometric elevation angles from about 10 to 3 degrees (about 7400 to 8100 km),

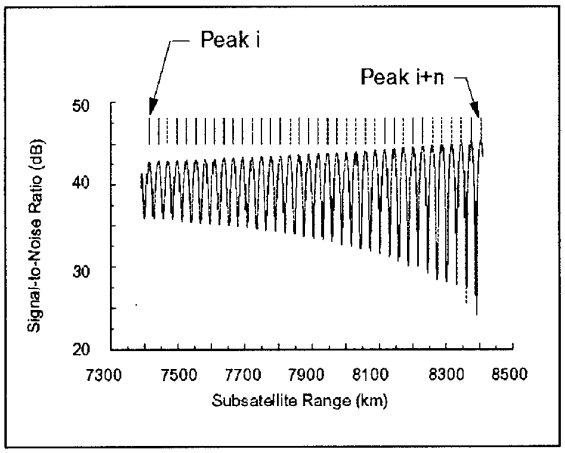


Figure 2. Calculated interference pattern and peak locations for the GPS Clear/Acquisition loop (1,575.42 MHz).

atmospheric refraction has a small effect on the position of the interference peak but it can be removed (or at least minimized) by considering the gradient of the peak number with respect to subsatellite range. If the gradient of the peak number with respect to subsatellite range is defined as *dP/dR*, a least squares regression analysis (and visual inspection of Figure 3) show that *dP/dR* is nearly linear so one can express the receiver height, z (in m), as a linear function of *dP/dR* (in km^{-1}). The regression analysis yields the relation

$$z = 594.518\ dP/dR + 0.01 \quad (7)$$

The standard deviation of the error is 0.0166 and the coefficient of determination is 0.999992, which validates the claim of linearity. A similar regression analysis for the L2 frequency (1,227.60 MHz) shows that the *dP/dR* multiplier is 762.964 which is in the proper ratio of 154/120 to the L1 multiplier.

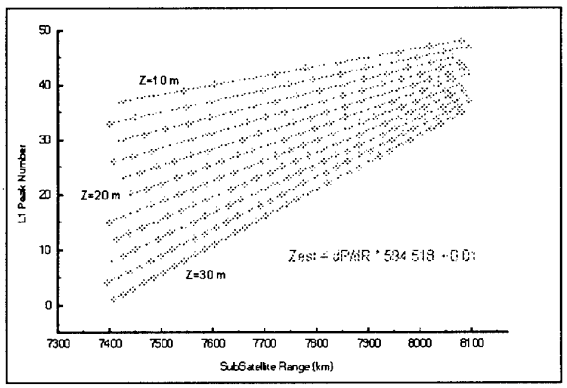


Figure 3. Computed interference peak location for antennas 10 to 30 m above the surface.

3. MEASUREMENTS

Figure 4 is an example of an interference pattern that was measured on 29 February 1996 using GPS satellite 27 (satellite vehicle number 27 broadcasting pseudo-random noise, PRN, code 27). The small crosses indicate the actual snr of the receiver's Clear Acquisition (CA) control loop, which was sampled at a 1 Hz rate; the dark solid line is a smoothed and filtered version of the CA snr. Procedures have been developed to automatically locate the interference peaks; the results are identified by the vertical hashmarks. It is obvious that the peak finding procedures have some shortfalls. If the vertical hashmarks (peak locations) are labeled from left to right starting at 0, marks 4 & 5 and 6 & 7 correspond to the shoulders of individual peaks, a peak is missed between 7 & 8 and another peak is missed between 14 & 15. There are two peaks erroneously declared and two peaks were missed. A regression fit to the data as it is with four errors yields an antenna height of 20.14 m, which is close to the measured height (based on the Scripps Pier tide gauge) of 21.25 m. An analysis of pure raw data with errors added in by the peak fit procedure gets to within about 5% of the true antenna height, which indicates that the technique has promise. Visually extracting the peak locations from the data results in marks 4 & 5 averaged to obtain peak 4 and marks 6 & 7 averaged to obtain peak 5. A regression fit is applied to peaks 0 to 5 and 7 to 13 (the missing peak, 6, is not used) and the calculated antenna height is 21.27 m, within 2 cm of the actual measured height. Clearly, as long as the *proper* peaks are used, the technique is remarkably accurate.

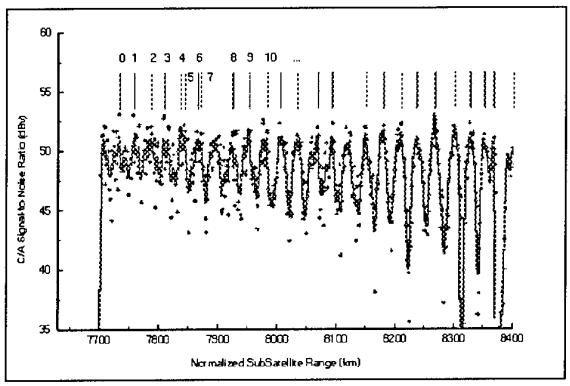


Figure 4. Peak locations automatically extracted for GPS satellite measurements made on 29 February 1996. Satellite 27 rising at 1633Z.

Results from a series of measurements made on 29 February 1996 are shown in Figure 5. The abscissa is time (major tics are spaced in 2 hour increments) and the ordinate is antenna height above the sea. The lighter curve with the vertical bars is the antenna height as measured by the tide gauge located at Scripps Pier (the vertical bars indicate the standard deviation of the measurements) and are reported every six minutes. The dark crosses are the results calculated from measurements of different GPS satellites rising and setting at the horizon. The number associated with each cross corresponds to the satellite PRN code (satellite 16 is shown twice, one measurement was at its rise and the other was made during its set). These results are all derived from a "hand" analysis of the data (along the lines of what was described in the previous paragraph) and are in good agreement with the Scripps Pier data. The error observed for PRN 09 was reduced to less than 10 cm by a computer search of the raw data to find a "best" fit (Scripps Pier measurement of 22.13 ∓ 0.08 compared to a calculated value of 22.03). However, the same application significantly increased the error for other satellites.

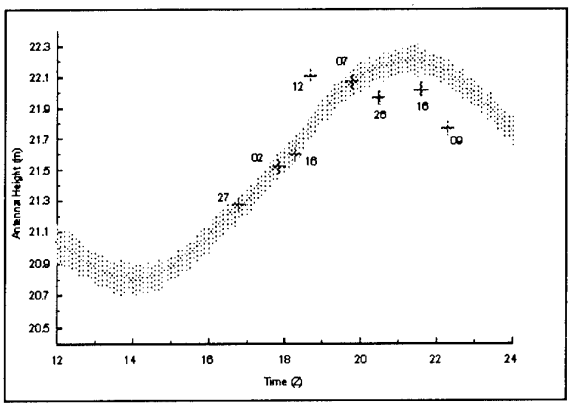


Figure 5. Calculated (cross) and measured antenna height (dots and bars) for a series of GPS satellites rising and setting on 29 February 1996.

Results from an earlier series of measurements [*Anderson*, 1995] are shown in Figure 6. Here, one satellite (PRN 12) was observed once per day for four days (5 through 8 December 1994) and the data were hand analyzed (these data are used to "train" both the automated peak location and the best fit procedures). The abscissa is the measured antenna height (based on the Scripps Pier tide gauge) and the ordinate is the calculated antenna height. Results from both frequencies (L1 and L2) are shown; the worst case error is less than 10 cm, which is about one standard deviation of the tide gauge measurements. Certainly, the technique has merit.

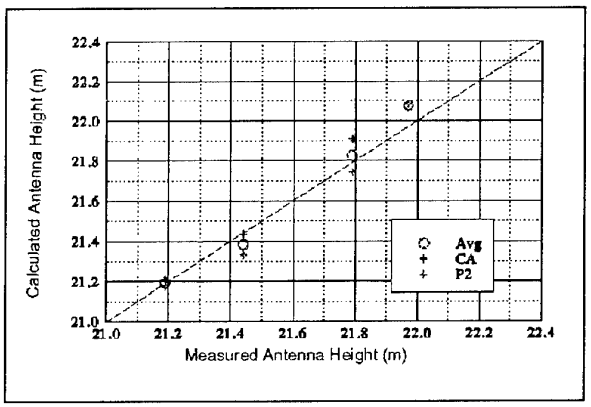


Figure 6. Calculated and measured antenna height for a series of data taken in December 1994.

4. CONCLUSIONS

A non radiating, non intrusive, remote sensing method to measure changes in water height is examined and the results to date indicate that the technique of determining water level from GPS measurements has promise and, with some additional effort, is a workable solution as a *remote sensing tide gauge*, which has a number of advantages. First, it can be easily set up and it can be operating in a short period of time. Second, the equipment package can be small, about the size of two briefcases, and it is easily transportable. Third, there is no requirement to install equipment in the water to measure the changes in water height. While not amenable for all locations because of the limitation that there must be a clear horizon between the receiver and the satellite, the size, ease of installation, and remote sensing aspects can significantly reduce the logistics cost of a tide gauge.

The reduction of GPS satellite signal measurements to antenna height above the water currently requires considerable human involvement (but not necessarily a high degree of skill) to sort out erroneous peaks and interpret the interference pattern. With additional effort it is strongly felt that the pattern recognition problems can be overcome and fully automated. It has been demonstrated that as long as the *proper* peaks in the interference pattern are used, the technique is remarkably accurate.

ACKNOWLEDGEMENT

This work is sponsored by Dr. Scott Sandgathe of the Office of Naval Research.

REFERENCES

Anderson, K.D., Tropospheric refractivity profiles inferred from low-elevation angle measurements of Global Positioning System (GPS) Signals, AGARD CP 576-2, Sep. 19-22 1994.

Anderson, K.D., A GPS Tide Gauge, *GPSWorld Showcase*, p. 44, August 1995.

Bean, B.R. and E.J. Dutton, *Radio Meteorology*, Dover Publications, Inc. 1968.

Gradshteyn, I.S., and I.M. Ryzhik, *Table of integrals, series, and products*, p. 80, Academic Press, 1980.

Hattan, C.P., Specification for a standard electromagnetic propagation model, NCCOSC RDTE DIV Tech. Doc. 1772, June 1990.

Paper 32

N. Farsaris (GR)

Can the TG work with the cheap versions of GPS receivers? (Unencrypted signals only?)

Author's reply:

Yes. The tide guage concept does not require the use of the encrypted GPS signals. Good results can be obtained by using just the C/A signal, which is not encrypted. The advantage of using the encrypted signals is that two measurements, using both GPS frequencies, can be made at the same time.

N. Douchin (FR)

You said that propagation effects (refraction) can be neglected for the elevation range considered. What about an elevated duct on the satellite-to-ground path?

Author's reply:

An elevated duct will slightly change the location of a peak in range, but the gradient of peak number with respect to range is not affected as much. There is some small effect but it can be ignored.

G. Brown (US)

Do you see any effects of the diffuse scattering from the surface roughness and, if so, can you comment on its magnitude.

Author's reply:

I am not sure how much system noise comes from ionospheric fluctuations, tropospheric fluctuations, or scattering from the surface. However, we will look at surface roughness effects to see if we can use the peak to null depth as an indication of wind speed.

C. Zeisse (US)

How does the accuracy of your measurement depend on receiver height?

Author's reply:

I have looked at receiver heights from 5 to 30 m above the water and the procedure is very workable. I think for heights less than 5 m the accuracies will begin to suffer.

Assessment of NAM, the navy aerosol model, for visibility determinations in inland seas such as the Persian Gulf

Stuart G. Gathman
Propagation Division
NCCOSC RDTE DIV 883
49170 Propagation Path
San Diego CA 92152-7385
USA

SUMMARY

The effects of aerosol on the transmission of electrooptical energy in the marine atmosphere are very important. The propagation losses caused by aerosols at shipboard level over the open ocean regions of the world are adequately described by the Navy Aerosol Model, NAM *Gathman* [1983]. However, inland seas are very different in their geographical situations and the blind application of open ocean developed models to these regions is not necessarily applicable. The US Navy SHAREM 110 exercise, which took place in the Arabian Gulf and the Gulf of Oman in Feburary 1995, offered an excellent opportunity to investigate aerosol characteristics in these inland sea regions and compare the results to the open ocean situations as represented by NAM.

1. INTRODUCTION

NAM represents the atmospheric aerosol size distribution by a tri-modal aerosol size distribution consisting of three lognormal components each parameterized with measurable meteorological variables. The open ocean NAM assumes that a three lognormal structure is adequate to describe the aerosol size distribution. In NAM, where the aerosol radius is "a", the aerosol size distribution, n(a) is described as:

$$\frac{dN}{dr} = n(a) = \sum_{i=1}^{3} \frac{k}{f} \cdot A_i \cdot \exp\left(-\ln\left(\frac{a}{f \cdot r_i} \right)^2 \right)$$

The relative humidity effects on the sizes and composition of hygroscopic aerosol in NAM are represented by the swelling factor, f, which is 1 at a relative humidity of 80%. The visibility correction factor, k, is 1 and the mode radii, r_1, r_2, and r_3 of the three species of aerosol are: 0.03, 0.24, 2.0 micrometers, respectively.

The second and third mode of the model respond to wind generated aerosol over the open ocean whereas the first mode refers to a background aerosol class probably of non-marine origin. Several methods of determining the magnitude of the first mode have been suggested which work quite well for open ocean work but fall short of ideal for coastal or inland sea situations. For instance the Radon 222 concentration technique makes the assumption that Radon 222 is a tracer for air that has come from the land and used to determine how long an air mass has been over water prior to coming to the measurement site. In an inland sea situation, however, all the air that is measured at the site has spent time over land only a few hundred kilometers prior to reaching the measurement site.

Recently a method utilizing the atmospheric condensation nuclei CN measurement for determining the magnitude of the first mode has been suggested by *Battalino* [1995]. CN is in principle a simple measurement dating back to *Aitken* [1880/81] and *Pollak and O'Connor,* [1955] which utilizes the principles of sudden expansion similar to that of a cloud chamber to grow all particulate in the sample to be visibly active. The US Navy SHAREM exercise in Feburary 1995 offered an opportunity to test the application of the CN method for determining the applicability of NAM for inland sea locations.

Paper presented at the AGARD SPP Symposium on "Remote Sensing: A Valuable Source of Information", held in Toulouse, France from 22-25 April 1996, and published in CP-582.

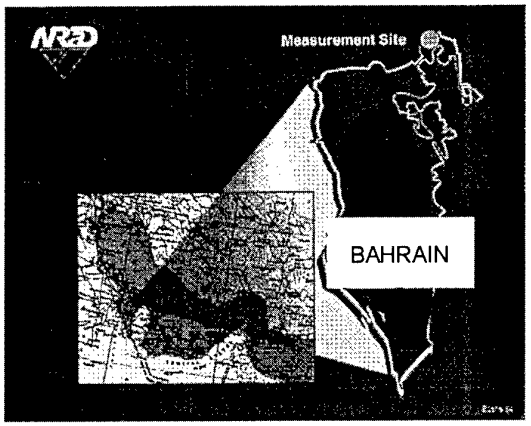


Figure 1 A Map showing the measurement site in relation tho the island of Bahrain and the Gulf region.

2. THE MEASUREMENTS

The location of the measurement site in Bahrain and its relation to the rest of the region is shown in figure 1. The aerosol size distribution measurements were made with an axially scattering spectrometer probe (ASSP-100) manufactured by Particle Measuring Systems, Boulder, CO. The total concentration of aerosol at the site was measured with a Rich-200 Condensation Nuclei Monitor produced by Environment One, Syracuse, NY. A single angle nephelometer scattering device on loan from Fysisch en Elektronisch Laboratorium, FEL-TNO, the Netherlands, was used to measure visibility. Data from these instruments, together with weather data collected by the Bahrain airport, constituted the island measurements portion of the SHAREM 110 aerosol experiment. The ASSP-100 aerosol spectrometer probe was mounted on the edge of the roof of a two-story office building at the airport in Bahrain. The nephelometer and the end of the CN sampling tube were located within a meter of the aerosol probe as shown in figure 2. The electronic processing instruments for the aerosol probe and the nephelometer and the CN monitor itself were located in a second-story office directly below the roof line. The instruments were set up to run continuously with a minimum amount of operator maintenance required.

The airport in figure 1 is located on a small island, Al Muharraq, connected to the main island of Bahrain by a causeway. The office building used is located within 500 meters of the water on the east side of the island, facing away from the land masses of Saudi Arabia and Bahrain. Any time land-based instruments are trying to measure conditions in nearby water, local contamination is a major problem. However, proper use of wind direction and filtering data for obvious locally-produced aerosol clouds from trucks and aircraft can assure that the data reasonably represents the aerosol in the Persian Gulf.

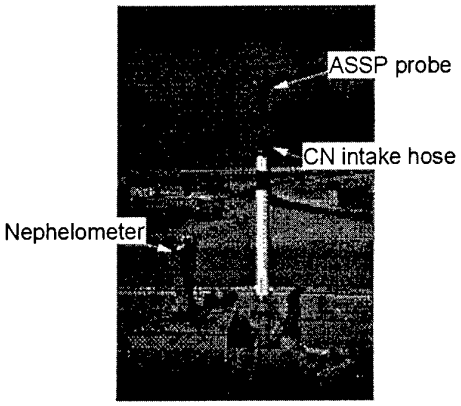


Figure 2 A photograph showing the location of the various measurement sensors on the roof of the building in Bahrain.

The purpose of this paper is to use the meteorological measurements from the SHAREM experiment to calculate the volume scattering coefficient at visible wavelengths using NAM (with the CN technique for determining A1) plus a molecular scattering component. This will in turn be compared with a direct scattering measurement made using the nephelometer and with volume scattering calculated with Mie theory using the aerosol data from the directly measured aerosol size distribution (for the larger sizes) and model estimates of dN/dr from the CN measurements for sizes below the threshold of the aerosol spectrometer. These results will show the applicability of using a model like NAM in an inland sea region if the appropriate filtering of the CN data is taken into account.

3 MODEL AND MEASUREMENTS COMPARISON IN THE VISIBLE

A comparison between the measurements and modeling of scattering coefficients at the visible wavelengths make up the data set by which this paper makes its conclusions. Two completely different instrumental measurements were made by the nephelometer on the one hand and the combination of the CN counts and the aerosol spectrometer data on the other. Of course they are not identical measurements and cannot be compared directly as this would be similar to comparing apples and oranges. Each of these measurements must be converted into the same physical property so that a direct comparison can be made.

The nephelometer measures the scatter of light directly whereas the aerosol spectrometer and the CN counter provide information about aerosol size distribution. The calculated extinction from an aerosol size distribution does not account for the extinction that is produced from the molecules in the air whereas the "extinction" or "scattering" obtained from the nephelometer cannot tell if the molecules or the aerosol are responsible for the scattering.

Normally in the visible wavelengths, the value of the scattering coefficient is due to the more or less constant mix of molecules in air, and is rather small in value compared to that of the aerosol and can be calculated from the LOWTRAN code when no aerosol is specified. This must be added to the aerosol scattering calculated from the aerosol size distribution to provide an equivalent quantity by which to compare the two results of the two instruments.

NAM uses the CN concentration and the standard airport meteorological measurements to calculate the model aerosol size distributions on which Mie calculations are made to obtain a model determined aerosol scattering coefficient. To this also the molecular scattering must be added.

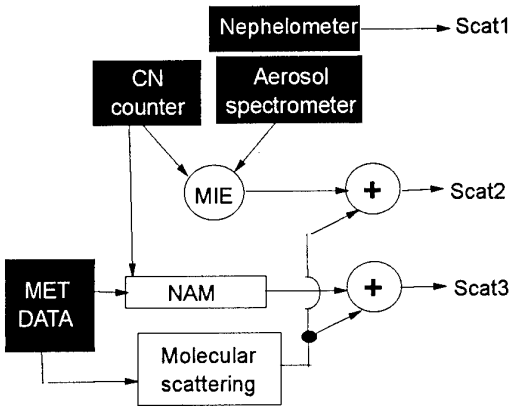


Figure 3 Flow chart showing how measured data from the black boxes are mixed with various calculations and models to all provide the scattering coeffients for comparison.

Figure 3 shows the data flow from the instruments available for this experiment and how the data was manipulated to obtain the same parameter at the measurement site. There is of necessity a common use of some of the data between various calculations. In the analysis section below, it is believed that there are sufficient tests of the data to show that this does not cause an interpretation problem with the data.

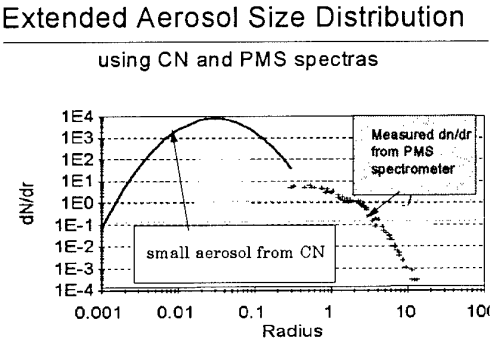


Figure 4 The extended aerosol size distribution is plotted here using both the CN count to define a lognormal curve for the smallest aerosol sizes (below 0.3 microns) and the directly obtained aerosol size distribution measured at sizes above 0.3 microns.

Figure 4 shows the buildup of a complete aerosol size distribution using the ASSP and the CN counter. In this figure, the dN/dr values obtained

from the aerosol spectrometer are plotted as individual points. Also in this figure is plotted a lognormal function which is fitted so as to represent all of the aerosol measured by the CN counter (greater than .003 microns) less the number of aerosol measured by the aerosol spectrometer. The log normal has the same width parameter and mode radius as does the first term of the NAM model. This is a method of extending the aerosol size data obtained from the aerosol spectrometer to smaller sizes which might be important in the short wavelength Mie calculations. The aerosol extinction at visible wavelength was then calculated over the entire range of data from .003 microns to the largest sizes which the ASSP was capable of seeing.

4. THE COMPARISON

In Figure 5, a time series of all three of these measures of extinction in the visible wavelengths is plotted.

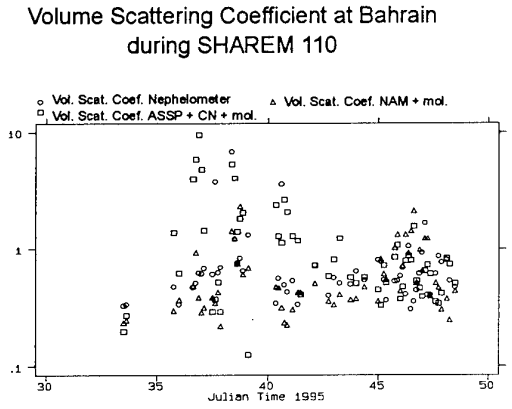


Figure 5 Plot of all of the scattering coefficients obtained from the measurements during the experiment plotted on a semi log scale. Note that the model is the triangle symbol in the plot and the more direct measurements are the circles and the squares.

It is seen that all of the data appears to fall into the same envelope of scatter. There does not appear to be a clear cut divergence of the data as would be expected if one of the measurements were clearly better or worse than the others. An occasional spike is seen which can be associated with local contamination times which were not isolated by the filtering techniques used. A look at the comparison in a more statistical sense is shown in figure 6 which shows the results of the comparison between the scattering coefficient from the nephelometer, Mie calculations from the aerosol measurements, and the NAM model.

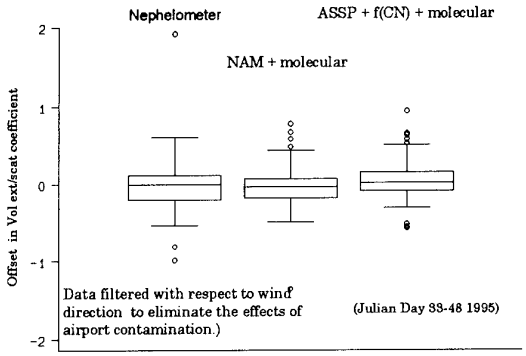


Figure 6 Box and whisker plot showing the model performance with respect to the measurements. The offset here is the difference between an individual observation of the volume scattering coefficient and the mean of all three for that time spot.

Because it is not known at any particular time what the visible extinction is exactly and there exist three different indicators of unknown accuracy of what it is, an average of all three gives the best value for the atmospheric parameter at that time. This comparison is expressed in the form of a "box and whisker" plot which shows at a glance various statistical quantities of the experiment.

The plot has the following graphical interpretation. Each of the box and whisker figures in the plot represents the offset of a particular measurement from the mean of all three of the measurements at a particular time. There is a box near the center of each figure which represents the interquartile range of the offset data for that particular technique. The horizontal line near the center of the box represents the median of the offset data while the upper and lower edges of the box represent the 75th and the 25th percentiles respectively. Extending from the boxes in the vertical direction are lines, the

extreme extent of which represent the upper and lower adjacent values of the data. (The adjacent value is 3/2 of the interquartile range.) Also indicated in the plot are outside values plotted as circles representing individual points obtained by measurement which are outside of the normal statistical measurements. From these data it is seen that NAM fits on the whole values of visibility as well or better than some of the direct measurements.

5. CONCLUSION

A tri-modal log normal aerosol model such as NAM was successful in predicting the electro-optical propagation characteristics of the aerosol loading over an inland sea provided a proper indicator of the very small aerosol is available. This experiment shows that in the visible wavelengths, the model compared well with the more direct measurements. It is also reasonable to predict that it would do equally as well in the IR bands if there were IR measurements available during the experiment to compare with. Several techniques are available for determining the A1 input parameter necessary for the successful application of NAM, but for the inland sea situation, a direct measurement related to the very small aerosol loading of the atmosphere is required. Measurements such as visible range or CN concentration seem to do well in this regard.

ACKNOWLEDGMENT

This work was sponsored by Dr. Andreas K. Goroch of NRL.

REFERENCES

Aitken, J., On dusts, fogs and clouds, *Trans. R. Soc. Edinb.*, **30**, p337, 1880//81.

Battalino T., Personal communication, 1995.

Gathman, S.G., Optical properties of the marine aerosol as predicted by the Navy aerosol model, *Optical Engineering*, **22**, p 057, 1983.

Pollak, L.W. & O'Connor, T.C., A photo-electric condensation nucleus counter of high precision, *Geofis. pura appl.*, **32**, p139, 1955.

Sensing Zero-Range Surface Ship Infrared Signatures

Charles P. McGrath
Carl R. Zeisse
Propagation Division
NCCOSC RDTE DIV 883
49170 Propagation Path
San Diego, CA 92152-7385
USA

SUMMARY

Many imaging and non-imaging ship signature prediction codes exist, but few have been validated. In order to properly evaluate target signature and background prediction codes, an accurate database of radiometric images is required. Accurate imaging radiometers provide the capability of remote sensing of marine targets to build such a database. However, measured images must be corrected for target emissivity and atmospheric influences to determine the apparent zero-range radiance values for the target and background. Presented here is a method for acquiring radiometric images of marine targets and correcting the temperatures for atmospheric influence. The measurements are made in both the 2.0-5.6 micron and 8.0-12.0 micron bands using an airborne AGEMA-900 thermal imaging system. The image acquisition is supported with meteorological measurements that provide the input for the LOWTRAN 7 [1] and the ship signature prediction codes. A modified version of the LOWTRAN 7 code containing the AGEMA system spectral response is used for making the atmospheric corrections. Statistical area analysis is performed on the graphic images to produce integrated target and background radiance values. The integrated radiance differences are converted to thermal contrasts for evaluating thermal contrast prediction codes, such as the Georgia Technology Research Institute model, TCM-2 [2], which is used in the Electrooptical Tactical Decision Aid (EOTDA)[3]. A primary use of the database will be to validate the generic ship model which is a part of the EOTDA.

LIST OF SYMBOLS

W	Planck's spectral radiant emittance
ν	wavenumber in cm^{-1}
R	slant path distance to target in km
T_S	zero-range average ship temperature in °K
T_A	measured average ship temperature
ε	ship emissivity
τ	atmospheric transmittance
N_P	path radiance
F	relative spectral response of imaging system

1. BACKGROUND

Effective strike warfare operations require weapons systems that are capable of delivering ordinance within a three meter window of accuracy from safe aircraft stand-off ranges. Target acquisition and designation ranges from optical sensors depend upon the amplitude of the zero-range contrast between the target and the background, and upon the attenuation from the intervening atmosphere. Mission planners rely on tactical decision aids for predicting these effects. Accurate predictions minimize pilot risk and maximize mission success by determining the optimum time over target and by selecting the most effective type of weapons to employ for a given scenario. Infrared technology was employed extensively in the Persian Gulf during Operation Desert Storm. Since then, a major emphasis of the modern military has been improving models that predict the performance of infrared sensors. In order to properly evaluate these models, accurate empirical data are needed. The measurement effort described in this report is primarily focused on improving the EOTDA.

The EOTDA is a tri-service decision aid which is implemented in major mission planning systems.

Paper presented at the AGARD SPP Symposium on "Remote Sensing: A Valuable Source of Information", held in Toulouse, France from 22-25 April 1996, and published in CP-582.

Although originally developed by the US Air Force in the early 1980s, it was later extended to include Army and Navy sensors and targets and is employed in the Navy Tactical Environmental Support System version 3.0 (TESS(3)) and the Tactical Aircraft Mission Planning System (TAMPS) version 6.0. The EOTDA was used occasionally during the Persian Gulf War by the US Air Force and favorably influenced mission planning [4]. The EOTDA is also being marketed globally by Hughes STX Corporation, who was a prime contractor in developing the original EOTDA for the US Air Force in the 1980s.

As depicted in figure 1, the EOTDA consists of three primary modules, the sensor performance model, the atmospheric transmission model, and the

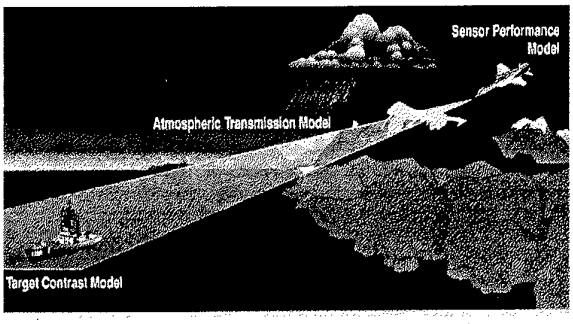


Figure 1. EOTDA scene composition: the sensor performance, atmospheric transmission, and target contrast models.

target contrast model. Ideally, the majority of uncertainty of EOTDA predictions should come from the sensor performance model, mostly because of human performance differences among operators. The sensor performance model uses the Johnson criteria [5] to convert thermal contrasts to an equivalent bar chart, based upon the target geometry, thermal properties, and human perceptions. If we assume input data are sufficiently accurate and also assume the scene complexity fits within the modeling constraints, then, attributing 20 percent accuracy to the transmission [1] model, 20 percent to the target contrast [2] model, and 25 percent to the sensor performance model [6] yields an expected accuracy of the overall EOTDA of 38 percent. However, the scatter diagrams of figures 2 and 3 show that in the marine environment, the error even exceeds the 50 percent error bars in both the coastal and open-ocean cases.

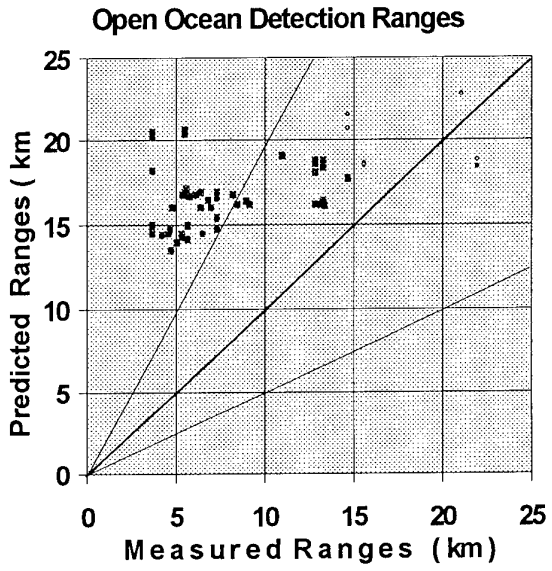


Figure 2. Open-Ocean Prediction Errors.

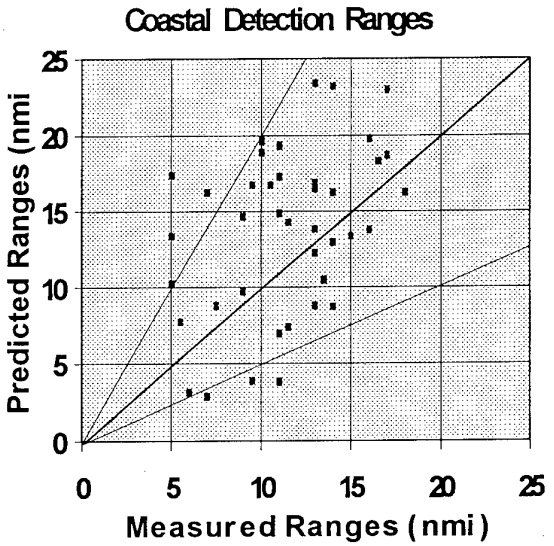


Figure 3. Coastal Prediction Errors.

Several reports [7, 8, 9] indicate the excessive error in applying the EOTDA to marine tactical scenarios is probably caused by the water background algorithm. Thus, a primary goal of the remote sensing measurement effort described in this paper is to test improvements to the water background model and the impact of improvements on the overall performance of the EOTDA. Currently, the Zeisse

SeaRad model [10] is being implemented into the EOTDA for comparison with the existing Shapiro model [11].

2. PLATFORM AND EQUIPMENT

To properly evaluate the EOTDA, calibrated infrared imagery, and meteorological and navigation data are required at the target site. To avoid the problems and expense associated with on-board ground-truth measurements, target and background properties are measured remotely using a Piper Navajo aircraft as the primary measurement platform.

2.1 Imaging System

The target imaging is accomplished with an airborne AGEMA model 900 Thermovision imaging system. The range specifications and accuracy are given in the table 1.

Waveband (microns)	Range (°C)	Accuracy (°C)	Type Det.
SW 2- 5.6 (no filter)	-10/+80°C	±1.0 °C	InSb
LW 8-12 (LPL filter)	0/+110 °C	±0.5 °C	MCT
LW 8-12 (no filter)	-30/+80 °C	±1.0°C	MCT

Table 1. AGEMA 900 Specifications.

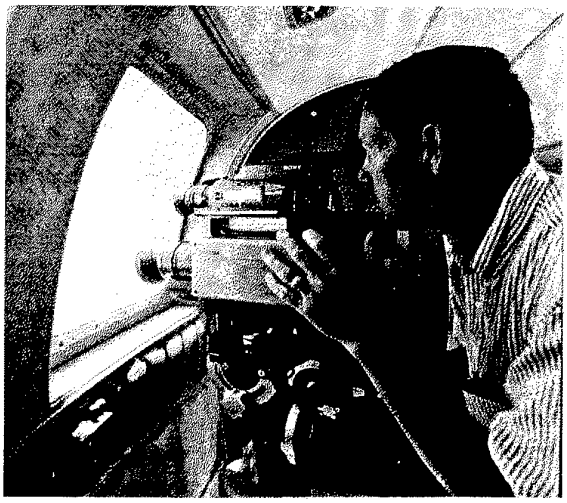


Figure 4. Aircraft use of IR and visible Cameras.

Figure 4 shows the mounting configuration of the three cameras (two AGEMA infrared cameras and one Sony video camcorder). The Sony Hi-8 camcorder is used for sighting and tracking the target during measurements and provides an image record in the visible wavelengths.

The AGEMA 10° x 5° (horizontal x vertical) field of view lenses are employed for all infrared measurements in the aircraft system. These provide a spacial resolution (IFOV) of 0.87 milliradians for the shortwave band (2-5.6 microns), and 0.76 milliradians for the longwave band (8-12 microns) at 50% modulation. A seven micron long-pass (LPL) filter is used for all longwave measurements to eliminate midwave influence onto the longwave detector. Since the instrument is Sterling cooled, the need for handling liquid nitrogen during the aircraft measurements is eliminated.

2.2 Meteorological Support

The NRaD-leased Piper Navajo twin engine aircraft is instrumented to collect and record meteorological and aerosol data. Most of the parameters required to run LOWTRAN, MODTRAN, and the EOTDA are measured aboard the aircraft. The measured parameters include air temperature, relative humidity, dew point, infrared sea surface temperature (8-12 micron band), air pressure, and aerosol size distributions.

Supplementary measurements from other platforms provide insight into weather variability and a history that cannot be obtained from the aircraft alone. Hourly surface observations are usually available from shipboard records. These also provide wind data not available from the aircraft.

2.3 Geometry Considerations

Navigation data are necessary to determine the distance and view angle between the aircraft sensor and the ship target. The position of the target is determined by manually noting the latitude and longitude on the airborne Trimble 3000 GPS during over-flight at the beginning and end of the measurement event. The GPS system is accurate to 15 meters and the position of the aircraft is recorded every five seconds. The indicated air speed, and the pressure and radar altitudes are also recorded. The Bonzer TRA-2500 radar altimeter is accurate to five percent and is used to calibrate the pressure

altimeter. The altitude and GPS positions allow the path geometry between the sensor and target to be calculated. Additionally, ship navigation and weather records are collected when the ship returns to port to compare with *in situ* measurements. From this information the target operational history is determined.

A second method of determining the view angle and distance from the sensor to the target is to use the image data itself along with the altimeter recordings. The scanner samples 272 pixels per line and displays 136 lines at a frame rate of 15 Hertz. Using the 10° x 5° lens, 0.646 milliradians are subtended per pixel. If the length of the target is known, then the distance to the target can be calculated from the number of pixels subtended.

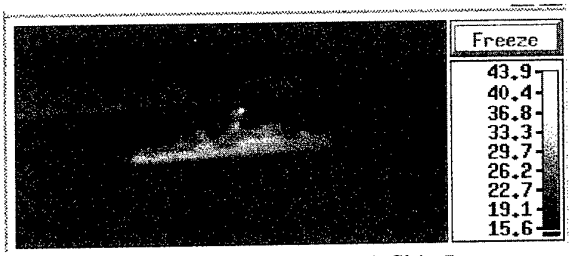
3. MEASUREMENTS

Three primary types of measurements are collected. First, meteorological input data needed for the LOWTRAN, MODTRAN, and EOTDA models are collected. This involves flying in a vertical upward spiral starting at an altitude of 15 meters, climbing to 1500 meters. The second type of measurement is close-up imaging of the target and background that can be later converted to zero-range measurements. Meteorological data are continuously recorded during the radiometric imaging. The third type of measurement is the detection range of the target, which is determined by flying away from the target until the target is barely discernible against the background and then noting the distance on the distance measuring equipment (DME) onboard the aircraft.

3.1 Close Range Measurements

In most cases the choice of flying radius around the target is that distance which allows the entire target to just fit on the image screen. However, to achieve confidence in the analysis and corrections, it is useful to measure at two or more distances along the same slant path. Typical flying radii when imaging ships as targets range from 500 to 3000 meters. At a distance of 800 meters from the target the target area per pixel is 0.257 x 0.257 meters per pixel. At 1200 meters it is 0.385 x 0.385 meters per pixel.

Figure 5 illustrates the typical scene resolution of both the long and short wave AGEMA cameras for a coastal scene at about 3 kilometers from the target.



Short Wave (2.0-5.6 microns) Ship Image



Long Wave (8-12 microns) Ship Image

Figure 5. Coastal IR scene.

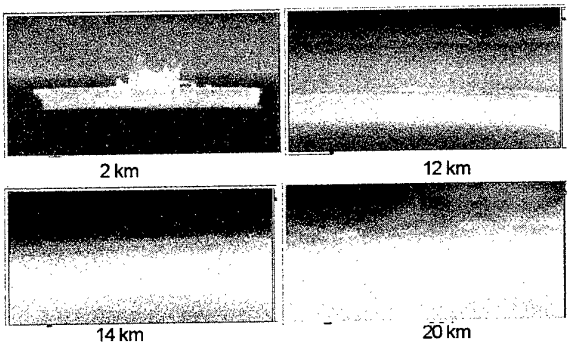


Figure 6. Determining detection range. From an aircraft altitude of 300 meters the ship was barely visible at 20 km and just disappeared at 23 km.

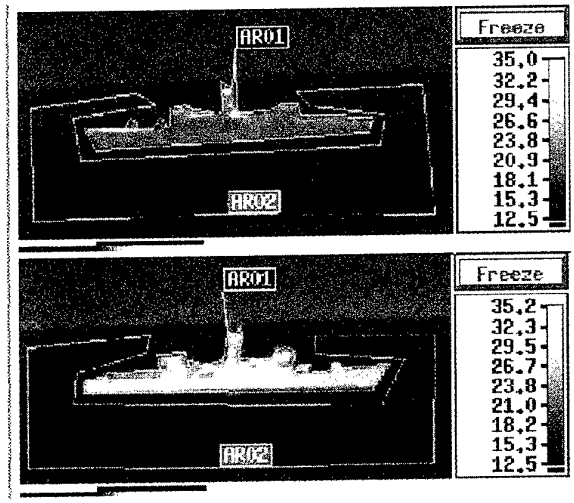


Figure 7. Target/background analysis. The port view (top) and starboard view (bottom) are both 8-12 micron images of same ship at same time.

Although, target resolution diminishes with distance, resolution is not a problem because the EOTDA analysis operates primarily on total average temperatures over the entire target area, rather than hot spot detection. The major problem resulting from this remote sensing technique is the error introduced when correcting for the influence of the intervening atmosphere.

3.2 Detection Range Measurements

Besides making close-range measurements to acquire the thermal signatures of the target and background, detection range measurements are also collected for comparison with predictions from the EOTDA. Figure 6 shows infrared images (8-12 micron band) of a ship target fading from view as the distance from the camera increases. The small bright spot representing the thermal signature of the ship is barely visible at the 20 kilometer range and just disappears at about 23 kilometers. The range is determined from the aircraft DME display which receives the tactical aircraft navigation beacon (TACAN) signal from the ship. This detection range will be compared to the EOTDA prediction once the generic ship model (an add-on to the EOTDA) is available to model this type of ship.

4. DATA INTERPRETATION TECHNIQUES

The AGEMA 900 system provides several tools for analyzing the recorded images. The total integrated radiance is of primary interest in evaluating the target and background models of the EOTDA. The total integrated radiance is found by using the area analysis capability of the AGEMA system. The system provides the capability of defining an irregular area and then performing statistical analysis on that area.

	Port View	Starboard
Ship	20.4 ° C	29.0 °C
Background	13.5 °C	13.2 °C
Difference	6.9 °C	15.8 °C

Table 2. Average temperature results from Figure 7 image area analysis statistics, showing temperature contrast is significantly greater for the starboard view.

Figure 7 shows long wave ship images of the port and starboard sides of a ship after the areas of interest have been outlined. In this example, there is a significant difference between the two sides. Table 2 shows that in this example, the thermal contrast is significantly greater (15.8 °C v. 6.9 °C) for the starboard view, making the ship much more vulnerable to an attack to the starboard. An attack aircraft could conceivably locate and deliver weapons at more than twice the stand-off range by approaching toward the sunlit (starboard) side instead of the shaded and wind cooled side. It is crucial that decision aid software be able to accurately predict these tactical conditions.

Because the distances from the camera to the target (500 to 3000 meters) include a significant atmospheric influence, the image analysis requires further correction. The atmospheric contribution to the scene displayed by the calibrated AGEMA 900 imaging system is shown in equation 1. Each side of the equation represents the total radiance upon the detector for a given pixel. The image is displayed as a 272 x 136 pixel array of spot temperatures determined in accordance with Plank's black body formula.

Equation 1 can also represent the total scene radiance arriving at the AGEMA detector for a given set of pixels. The left side shows the total radiance in terms of the apparent target temperature, T_A. The right side shows the break-down of the apparent temperature, T_A, into its atmospheric components and the actual (zero-range) target temperature, T_S, which is the desired result.

$$\frac{1}{\pi}\int_{\nu_1}^{\nu_2} W(T_A,\nu)F(\nu)d\nu = \int_{\nu_1}^{\nu_2}\left[\varepsilon(\nu)\tau(\nu,R)\left(\frac{1}{\pi}\right)W(T_S,\nu) + N_p(\nu,R)\right]F(\nu)d\nu \quad (1)$$

The target emissivity, ε, is estimated from empirical data and assumed constant over the 8-12 micron band (wavenumbers 1250 cm^{-1} to 830 cm^{-1}). The AGEMA system relative spectral response, $F(\nu)$, is a measured parameter, that is inserted into LOWTRAN or MODTRAN by modifying the code. Certain meteorological parameters, such as, air mass parameter and visibility are operator determined

rather than instrument measured. The slant-path range, R, transmittance, τ, and path radiance, N_P, and the zero-range target radiance, N_S, are determined directly from the output of the LOWTRAN or MODTRAN codes. The final zero-range temperature, T_S, is calculated from the zero-range radiance, N_S, using Plank's blackbody formula.

Of particular interest is the effect of the AGEMA relative spectral response, $F(\nu)$, on the voltage produced by the detector and displayed as a pixel with of temperature T_A. Both the LOWTRAN and MODTRAN codes assume a flat spectral response (i.e., $F(\nu)$ equal to one) across the band of interest. This approximation may work for most cases, but accuracy can be improved by inserting the measured instrument response into the LOWTRAN or MODTRAN codes. Figure 8 shows the measured relative spectral response for our particular AGEMA system in the long wave band. Assuming a flat response in the long wave band (rather than the measured response) would likely produce reasonable results since the out-of-band energy nearly offsets the error in the in-band (8-12 micron) range of Figure 8. However, in the short wave band, assuming a flat response across the 3-5 micron band would produce much more inaccurate results as illustrated by the difference in the areas under the two curves in Figure 9. Thus, we consider it worthwhile to modify the LOWTRAN or MODTRAN codes to include the relative spectral response of the remote sensing system.

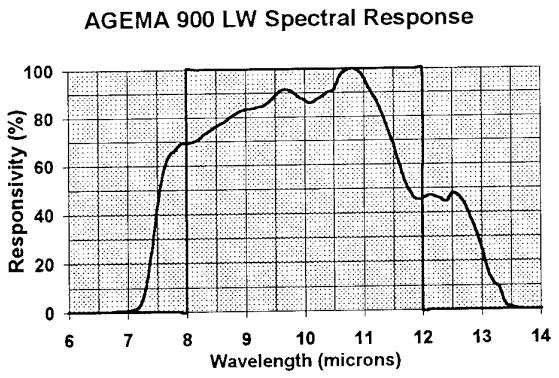


Figure 8. Relative spectral response of the AGEMA 900 long wave imaging system.

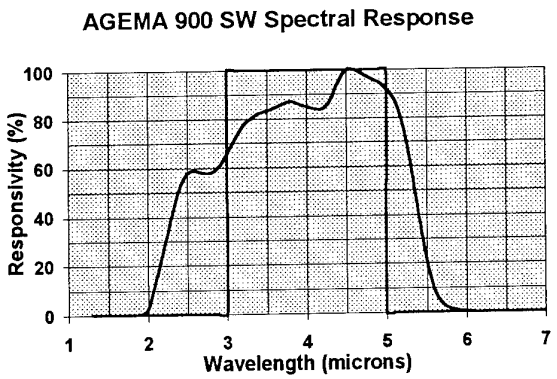


Figure 9. Relative spectral response of the AGEMA 900 short wave imaging system.

The general procedure for correcting the apparent spot temperatures or area averaged temperatures to zero-range values is a trial and error method that uses either the LOWTRAN or MODTRAN codes. The meteorological data, path geometry, target emissivity, and estimated target temperature are entered into an input file. The code is run as many times as necessary varying the estimated target temperature (denoted as parameter TBOUND in the LOWTRAN manual [1]) until the output agrees with the integrated radiance value corresponding to the remotely measured value of the AGEMA system. The dependence and variability of this technique on various parameters (e.g., emissivity) is described in a separate report, number 36, in these conference proceedings.

5. CONCLUSION

Low-cost remote sensing of infrared signatures is possible using a calibrated laboratory instrument like the AGEMA 900. The biggest problem is correcting for the intervening atmosphere. By using a trusted atmospheric code, such as LOWTRAN or MODTRAN, corrections can be made. And, by making remote measurements at two or more distances, the corrections can be fine-tuned to further improve accuracy.

It is useful to know the instrument spectral response to relate the measured and corrected values to other types of systems. Once the response is measured accuracy can be improved by modifying the LOWTRAN or MODTRAN atmospheric codes to account for the instrument response.

ACKNOWLEDGMENTS

This work is funded by the Office of Naval Research, code 322. Thanks to Herbert Hughes for providing technical assistance pertaining to LOWTRAN and analysis of the data. Thanks to Kathleen Littfin for assisting in the data acquisition. And thanks to Gibbs Flying Service for the aircraft support and the expert flying of our pilot, Robert Akin.

REFERENCES

1. Kneizys, F. X., E. P. Shettle, L. W. Abreu, J. H. Chetwynd, Jr., G. P. Anderson, W. O. Gallery, J. E. A. Selby and S. A. Clough, "Users Guide to LOWTRAN 7," Air Force Geophysics Laboratory Technical Report No. 88-0177 (1988).

2. Blakeslee, L. and L.J. Rodriguez, "User's Manual for TCM2," Georgia Tech Research Institute (GTRI), Georgia Institute of Technology, Interim report for period Jan-June 1993 under Wright-Patterson AFB Contract F33615-88-1865 (July 1993).

3. Freni, J. M. L., M. J. Gouveia, D. A. DeBenedictis, I. M. Halberstam, D. J. Hamann, P. F. Hilton, D. B. Hodges, D. M. Hoppes, J. J. Oberlatz, M. S. Odle, C. N. Touart, and S-L Tung, "Electro-Optical Tactical Decision Aid (EOTDA) User's Manual, Version 3," Phillips Laboratory Technical Report PL-TR-93-2002 Vol. I & II (11 January 1993).

4. Havener, K. and S. Funk, "The Utility of Electro-Optical Tactical Decision Aids (EOTDAs)," USAF Air Weather Service Report No. AWS/XTA 91-1 (10 September 1991).

5. Johnson, John, "Analysis of Image Forming Systems," *Image Intensifier Symposium* proceedings, (1957).

6. Shumaker, D., "Target Detection and Recognition Revisited," ERIM Spectral Reflections Technote 95-03 (October 1995).

7. Dreksler, S. B., "Electro-Optical Tactical Decision Aid Sensor Performance Model Evaluation," NRL Monterey Technical Report NRL/MR/7543-94-7216, Naval Research Laboratory, Monterey, CA (April 1995).

8. McGrath, C. P., "Open Ocean Effectiveness of the Electro-Optical Tactical Decision Aid Mark III," NRaD Technical Report 1664, Naval Command Control and Ocean Surveillance Center, RDT&E Division, San Diego, CA (June 1994).

9. Schemine, K.L., Dunham B. M., "Infrared Tactical Decision Aid Background Signature Model Assessment," Technical Report WL-TR-94-1064, Wright Laboratory, Wright Patterson AFB, OH (September 1993).

10. Zeisse, C. R., "SeaRad, A Radiance Prediction Code," NRaD Technical Report 1702, Naval Command Control and Ocean Surveillance Center, RDT&E Division, San Diego, CA (November 1995).

11. Shapiro, R., "Water Backgrounds in the Infrared and Visible Tactical Decision Aids," Technical Report AFGL-TR-87-0254, Air Force Geophysics Laboratory, Hanscom AFB, MA (August 1987).

12. Zeisse, C. R., "Relative Spectral Responsivity of Two AGEMA Infrared Scanning Cameras," NRaD Technical Report 1699, Naval Command Control and Ocean Surveillance Center, RDT&E Division, San Diego, CA (May 1995).

Paper 34

M. Acheroy (BE)

Is it possible to adapt your LOWTRAN version to any IR radiometer (with regard to the spectral response)? Is this LOWTRAN version (adapted) available?

Author's reply:

Yes, MODTRAN/LOWTRAN is available (including source code) from an FTP site maintained by the Air Force Phillips Laboratory, Hanscom Air Base, Massachusetts. Contact Gail Anderson. The modifications involve spectrally correcting the MODTRAN/LOWTRAN radiance calculations as the code steps through its integration. Our modifications to the code make MODTRAN/LOWTRAN correct the radiance by a spectral look-up table of values measured for our imaging system.

C.R. Philbrick (US)

Comment: I think the references to LOWTRAN 7 should be updated to be MODTRAN 3, which is available on internet. For those interested, a workshop on the transmission models will be during the first week of June at Hanscom AFB. Contact: Gail Anderson at Phillips Laboratory.

Question: What was the basis of the percentage accuracy that you gave for the sensor, atmosphere and target, and the percentage given in the performance figures?

Author's reply:

The percentage accuracy is empirical. Experiments are run with operational FLIR systems. Reported detection ranges from the pilots are compared with EOTDA predictions. The error and scatter are within the theoretical expectation of 38% for land background problems, but exceed 38% with marine backgrounds and targets.

Radiance of the Wind-Ruffled Sea

C. R. Zeisse, C. P. McGrath, and K. M. Littfin
Propagation Division
NCCOSC RDTE DIV 883
49170 Propagation Path
San Diego, CA 92152-7385 USA

and

H. G. Hughes
Science and Technology Corporation
101 Research Drive
Hampton, VA 23666-1340 USA

SUMMARY

The important contributions to sea radiance are: (1) path emission, (2) sky reflection, (3) sun reflection, and (4) thermal black body emission. The spectral nature of each contribution can be understood from the infrared properties of the atmosphere and the ocean, and regularities among them can often be explained by Kirchoff's Law. *SeaRad*, a computer model of sea radiance made by modifying MODTRAN2, takes source and path terms from that code, performs a statistical average over capillary wave slopes, and arrives at a prediction which agrees to within about 1 C with sea radiance measurements in the infrared.

LIST OF SYMBOLS

(Note: All radiances are *spectral* radiances.)

f	Receiver relative spectral responsivity.
N	Total sea radiance.
N_{bb}	Black body radiance.
N_o	Incident solar radiance.
N_s	Incident sky radiance.
N_{path}	Path radiance.
N_{sea}	Thermally emitted sea radiance.
N_{sky}	Reflected sky radiance.
N_{sun}	Reflected solar radiance.
P	Occurrence probability.
p	Occurrence probability density.
Q	Interaction probability.
q	Interaction probability density.
T	Physical absolute temperature.
t	Equivalent absolute temperature.
$\mathbf{U}$	Unit vector.
$\mathbf{U}_n$	Unit vector normal to facet at origin.
$\mathbf{U}_r$	Unit vector pointing from origin to receiver.
$\mathbf{U}_s$	Unit vector pointing from origin to source.
W	Wind speed.
X	Coordinate axis pointing up wind.
Y	Coordinate axis pointing cross wind.
Z	Coordinate axis pointing to zenith.
ϕ	Azimuth of $\mathbf{U}$.
ν	Wave number.
θ	Zenith angle of $\mathbf{U}$.
ρ	Fresnel reflectivity of sea water.
σ_c	Standard deviation of ζ_y.
σ_u	Standard deviation of ζ_x.
τ_{path}	Path transmission.
ω	Angle of incidence; angle of reflection.
ζ_x	Facet slope in X direction.
ζ_y	Facet slope in Y direction.

1. SEA RADIANCE

In the infrared and visible regions of the spectrum there are four major contributions to sea radiance, each of which is shown in figure 1.

The first contribution is path radiance, shown at the top of figure 1. The footprint of a single pixel in an image of the sea is indicated by the wavy line. The footprint is observed by a receiver at the end of a ray whose zenith angle at the footprint is θ_r. Let N_{path} designate the spectral radiance in W m^{-2} sr^{-1} $(cm^{-1})^{-1}$ along the path[1] from the footprint to the receiver.

The second contribution is reflected sky radiance. Spectral radiance N_s from a portion of the sky arrives at the footprint along a ray whose zenith angle there is θ_s. The footprint contains wave facets of different slopes, many that reflect the incoming sky radiance away from the receiver toward other parts of the sky. These facets are ignored. However, the footprint will also contain some facets whose slope is correct for reflecting the incoming sky radiance toward the receiver along the path defined by the zenith angle θ_r. These facets are retained. The contributions from all portions of the sky are summed together after specular reflection by the appropriate facets within the footprint, and the sum leaving the footprint at zenith angle θ_r is designated N_{sky}. During its path to the

[1] In this paper, the word path refers only to the optical path between the footprint and the receiver.

Paper presented at the AGARD SPP Symposium on "Remote Sensing: A Valuable Source of Information", held in Toulouse, France from 22-25 April 1996, and published in CP-582.

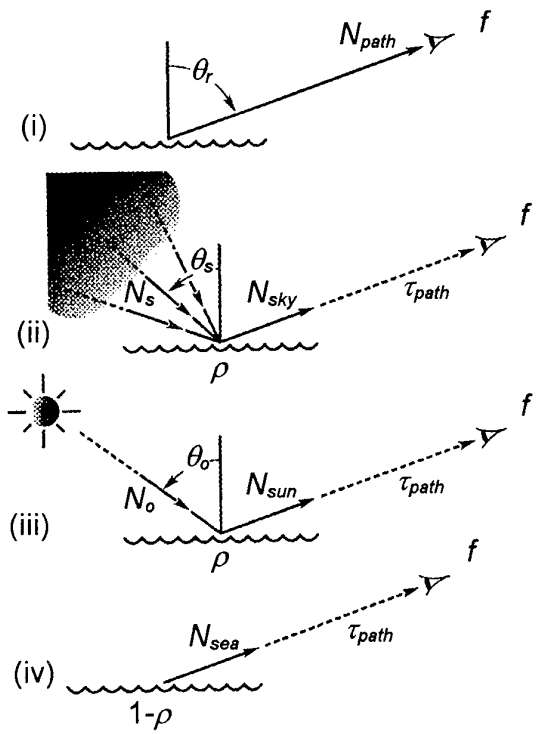


Figure 1. Four contributions to sea radiance: (i) path radiance, (ii) reflected sky radiance, (iii) reflected solar radiance, and (iv) thermal black body radiance.

receiver, the reflected sky radiance is attenuated by the path transmission τ_{path}.

The third contribution is reflected solar radiance, sun glint. The mechanism is analogous to the mechanism of sky radiance. Spectral radiance N_o from the solar center arrives at the footprint along a path whose zenith angle there is θ_o. Within the footprint most facets deflect the solar ray away from the receiver and

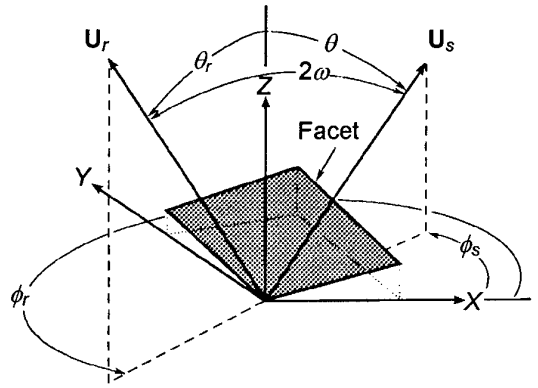


Figure 2. The geometry of facet reflection. The facet normal U_n has been left out of the figure for clarity. Azimuths are considered positive when measured counterclockwise looking toward the nadir.

are rejected, but some facets are retained because they deflect the ray specularly toward the receiver along a path with zenith angle θ_r. N_{sun} is the spectral radiance leaving the footprint after summation over rays arriving from all portions of the solar disk. The reflected solar radiance is also attenuated by the path transmission τ_{path} before final reception.

The fourth contribution is thermal black body emission. Each facet emits a spectral radiance N_{bb} given by Planck's equation for a black body whose temperature is the sea temperature. The spectral emissivity of a given facet in the direction of the receiver is specified by the slope of that facet and the value of θ_r. N_{sea} is the thermal spectral radiance leaving the footprint for the receiver after summation over all facets within the footprint. As before, N_{sea} is attenuated by path transmission after leaving the footprint.

Optical reflection governs the second and third contributions and optical emission governs the fourth. By application of Kirchoff's Law to sea water, an opaque medium, the spectral emissivity is one minus the spectral reflectivity.

The total spectral radiance N arriving at the receiver is given by the sum of the four contributions just described,

$$N = N_{path} + \left(N_{sky} + N_{sun} + N_{sea} \right) \tau_{path}, \quad (1)$$

and if f is the relative spectral responsivity of the receiver then the measured sea radiance will be

$$N f = N_{path} f + \left(N_{sky} + N_{sun} + N_{sea} \right) \tau_{path} f. \quad (2)$$

Every quantity in (2) is a spectral quantity which depends on the value of wave number[2] ν (cm^{-1}).

2. THE COORDINATE SYSTEM

The previous discussion was a two dimensional description of sea radiance which neglected the azimuthal dependence of rays arriving and leaving the footprint. The full three-dimensional geometry will now be introduced.

[2] The wavelength in μm is given by dividing the wave number in cm^{-1} into 10,000. For example, 2000 cm^{-1} corresponds to 5 μm. Although wavelengths are almost universally used by the infrared community, we prefer wave numbers which are proportional to energy.

Figure 2 shows the geometry of reflection. A coordinate system was chosen whose origin is the point of reflection with the X-axis pointing upwind, the Z-axis pointing toward the zenith, and the Y-axis pointing crosswind such that a right-handed system is formed. The X-Y plane is therefore horizontal at the point of reflection. The tilted facet passes through the origin.

Two unit vectors are shown in figure 2: $\mathbf{U}_s$, pointing from the origin to the source, and $\mathbf{U}_r$, pointing from the origin to the receiver. A third unit vector, $\mathbf{U}_n$, is normal to the facet at the point of reflection but was removed from the figure for clarity[3].

If a specular reflection occurs, the three vectors for source, receiver, and facet normal will be connected by the law of reflection:

$$\mathbf{U}_s + \mathbf{U}_r = 2\cos\omega\,\mathbf{U}_n \tag{3}$$

3. FACET AVERAGING

The three quantities within the parenthesis on the right hand side of (1) are formed by taking a statistical average, over all facets, of the reflected or emitted radiance of a single facet:

$$\begin{aligned} N_{sky} &= \left\langle \rho\, N_s \right\rangle_r \\ N_{sun} &= \left\langle \rho\, N_o \right\rangle_r \\ N_{sea} &= \left\langle (1-\rho)\, N_{bb} \right\rangle_r \end{aligned} \tag{4}$$

The angular brackets in (4) represent the following integral over slope space [1]:

$$\langle g \rangle_r \equiv \frac{\iint\limits_{\substack{\omega \le \pi/2 \\ \mathbf{U}_r = const}} g \frac{\cos\omega}{\cos\theta_n} p(\zeta_x, \zeta_y, W)\, d\zeta_x\, d\zeta_y}{\iint\limits_{\substack{\omega \le \pi/2 \\ \mathbf{U}_r = const}} \frac{\cos\omega}{\cos\theta_n} p(\zeta_x, \zeta_y, W)\, d\zeta_x\, d\zeta_y} \tag{5}$$

The averages in (4) are not quite as simple as they appear. First, the integrals in (5) are only defined for $\omega \le \pi/2$, that is, only the front of each facet is used. Second, the receiver direction $\mathbf{U}_r$ is held fixed during the integration over sea slopes. The subscript "r" on the angular bracket serves as a reminder of this restriction. Third, different conditions apply during the average in each of the three cases. When the radiance is from the sky, there is specular reflection within the footprint and the integration is over all slopes in the sea up to and including sky locations[4] on the horizon ($\theta_s \le \pi/2$). When the radiance is from the sun, there is specular reflection within the footprint, but integration is only over the tolerance ellipse (those slopes capable of reflecting part of the solar disk toward the receiver). When the radiance is from the sea, no specular reflection is considered to occur and integration is over all slopes in the sea.

4. WAVE SLOPE PROBABILITIES

The function p in equation (5) still requires definition. Following Cox and Munk [2, 3] let

$$P = p\left(\zeta_x, \zeta_y, W\right) d\zeta_x\, d\zeta_y \tag{6}$$

stand for the probability that a wave facet will occur whose slope is within $\pm\, d\zeta_x/2$ of ζ_x and $\pm\, d\zeta_y/2$ of ζ_y when the wind speed is W. Cox and Munk obtained an expression for p, the occurrence probability density, whose lowest order term is a two dimensional Gaussian:

$$p\left(\zeta_x, \zeta_x, W\right) \approx \frac{1}{2\pi\sigma_u\sigma_c} \exp\left\{-\frac{1}{2}\left(\frac{\zeta_x^2}{\sigma_u^2} + \frac{\zeta_y^2}{\sigma_c^2}\right)\right\},$$

$$\sigma_u^2 = 0.000 + 3.16 \cdot 10^{-3} W,$$

$$\sigma_c^2 = 0.003 + 1.92 \cdot 10^{-3} W. \tag{7}$$

The standard deviations in (7) will be numerically correct when W is given in m s^{-1}. Figure 3 shows how the occurrence probability density depends on slope when the wind speed is 10 m s^{-1}. The coordinate system of figure 2 has been inserted at the top of figure 3 to illustrate the relation between coordinates and slopes. Note that when the azimuth of the facet normal, ϕ_n, is in the first X-Y quadrant both slopes are negative.

The meaning of the word "occur" requires elaboration. The footprint is a fixed flat surface in the X-Y plane, of total area A, above which the fluctuating sea surface can be found. If we select an infinitesimal horizontal area dA within the footprint and ask how likely it is for

[3] The zenith angle of $\mathbf{U}_n$ is the same as the tilt of the facet. The tilt is the angle of the steepest ascent within the facet.

[4] With the receiver location fixed, selection of an ocean slope defines a unique source location for a specular reflection.

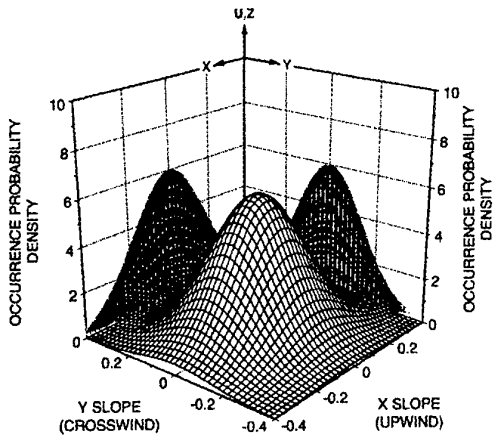


Figure 3. Occurrence probability density versus capillary wave slope for a wind speed of 10 m s^{-1}.

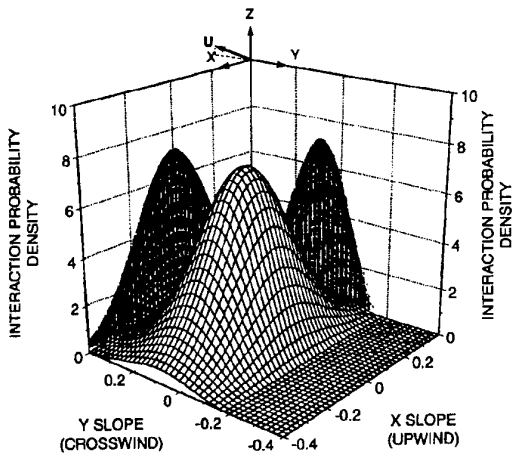


Figure 4. Interaction probability density versus capillary wave slope for a wind speed of 10 m s^{-1} and a ray pointing toward (80°, 270°).

the sea surface above it to have a given instantaneous slope, then the answer is provided by the function $p(\zeta_x, \zeta_y, W)$. If we wish, we may group together all facets with the same slope and ask instead what fractional area of the footprint they occupy together. The answer will be the same. The point is that "occurrence" is judged by comparing *horizontal* areas inside the footprint. Therefore p is proportional to the *horizontal* area occupied by all facets with a given slope.

Next we introduce the concept of the "interaction" of a facet with a ray. Following Plass et al [4], we assume that a facet interacts with a ray pointing in the direction **U** according to the area of the facet "seen" from **U**, that is, according to the facet area projected toward **U**. Let

$$Q = q\left(\theta, \phi, \zeta_x, \zeta_y, W\right) d\zeta_x d\zeta_y \tag{8}$$

stand for the probability that a facet whose slope is within $\pm d\zeta_x/2$ of ζ_x and $\pm d\zeta_y/2$ of ζ_y will interact with a ray pointing in the direction **U** = (θ, ϕ) when the wind speed is W. The wave slope interaction probability density is q. Then

$$q\left(\theta, \phi, \zeta_x, \zeta_y, W\right) = \frac{\frac{\cos\omega}{\cos\theta_n} p\left(\zeta_x, \zeta_y, W\right)}{\iint\limits_{\substack{\omega \le \pi/2 \\ U=const}} \frac{\cos\omega}{\cos\theta_n} p\left(\zeta_x, \zeta_y, W\right) d\zeta_x \, d\zeta_y} \tag{9}$$

because the factor $\cos\omega/\cos\theta_n$ projects the horizontal area represented by p toward **U** so that the numerator is proportional to the amount of the facet "seen" from **U**. The denominator normalizes the slope space volume under q to one. Figure 4 is a graph of (9) for a wind speed of 10 m s^{-1} showing how facets with a specified slope interact with a ray pointing in the direction (80°, 270°), that is, with a ray elevated by 10° along an azimuth in the negative-Y direction.

With these definitions it can be seen that (5) is merely the average over all slopes of the function g weighted by the function q for a ray pointing toward the receiver.

5. THE INFRARED ATMOSPHERE

In this section we describe the spectral behavior of all terms in equation (2). These examples, except for figure 8, are taken from LOWTRAN7 [5, 6] calculations at a spectral resolution of 20 cm^{-1} using the 1976 standard atmosphere, multiple scattering,

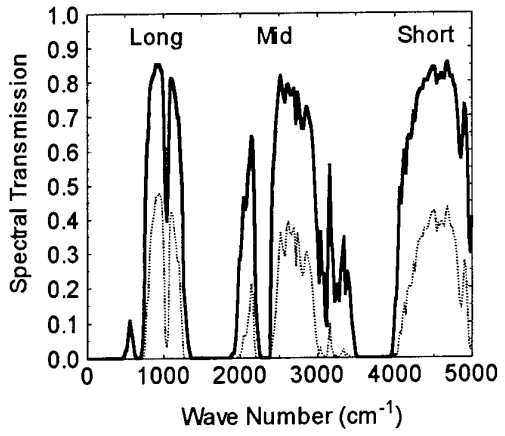


Figure 5. Spectral atmospheric transmission for a slant path to space at a zenith angles of 0° (solid curve) and 80° (dashed curve).

and the Navy aerosol model with the following parameters: moderate coastal aerosol influence (ICSTL = 5), visibility 20 km, and current and 24 hour wind speeds 10 m s^{-1}.

Figure 5 shows the transmission of the atmosphere for two different slant paths to space: a zenith path, and a path elevated above the horizon by 10°. These paths are on the order of 100 km long, and the strong absorption in certain spectral regions, chiefly due to water, has created the three major bands shown here, namely the long, mid, and short wave bands, named for their wavelength content. Note the deep notch near 2300 cm^{-1} in the mid wave band. It is due to carbon dioxide absorption.

Figure 6 shows spectral sky radiance as a function of zenith angle for two different wave numbers. The right curve, showing very little dependence on zenith angle, is for 700 cm^{-1} where the absorption is large and the transmission (at the notch in the long wave band in figure 5) is small regardless of zenith angle. The left curve, showing a large dependence on zenith angle, is for 950 cm^{-1} where the absorption is small and the transmission (at the left peak in the long wave band in figure 5) is large but dependent on zenith angle. The difference between these two curves can be understood in terms of Kirchoff's Law [7], which loosely says that a good absorber is a good emitter. At 700 cm^{-1} the cross section is large, energy is absorbed close to the point of observation at all zenith angles, and sky radiance is determined by the temperature of the atmosphere surrounding the point of observation, which, in the 1976 standard atmosphere, happens to be 288 K near the ground. At 950 cm^{-1}, on the other hand, the cross section is smaller, absorption is weak toward the zenith, and sky radiance is relatively small in that direction. Toward the horizon, at larger zenith angles, the absorption is greater because the path is longer, and, furthermore, the more horizontal path spends much of its time in the lowest layers of the atmosphere whose temperature is near 288 K. Hence for 950 cm^{-1} sky radiance increases with the zenith angle. The solid circles on the abscissa in figure 6, also shown in figure 7, are the spectral radiance of a 288 K black body for each of these wave numbers. In each case, whether by virtue of a large cross section or a long path, sky radiance along a slant path toward the horizon approaches the radiance of a black body at the temperature of the lowest atmospheric levels.

Figure 7 shows the spectral radiance of two black bodies, one at 288 K near the temperature of the sea and one at 5900 K near the temperature of the sun. Theses curves contain information on the spectral behavior of sea emission and sun glint. Each is given by Planck's formula which is

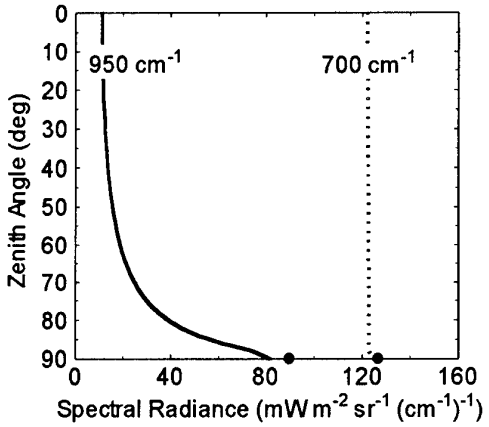


Figure 6. Spectral sky radiance in regions of strong absorption (dashed curve) and weak absorption (solid curve).

$$N_{bb}(T) = \frac{2 \cdot 10^8 c^2 h \nu^3}{\exp(100\, hc\nu/kT) - 1} \qquad (10)$$

when ν is in cm^{-1}, T is in K, and the natural constants are in SI units. Planck's formula can be approximated by the Rayleigh-Jeans Law

$$N_{bb} \approx 2 \cdot 10^6 \cdot ckT\, \nu^2 \qquad (11)$$

at low frequencies, has a maximum at

$$\nu = \nu_{\max} \equiv 1.96108\, T, \qquad (12)$$

and can be approximated by Wien's Law

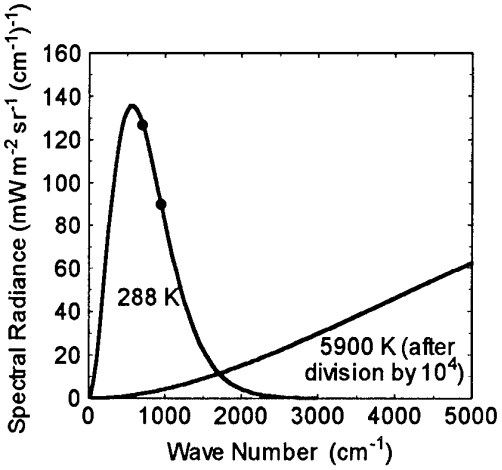


Figure 7. Spectral radiance of black bodies near room temperature (left) and near the temperature of the sun (right, divided by 10^4).

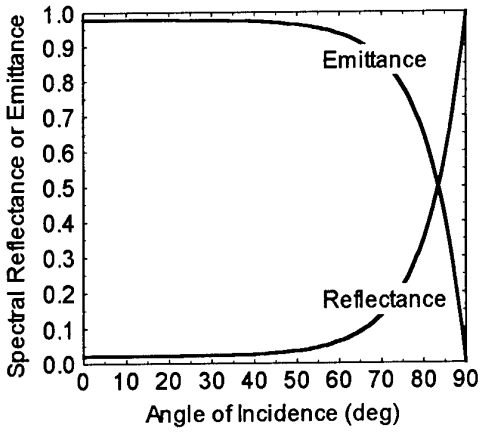


Figure 8. Fresnel spectral reflectance and emittance at 2500 cm^{-1}.

$$N_{bb} \approx 2 \cdot 10^{8} c^{2} h \nu^{3} \exp(-100 hc\nu / kT) \quad (13)$$

at high frequencies. The division of solar radiance by four orders of magnitude in figure 7 allows both curves to be plotted linearly on the same figure, but also has physical motivation in the fact that the volume of a glint column [8] is often three or four orders of magnitude smaller than unity, which is the total volume under the interaction probability density. In other words, the number of capillary wave slopes capable of reflecting part of the solar disk into the receiver is typically three to four orders of magnitude smaller than the total number of capillary wave slopes in the sea. Essentially, the sea carries out a forceful *mechanical* division of solar radiance. Hence the curves as they stand in figure 7 give a good indication of the relative importance of thermal sea emission and sun glint in the infrared region. It can be seen that sea emission dominates the long wave band, sun glint dominates the short wave band, and the two can be comparable in the mid wave band.

Figure 8 shows the reflectance of sea water as a function of angle of incidence for a wave number of 2500 cm^{-1}. Values of the reflectance at other wave numbers are close to the values shown here, differing from one another throughout the infrared by less than 10 %. The emittance is given by one minus the reflectance. Therefore, for those small angles of incidence obtained when viewing from a satellite or aircraft along a nadir path, processes of emission dominate over processes of reflection. This enables the remote sensing of ocean temperature from a satellite or aircraft.

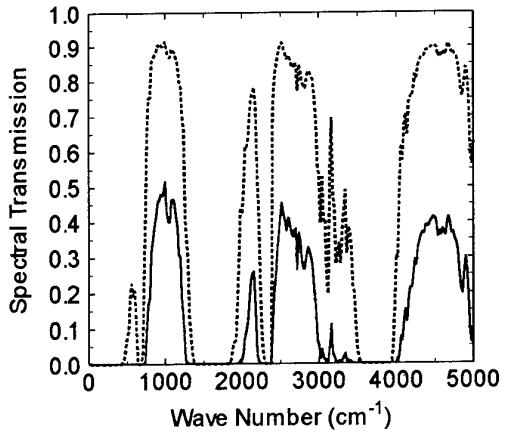


Figure 9. Spectral transmission of a 1 km (dashed curve) and a 10 km (solid curve) path. Each path originates at the ocean surface and terminates at a receiver whose altitude is 10 m.

Figure 9 shows representative values of path transmission for 1 km and 10 km paths from a footprint on the ocean surface to a receiver at an altitude of 10 m. The close resemblance of this figure to figure 5 is coincidental; the geometry is quite different in the two cases.

Representative values of path radiance are shown in figure 10 for the same geometry as figure 9, namely two almost horizontal paths which might occur from an ocean footprint to a receiver as shown in figure 1. The dashed curve is for a 1 km path and the solid curve is for a 10 km path. As the path length increases, the spectrum approaches that of a 288 K black body (see figure 7). The reason, as before, is that the longer path is accompanied by more absorption, all

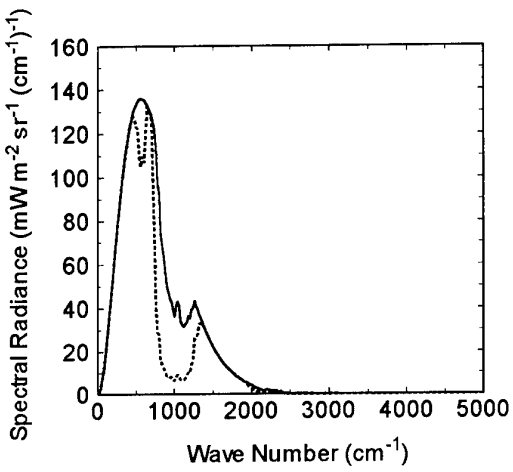


Figure 10. Spectral radiance of a 1 km (dashed curve) and a 10 km (solid curve) path originating at the ocean surface and terminating at a receiver whose altitude is 10 m.

in the lowest 288 K level of the 1976 standard atmosphere. By Kirchoff's Law the longer paths generate greater emission at 288 K, the temperature of the medium. Consequently, as a comparison of figures 9 and 10 will show, path transmission and path radiance vary in opposite directions: when one is large the other is small. Finally, the scarcity of appreciable path radiance in the mid and long wave bands is due to the rapid fall-off of room temperature black body radiance at high wave numbers according to Wien's Law, equation (13).

The function f, representing the frequency response of the receiver, is often taken to be a so-called "top hat" filter which is unity between lower and upper wave numbers ν_1 and ν_2 and zero at all others. Experience has shown that for various receivers and atmospheric conditions reasonable choices are

$$\begin{aligned} \nu_1^{LW} &= 830\ cm^{-1} \\ \nu_2^{LW} &= 1250\ cm^{-1} \end{aligned} \tag{14}$$

for the long wave band and

$$\begin{aligned} \nu_1^{MW} &= 2000\ cm^{-1} \\ \nu_2^{MW} &= 3330\ cm^{-1} \end{aligned} \tag{15}$$

for the mid wave band. A more accurate approach is to measure receiver relative spectral responsivity. Figure 11 shows values of f measured [9] for two infrared scanning cameras. The shape of each curve approximates an ideal photon detector [10] made from HgCdTe for the "LW" camera and InSb for the "SW" camera.

The spectral quantities shown in figures 6 through 10 must be averaged over slope space according to (4), added together according to (2), multiplied by a responsivity such as that shown in figure 11, and integrated over wave number in order to predict the result

$$\int_0^\infty N\, f\, dv \equiv \int_0^\infty N_{bb}(t)\, f\, dv \tag{16}$$

of a broad band measurement. Equation (16) also shows how a broad band measurement can be associated with a single temperature t. This association is not unique: radiant signals with different spectral content can give the same band integrated response (and the same equivalent temperature) just as the area under two different curves can be equal.

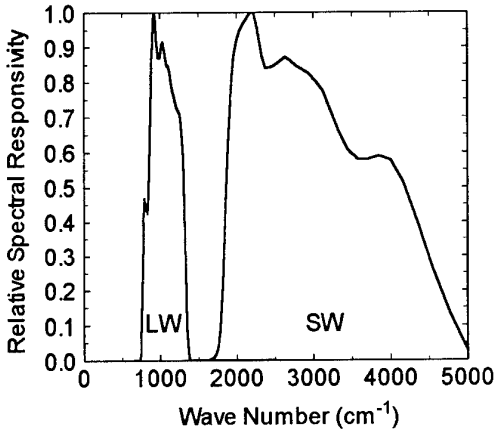


Figure 11. Relative spectral responsivity of AGEMA model 900 LW and 900 SW scanning cameras.

6. *SEARAD*

Broad band sea radiance can be computed by *SeaRad* [11], a computer code obtained by modifying the atmospheric radiance and transmittance code MODTRAN2 [5, 6].

SeaRad is organized as follows. New FORTRAN subroutines perform the capillary wave averages indicated by (4). The reflectivity is calculated from Fresnel's equations [12] with a complex optical index[5] taken from the literature [13, 14]. Values for source radiance (N_s, N_o, and N_{bb}), path radiance (N_{path}), and path transmission (τ_{path}) are provided by the original MODTRAN2 code[6].

SeaRad operates exactly like MODTRAN2 except that an additional logical switch is required in the input file to turn the sea radiance calculation on or off. *SeaRad* is designed for a single pixel (rather than an image) and includes polarization and the responsivities shown in figure 11 but does not include multiple reflections, shadowing, or gravity waves, nor does it include the effects of refractivity or turbulence. When compiled for a personal computer *SeaRad* executable code occupies 0.8 MB of memory. On a 100 MHz Pentium processor, a multiple scattering radiance calculation for one pixel takes 5 s when the sun is not involved (IEMSCT = 1) and 17 s when the sun is involved (IEMSCT = 2).

---

[5] These data for the index, available between 52.63 cm^{-1} and 25000 cm^{-1}, set the spectral range of *SeaRad*.

[6] MODTRAN2 has a maximum spectral resolution of 2 cm^{-1}. By changing a logical flag in the input file, the lower resolution model LOWTRAN7 will run instead.

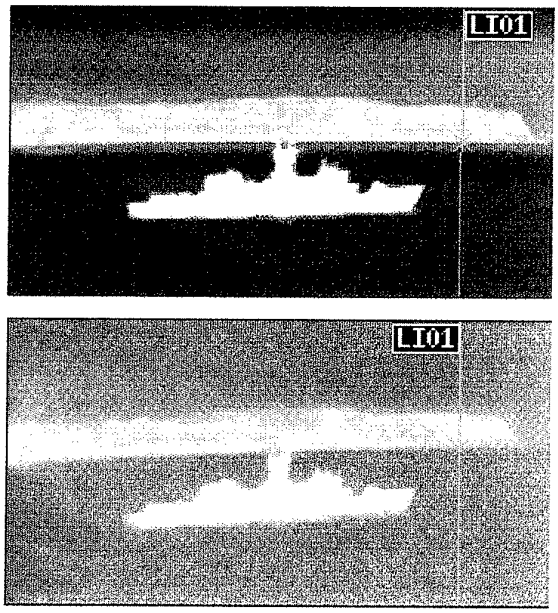


Figure 12. Airborne view of frigate near Point Loma, California, from an altitude of 51 m on 13 September 1994. The vertical lines "LI01" mark the location from which data for the following figures have been taken. Upper: long wave image. Lower: mid wave image.

7. COMPARISON WITH DATA

SeaRad calculations have been compared with sea radiance data acquired by two cameras manufactured by AGEMA Infrared Systems [15]. Apart from their spectral responsivities, given in figure 11, the cameras are similar. They have a dynamic range of 12 bits and are calibrated with black bodies to an absolute accuracy of ½ K. Each image is 5° high and 10° wide. The horizontal instantaneous field of view (90% slit response) is 2.1 mrad for the 900 SW/ST and 1.75 mrad for the 900 LW/ST. Each line is sampled 272 times horizontally and each frame contains 136 vertical lines. The frame rate is 15 Hz. Up to 15 minutes of real time data may be stored in the computer that forms a part of the AGEMA 900 system.

The AGEMA cameras were used to acquire radiometric data off the coast of southern California from a Piper Navaho aircraft on 13 September 1994 [16]. Figure 12 contains two simultaneous images, one from each camera, showing a frigate with Point Loma in the background at a distance of approximately 8 km. In these images the cameras were directed 31° (true) when the sun's azimuth was 233° (true) and solar glint was not apparent on the surface. Data from the vertical row of pixels indicated by the line "LI01" in each image were chosen for the purpose of making a comparison with *SeaRad*, and these data are shown as the solid lines in figures 15 through 18.

A vertical profile of meteorological parameters (temperature, relative humidity, and pressure) was measured from the aircraft at the same time these infrared data were taken. The profile was extended from the lowest aircraft altitude to an altitude of 1 m assuming the potential temperature and specific humidity were constant within the well-mixed boundary layer. The air temperature rose from 19.8 C at an altitude of 52 m to 20.4 C at an altitude of 1 m. The sea surface temperature of 21.3 C was measured from low altitude flights (≈ 30 m) using an Everest Infrared Temperature Sensor (Model 4000ALCS). That day the following surface wind measurements were obtained from the Naval Air Station North Island: current wind speed 6.2 m s^{-1}, 24 hour average wind speed 3.6 m s^{-1}, wind direction 280° T. For *SeaRad* calculations each profile was also extended above the highest altitude attained by the aircraft to an altitude of 40 km using a standard atmosphere. The skies were free of clouds.

The spectral output of *SeaRad* for the 13 September 1994 inputs just described are shown for the long and mid wave bands in figures 13 and 14 respectively. For these figures the zenith angle at the receiver (ANGLE in the *SeaRad* input file) is 91° and the computed range to the ocean surface is 2.9 km.

The band integrated predictions of *SeaRad* for the 13 September 1994 data are shown dashed as a function of zenith angle in figures 15 and 16 for the long and mid wave bands respectively. In these figures the data are shown by the solid line. Zenith angles less than 90.21° describe the sky; larger zenith angles describe the sea. In the sky, *SeaRad* is identical to MODTRAN2, and comparison of the calculation with the sky data shows the agreement MODTRAN2 is able to attain in this case. The radiance step in the data just above the horizon is Point Loma, which is not modeled by either *SeaRad* or MODTRAN2. In the sea, MODTRAN2 alone would generate the dashed curves labeled "Path" in each figure. The dashed curves labeled "Sea" and "Sky" are contributed by *SeaRad*, and, when added in the sea to the path contribution from MODTRAN2, result in the unlabeled dashed curve showing the agreement *SeaRad* is able to attain in this case.

The integral of each curve in figure 13 is shown by the solid circle in figure 15. The shapes of the sea contributions in figure 15 are determined by the way the path transmission changes with zenith angle. Near the horizon the path is long, path transmission is small, and the contributions leaving the footprint are strongly attenuated before reaching the receiver. The path radiance, on the other hand, is largest when the

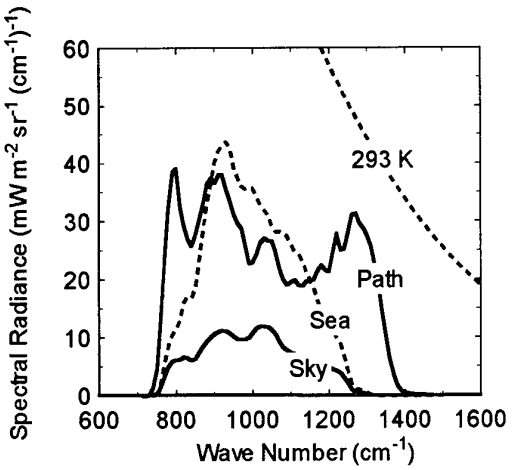


Figure 13. Spectral radiance from the ocean computed in the long wave band for 13 September 1994. The observation angle is 91° and the range is 2.9 km.

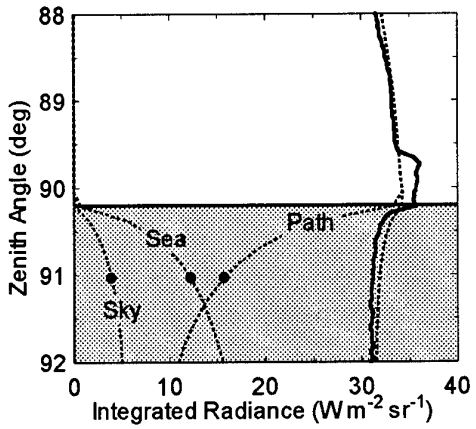


Figure 15. Vertical radiance profile in the long wave band for 13 September 1994. Data are shown by the solid line; the *SeaRad* calculation is shown by the dashed lines. Each solid circle is the integral of the corresponding curve in figure 13.

path is longest. Consequently the radiance of a footprint on the horizon usually can't be seen; the receiver is blinded by path radiance. As the zenith angle increases, the footprint rapidly approaches the receiver, path length and path radiance diminish together, and path transmission increases, all of which allow much more of the footprint to be seen at the receiver. The same shapes occur in the sea in figure 16, and they have the same explanation.

In figure 13 the low and high frequency edges of the calculation, most clearly seen in the curve for path radiance, are dictated by the responsivity of the receiver. In figure 14 the low frequency edge of the path radiance is due to the receiver responsivity, but the high frequency fall-off is due to Wien's Law for objects near room temperature. On 13 September 1994 the air temperature near the ground was 293 K. The dashed curve labeled 293 K in figure 14 is Planck's curve for that temperature. Notice that the path radiance approaches the black body radiance in spectral regions of strong absorption; for example, in the CO_2 band between 2200 and 2400 cm^{-1}. On this day, and for the sea, sky, and path contributions in general, the large atmospheric transmission and high instrumental responsivity out to wave numbers close to the "top hat" upper limit of 3300 cm^{-1} have no impact: the radiance has already died away by mid-band. However, that would emphatically not be the case for sun glint as a quick glance back at figure 7 will show.

Radiance values add but equivalent temperatures do not, so the relationship of one sea contribution to another and to their total must be shown on a radiance scale as in figures 15 and 16. Figures 17 and 18 show how *SeaRad* and the data compare on the basis of

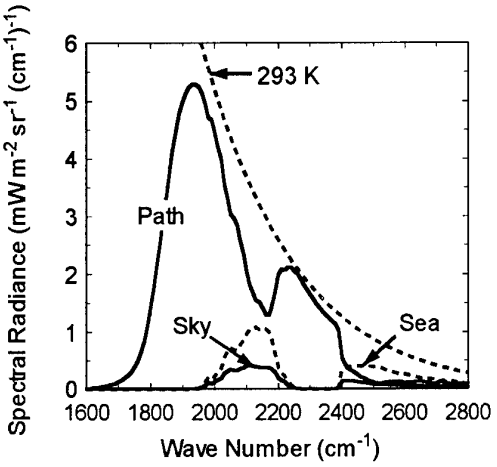


Figure 14. The same as figure 13 for the mid wave band.

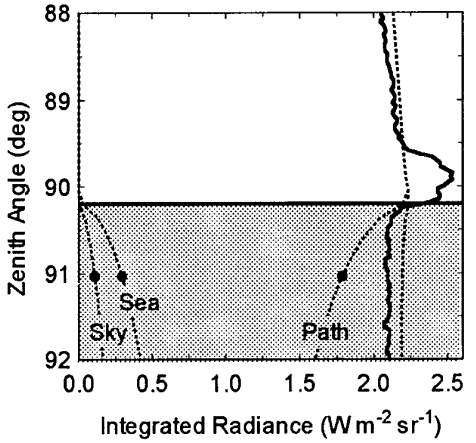


Figure 16. The same as figure 15 for the mid wave band.

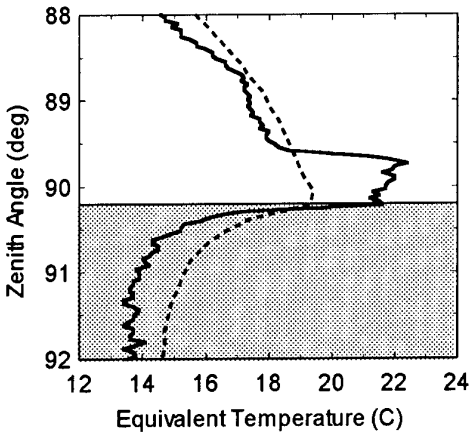


Figure 17. Long wave comparison of *SeaRad* with 13 September 1994 data.

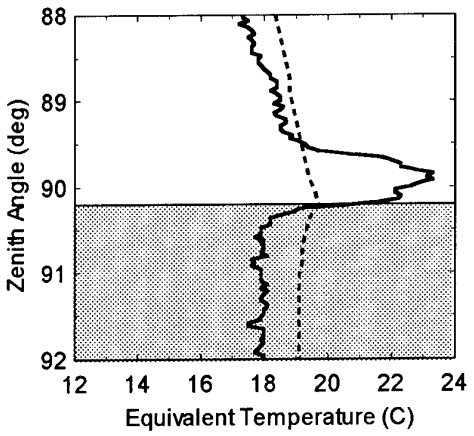


Figure 18. Mid wave comparison of *SeaRad* with 13 September 1994 data.

equivalent temperature. The agreement is typical of what we have found in many cases we have examined, namely, about 1 C. The cause of this persistent discrepancy is unknown. In the past we have occasionally forced a fit between the prediction and the data at one particular zenith angle, usually the horizon, by treating the air mass parameter (ICSTL) and visibility (VIS) as adjustable parameters in *SeaRad*. That practice was not adopted here, but would probably have improved the agreement in figures 17 and 18 at the expense of disguising the general agreement that can be expected from these kinds models at this stage of their development.

8. SEA RADIANCE AND REMOTE SENSING

We conclude with some brief comments on the role of sea radiance in remote sensing. From the preceding text, it is easy to appreciate that the relative importance of each sea radiance contribution depends on the spectral band. In the long wave band, thermal black body emission dominates. This is especially true for short nadir paths, because (1) short paths minimize path radiance and (2) nadir paths minimize sky reflection and maximize sea emission because their geometry involves small angles of incidence where, as shown in figure 8, the Fresnel reflection coefficient is small and the Fresnel emission coefficient is large. Hence, remote nadir observation of long wave infrared radiance can be used to infer the physical temperature of the sea or of a ship. In the mid wave band remote sensing of sea properties suffers from complications due to competition between the four radiance contributions, spectral emissivity variations, and strong molecular absorption of carbon dioxide [17, 18]. In the short wave band and the visible region sun glint is a major ingredient of sea radiance. Therefore, remote observation of visible glint patterns, known to be very sensitive to wind speed, can be used to infer the wind speed above the sea surface [19] and, hence, the distant sea state.

ACKNOWLEDGMENT

This work was supported by the Office of Naval Research.

REFERENCES

1. Zeisse, C. R., "Radiance of the Ocean Horizon," *Journal of the Optical Society of America* **A**, 1995, 2022.
2. Cox, C. and Munk, W., "Measurement of the Roughness of the Sea Surface from Photographs of the Sun's Glitter," *Journal of the Optical Society of America,* vol. 44, 1954, p. 838.
3. Cox, C. and Munk, W., "Slopes of the Sea Surface Deduced from Photographs of Sun Glitter," *Scripps Institution of Oceanography Bulletin,* vol. 6, 1956, p. 401.
4. Plass, G. N., Kattawar, G. W., and Guinn, J. A., "Radiative Transfer in the Earth's Atmosphere and Ocean: Influence of Ocean Waves," *Applied Optics,* vol. 14, 1975, p. 1924.
5. Berk, A., Bernstein, S., and Robertson, D. C., "MODTRAN: A Moderate Resolution Model for LOWTRAN 7," Report GL-TR-89-0122, Air Force Geophysics Laboratory, Bedford, MA, 1989.
6. Kneizys, F. X., Shettle, E. P., Abreu, L. W., Chetwynd, J. H., Anderson, G. P., Gallery, W. O., Selby, J. E. A., and Clough, S. A., "Users Guide to LOWTRAN 7," Report AFGL-TR-88-0177. Air Force Geophysics Laboratory, Bedford, MA, 1988.
7. Born, M., and Wolf, E., "*Principles of Optics*," 4th edition, Pergamon Press, New York, 4th ed., 1970, p. 622.
8. Zeisse, C. R. "Radiance of the Ocean Horizon," NCCOSC RDT&E TR-1660, Naval Command,

Control and Ocean Surveillance Center, RDT&E Division, San Diego, CA, 1994.

9. Zeisse, C. R., "Relative Spectral Responsivity of Two AGEMA Infrared Scanning Cameras," NCCOSC RDT&E TR-1699, Naval Command, Control and Ocean Surveillance Center, RDT&E Division, San Diego, CA, 1995.
10. Hudson, R. D. Jr., "*Infrared System Engineering*," Wiley, New York, 1969, p. 287 ff.
11. Zeisse, C. R., "SeaRad, A Sea Radiance Prediction Code," NCCOSC RDT&E TR-1702, Naval Command, Control and Ocean Surveillance Center, RDT&E Division, San Diego, CA, 1995.
12. Stratton, J. A., "*Electromagnetic Theory*," McGraw-Hill, New York, NY, 1941, pp. 505 ff.
13. Hale, G. M. and Querry, M. R., "Optical Constants of Water in the 200 nm to 200 μm Wavelength Region," *Applied Optics* **3**, 1973, p. 555.
14. Querry, M. R., Holland, W. E. Waring, , R. C., Earls, L. M., and Querry, M. D. "Relative Reflectance and Complex Refractive Index in the Infrared for Saline Environmental Waters," *Journal of Geophysical Research* **82**, 1977, p. 1452.
15. AGEMA Infrared Systems, 550 County Avenue, Secaucus, NJ, 07094.
16. Hughes, H. G., "Infrared Ship and Background Signatures in a Coastal Environment," Science and Technology Corporation Technical Report 2937, Science and Technology Corporation, Hampton, VA, 1995.
17. Littfin, K. M., McGrath, C. P., Jensen., D. R., and Hughes, H. G., "Dependence of Mid and Far Infrared Sea Radiances on Viewing Altitude," *Proceedings of the 1995 SPIE Orlando Symposium*, Orlando, FL, 1995.
18. Hughes, H. G., "Dependence of Mid and Far Infrared Sea Radiances on Viewing Altitude," Science and Technology Corporation Technical Report 2869, Science and Technology Corporation, Hampton, VA, 1994.
19. Plass, G. N., Kattawar, G. W., and Guinn, J. A., "Isophotes of Sunlight Glitter on a Wind-Ruffled Sea," *Applied Optics* **16**, p. 643, 1977.

AN ALGORITHM FOR SHIP TEMPERATURE RETRIEVAL FROM AIRBORNE INFRARED RADIOMETRIC MEASUREMENTS

H. G. Hughes
Science and Technology Corporation
101 Research Drive
Hampton, Virginia 23666-1340 USA

C. R. Zeisse and C. P. McGrath
Propagation Division
NCCOSC RDTE DIV 883
49170 Propagation Path
San Diego, California 92152-7385 USA

1. SUMMARY

A remote sensing technique is described which enables the zero-range average temperature of a ship to be determined. This temperature is determined from airborne infrared radiometric measurements and vertical profiles of meteorological parameters with a modified version of the LOWTRAN 7 computer code to remove the effects of the atmosphere and the sensor responsivity. Using long wave (LW) infrared ship signatures obtained with an airborne thermal imager (AGEMA 900) at different ranges, the algorithm's ability to remove the effects of the atmosphere is examined. The sensitivity of the inferred zero-range temperatures to uncertainties in the ship's surface emissivity and the resulting effects on system performance predictions are discussed.

2. INTRODUCTION

The Propagation Division (Code 88) of the Naval Command, Control and Ocean Surveillance Center, Research, Development, Test and Evaluation Division (NRaD) is funded by the Office of Naval Research to evaluate an Electro-Optical Tactical Decision Aid (EOTDA). This effort involves (1) determining the zero-range ship temperatures of different types of combatants and their sea backgrounds to evaluate the performance ranges of airborne FLIR systems, and (2) establishing a data base of the thermal characteristics of different types of ships. For this effort, airborne measurements of infrared ship signatures and their sea backgrounds have been obtained using an AGEMA 900 thermal imaging system. An algorithm has been developed using a modified version of the Atmospheric Transmission/Radiance computer code LOWTRAN 7 [1] to retrieve the zero-range temperature. This algorithm uses airborne measurements of the vertical profiles of meteorological parameters with the LOWTRAN 7 code to remove the effects of the atmosphere and sensor responsivity on the measured ship temperature. Also included in the modified computer code is a sea radiance model based on the Zeisse-Cox-Munk [2] wave slope statistics for capillary waves for determining the zero-range background temperatures in a wind-ruffled sea. In this paper, a case study is presented to demonstrate the algorithm's ability to remove the atmospheric effects and to investigate the sensitivity of system performance predictions to uncertainties in the ship's surface emissivity.

3. SHIP TEMPERATURE RETRIEVAL ALGORITHM

The radiance corresponding to the temperature, T_A, measured by the AGEMA at a range R from the ship is given by

$$\frac{1}{\pi}\int_{\nu_1}^{\nu_2} W(T_A,\nu)F(\nu)d\nu = \int_{\nu_1}^{\nu_2} N_p(\nu,R)F(\nu)d\nu + \int_{\nu_1}^{\nu_2} \varepsilon(\nu)\tau(\nu,R)\left(\frac{1}{\pi}\right)W(T_S,\nu)F(\nu)d\nu \quad (1)$$

where T_S is the ship's temperature at R = 0 and W(T,ν) is Planck's spectral radiant emittance at a wave number ν. ε(ν) is the ship's spectral

Paper presented at the AGARD SPP Symposium on "Remote Sensing: A Valuable Source of Information", held in Toulouse, France from 22-25 April 1996, and published in CP-582.

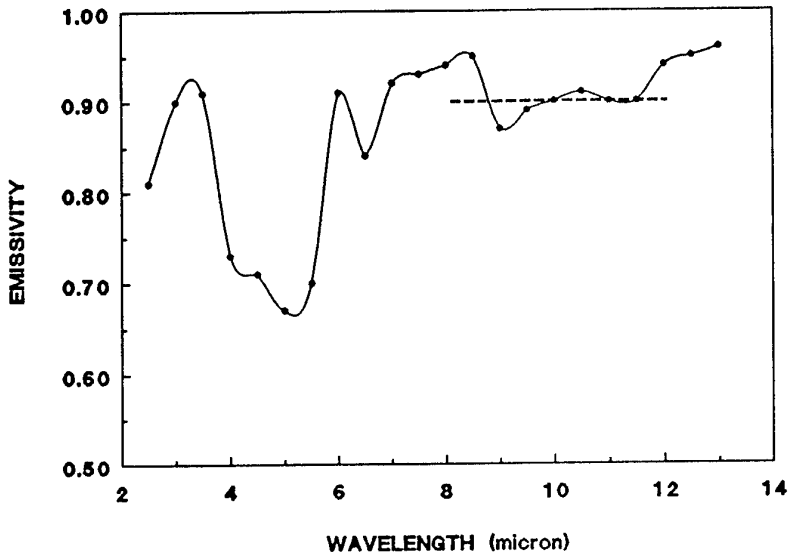


Figure 1. Spectral emissivity of light gray equipment enamel (Spec. MIL-E-15090).

emissivity, $\tau(\nu)$ is the spectral atmospheric transmissivity, $N_P(\nu)$ is the thermal radiance of the atmosphere, and $F(\nu)$ is the system responsivity.

The LOWTRAN 7 code at present allows only the use of a black or graybody emissivity. However, the surface paints of navy ships are selective radiators in the infrared bands. An example of the spectral emissivity for a light gray equipment enamel [3] is shown in Fig. 1. In the short wave (SW) band, the emissivities vary about a value of 0.65 at 5 μm to 0.9 at 3 μm and 6 μm. In order to retrieve the ship temperatures in this band the LOWTRAN 7 code needs to be modified. The LW band emissivity, however, more closely approaches a graybody as indicated by the dashed line in the figure. Assuming the ship's emissivity to be that of a blackbody or graybody for the LW band, then the modified LOWTRAN 7 code calculates directly the quantity inside the brackets of the right-hand side of Eq. 1 when the code input (TBOUND) is taken as T_S. The left-hand side of Eq. 1 is a constant for a fixed value of T_A, and a look-up table has been included in the code to express the radiance as an equivalent temperature. To solve for T_S for a given wavelength band, T(BOUND) is adjusted until the output file yields a temperature equal to T_A. The input requirements to LOWTRAN 7 are (1) vertical profiles of atmospheric pressure, temperature and relative humidity, (2) surface wind speeds (24-hour average and current), (3) Navy Aerosol Model continental influence parameter (ICSTL), (4) the surface visibility, and (5) the ship's surface emissivity.

4. MEASUREMENTS

The ship temperature measurements used in this study were obtained on 12 April 1995 during airborne measurements with the AGEMA 900 system off the southern California coast. The system was mounted in a Piper Navajo aircraft that was flown at low altitudes at different ranges from the ship. Vertical profiles of meteorological parameters (temperature, relative humidity and pressure) were also measured in the vicinity of the ship. The measured profiles of temperature and relative humidity are shown in Fig. 2. Throughout the day, the conditions were characterized by a low-level temperature inversion near 200 m with scattered cumulus clouds. The winds had remained westerly from 280°T for at least 48 hours preceding the measurements. At the time of the measurements, the 24-hour (WHH) and current (WSS) wind speeds were determined from the target ship to be 3.8 m/s and 5.2 m/s, respectively. The sea-surface temperature (15.6°C) was measured from an altitude of 11 m using an Everest (Model 4000ALCS) infrared temperature sensor mounted underneath the aircraft.

The dual-wavelength AGEMA 900 is an internally calibrated thermal imaging system with identical

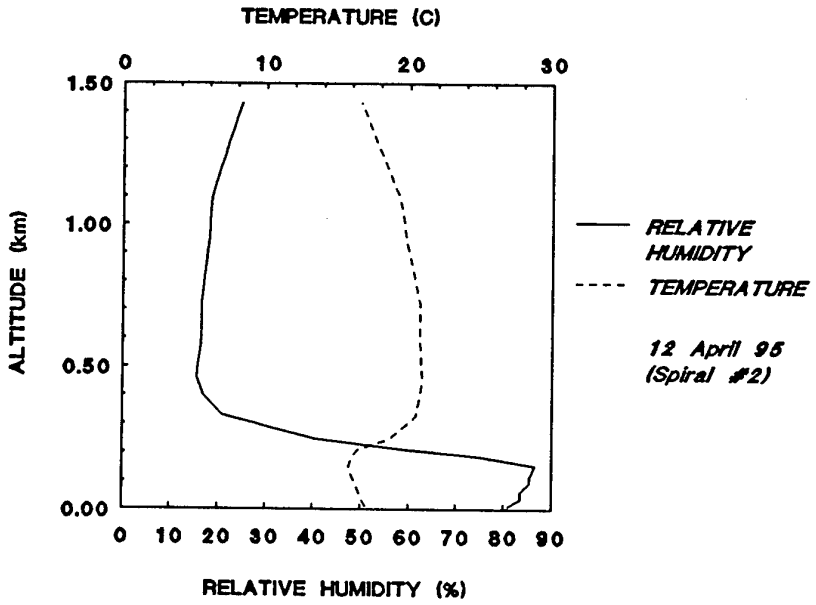


Figure 2. Measured profiles of relative humidity and temperature (12 April 1995).

vertical and horizontal fields-of-view of 5 and 10 degrees, respectively, for each band. The SW [indium antimonide (InSb)] and LW [mercury cadmium telluride (HgCdTe)] scanner detectors are Sterling cooled, and both approximate an ideal photon detector between the lower and upper limits of 6.6 μm and 14.29 μm for the LW scanner, and 1.89 μm and 6.29 μm for the SW scanner. As stated above, because the LOWTRAN 7 code does not accept selective radiator emissivities, only the ship signatures obtained with the LW scanner will be presented in this paper.

The AGEMA 900 data processing software (ERIKA) allows the digitized thermal scene to be displayed on a computer terminal in a format consisting of 136 vertical and 272 horizontal pixel lines. Selected areas in a scene can be outlined and the temperature distribution of pixels enclosed within the area determined. Examples of this display option is shown in Fig. 3. The images are of a surface vessel operating off the southern California coast. The upper signature was obtained from an aircraft altitude of 90 m at a range of 1.75 km (determined from the angular subtense of the known length of the ship). The lower signature shows the midship section of the same ship obtained from an altitude of 12 m at a range of 0.365 km. This range was determined from the known length (11 m) and the angular subtense of the small boat (LCPL) appearing in the lower left-hand corner of the figure. The upper and lower signatures were obtained approximately 3 min apart. To the right of each ship signature are the results of a temperature analysis of the pixels enclosed in the same area (AR01) in each scene. The average ship temperatures (STAT Avg) are 20.92°C and 23.57 °C at the 1.75 km and 0.365 km ranges, respectively.

5. SHIP TEMPERATURE RETRIEVAL

The ability of the algorithm to retrieve the zero-range temperature could best be tested by having *in-situ* thermal couple measurements on board the ship. However, permission to implement these types of measurements on naval combatants is difficult to obtain. Assuming the atmosphere to be horizontally homogeneous, close-up and far-away images (as shown in Fig. 3) provide an alternate method to investigate the removal of the atmospheric effects. This is done by insuring that the retrieved ship temperature is the same at the different ranges. For this purpose, the vertical profiles of meteorological parameters shown in Fig.2 were used with the surface wind speeds listed

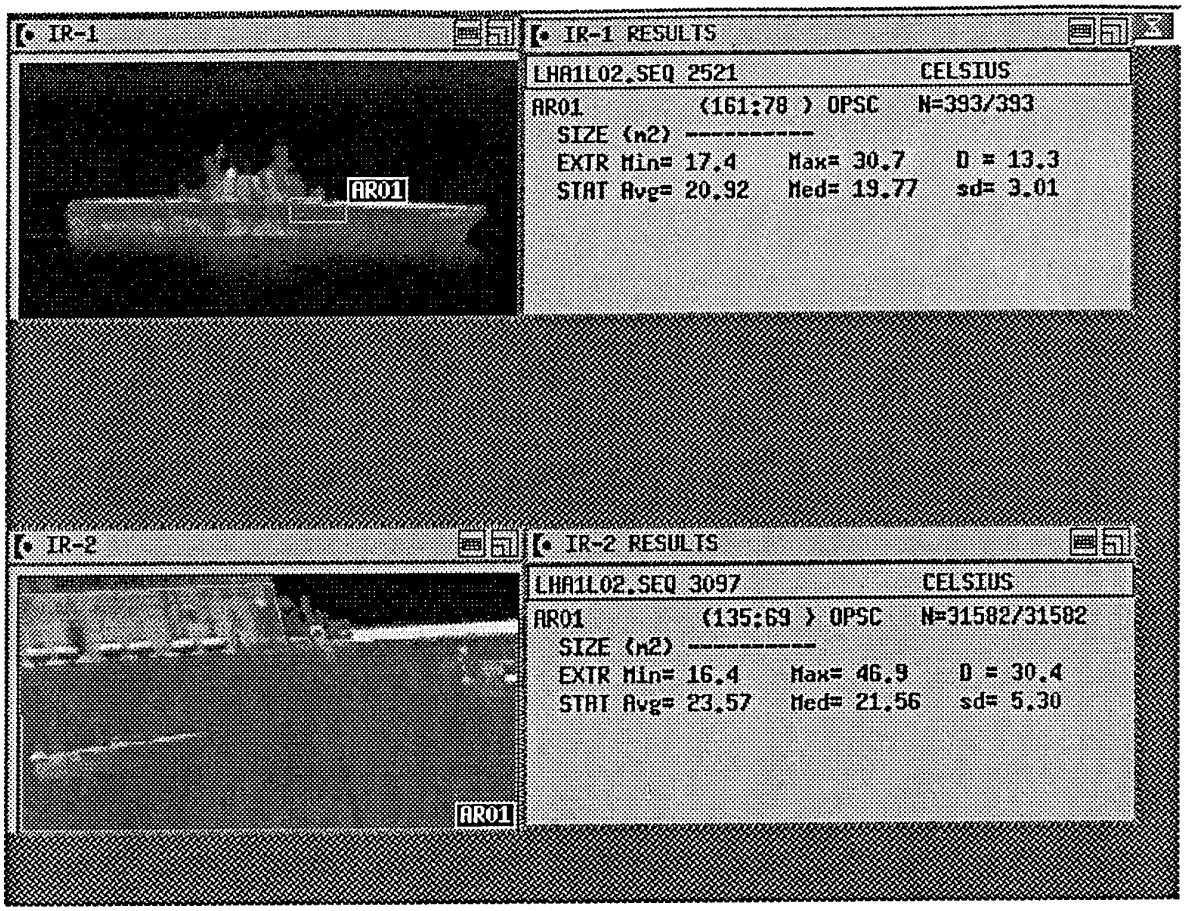


Figure 3. AGEMA display of the starboard side of the ship from an altitude of 90 m at a range of 1.75 km (top), and its midship section from an altitude of 12 m at a range of 0.365 km (bottom). The temperature analysis of the same outlined area (AR01) is shown to the right in each scene.

earlier in the modified LOWTRAN 7 code to determine the average zero-range temperature of the midship section within the enclosed areas (AR01) of Fig. 3. For these measurements, the air mass factor (ICSTL) and surface visibility (VIS) were not available. ICSTL has integer values from 1 (indicating maritime air) to 10 (indicating continental air). For a given combination of air mass factor and visibility the zero-range ship temperature (TBOUND) was adjusted to give the AGEMA measured values at each of the ranges. In Fig. 4, the zero-range midship temperatures of the area AR01 are plotted versus visibility for different values of ICSTL. For ICSTL = 1 (indicated by the dashed line connecting the open circles and dots in the figure), the zero-range temperatures determined at both ranges coincide for a visibility of 19 km. It is interesting to note that the zero-range temperatures for the shorter range (0.365 km) show little dependence on the visibilities greater than 19 km, indicating a negligible affect of aerosols on the atmospheric transmittance. Different combinations of ICSTL and VIS will give identical results as indicated in the figure for ICSTL values of 5 and 10. In these cases, only the cross-over points of the zero-ranges calculated versus visibility are shown. For this set of data, it is only known that the surface winds were westerly for at least 48 hours preceding the measurements and the air mass was most likely of maritime origin, i.e., ICSTL = 1. The surface visibility reported by the ship was 10 nmiles, making the lower visibilities required for the other two values of ICSTL unrealistic.

These calculations were performed assuming the ship's emissivity to be that of graybody equal to 0.9. Similar plots can be obtained for different graybody values as shown in Fig 5. Here, the midship zero-range temperatures determined from

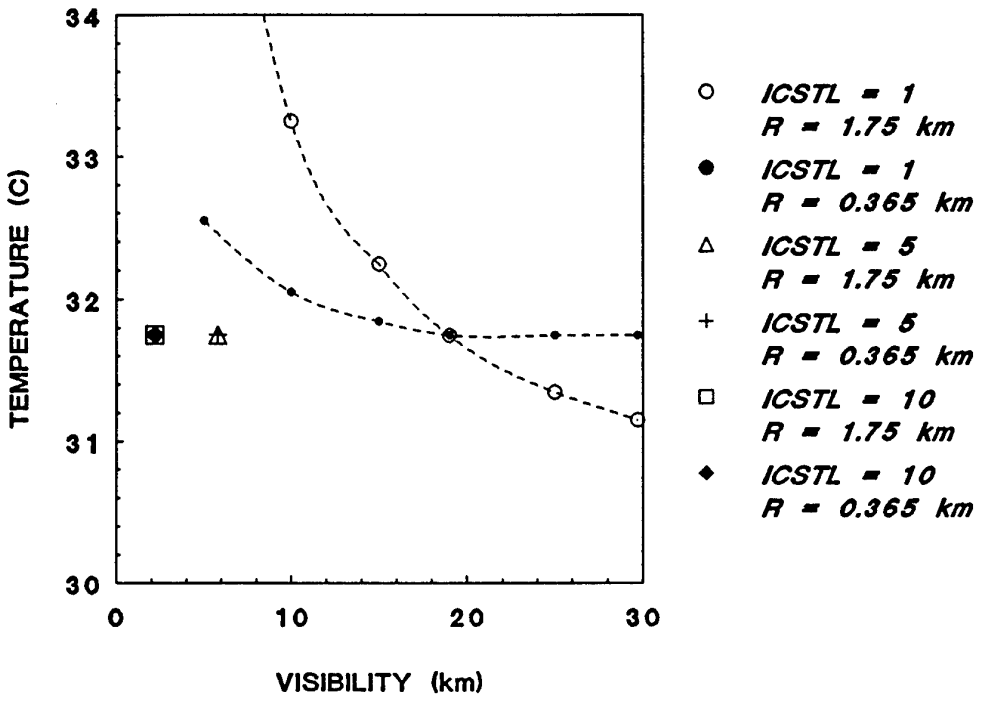


Figure 4. Zero-range midship temperatures versus visibility reguired to give AGEMA measurements.

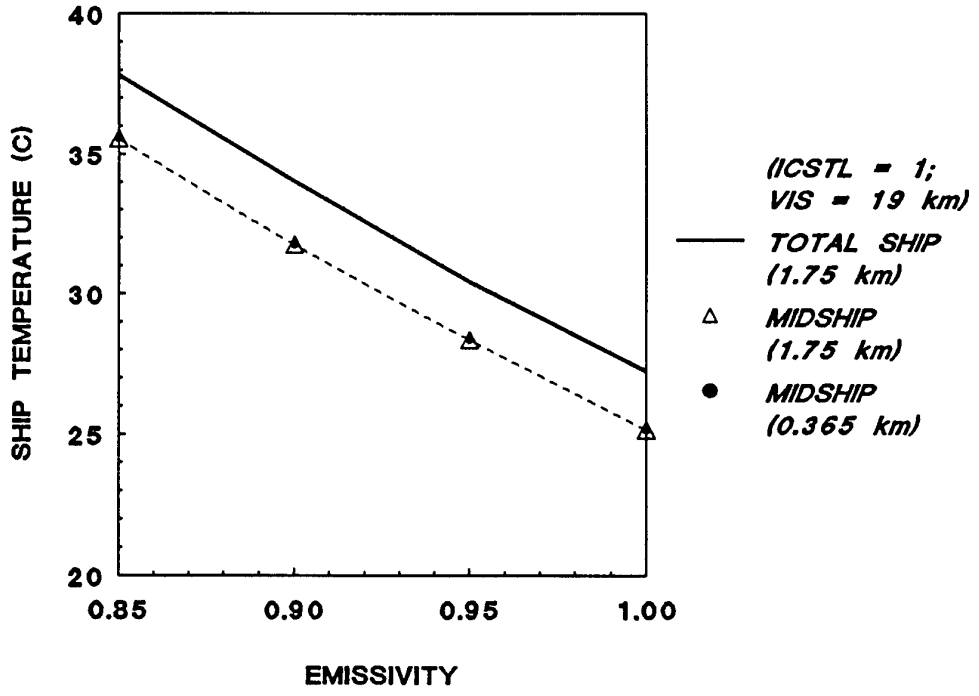


Figure 5. Zero-range temperatures versus emissivity required to give AGEMA measurements.

the two different ranges coincide for emissivity values from 0.85 to 1.0. Also shown in the figure are the zero-range temperatures for the total ship determined from the 1.75 km signature. The significant feature of the graph is that a 15% uncertainty in the ship's emissivity results in approximately a 10 °C uncertainty in the retrieved zero-range ship temperature.

6. EMISSIVITY EFFECTS ON SYSTEM PERFORMANCE PREDICTIONS

The ultimate test of the accuracy of the retrieved zero-range temperatures is how well they can be used to predict systems performance range. As performance range observations using operational systems were not available, the AGEMA system was flown away from the ship on a slant path until the ship was not discernible from the sea background. This range (accurate to 160 m) was determined from the ship's transmitted TACAN (Tactical Air Navigation) beacon received onboard the aircraft. In Fig. 6, the upper thermal image of the port side of the ship was obtained from an altitude of 99 m at a range of 1.77 km. The lower image in the figure was obtained from an altitude of 211 m at a range of 15 km. This image was the last recorded before the ship could not be detected on the AGEMA video display by the operator when the ship was at a range of 18.6 km.

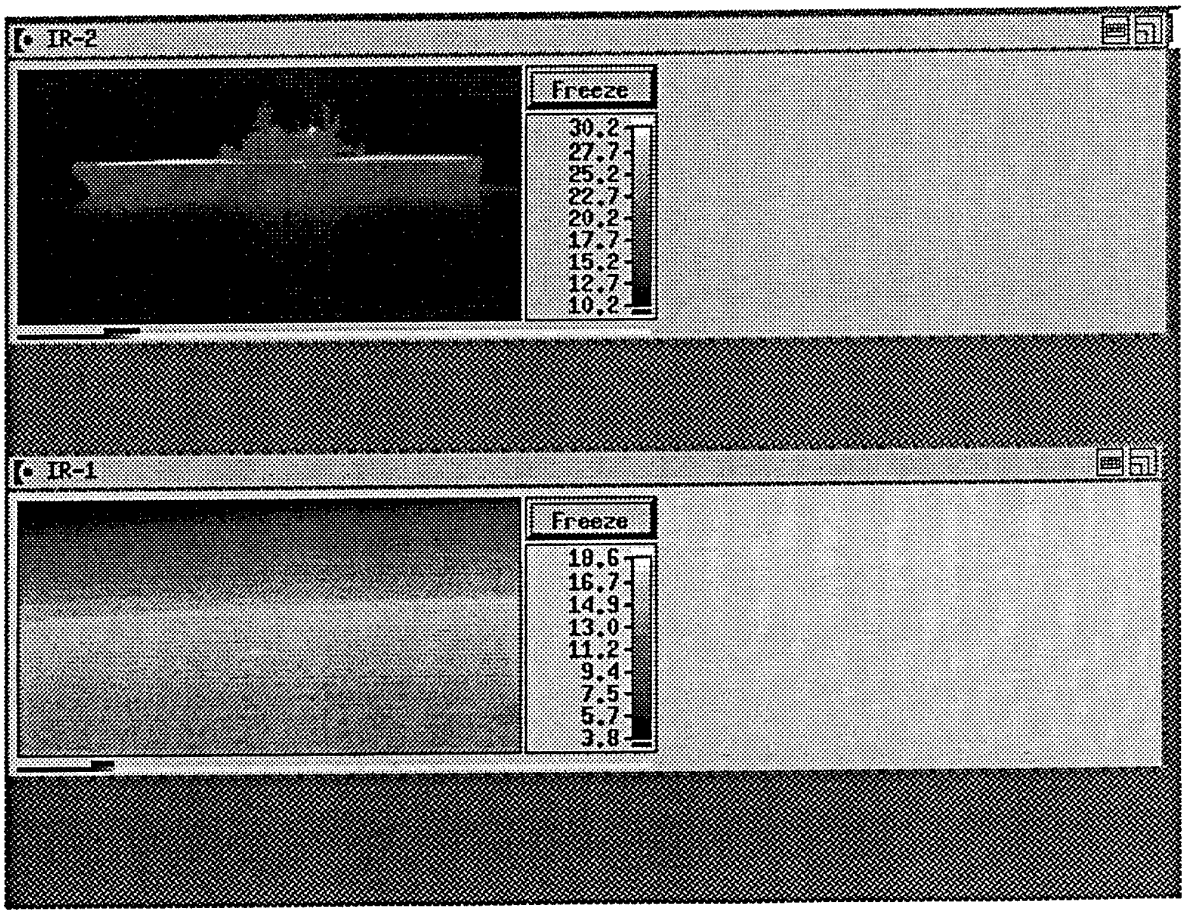


Figure 6. AGEMA display showing the port side of the ship from an altitude of 99 m at a range of 1.77 km (top) and from an altitude of 211 m at a range of 15 km (bottom).

In order to determine the accuracy by which the retrieved zero-range ship temperature can predict systems performance, a simple detection task is chosen here. For this purpose, the spatial frequency of the targeted ship has to be determined. The spatial frequency (ν) of a target in terms of its angular subtense (Θ) at a range R is defined as

$$\nu = (2\Theta)^{-1} = R(2000D_c)^{-1} \text{ (cycles/mrad)} \quad (2)$$

where D_c is the ship's *critical dimension* in meters. D_c is related to the ship's projected area (A_p) viewed by the sensor by

$$D_c = \sqrt{A_p} \quad (3)$$

The projected area of the ship is determined assuming the ship to be represented by an equivalent parallelepiped given by

$$A_p = lw\sin\theta\cos\phi + hw\cos\theta\cos\phi + hl\cos\theta\sin\phi \quad (4)$$

where h is the unobscured target height, w is the target width, l is the target length, θ is the viewing elevation angle, and ϕ is the viewing azimuth angle. The AGEMA's *minimum detectable temperature difference* (MDTD) response versus spatial frequency was determined by laboratory measurements using bar targets of known temperatures differences. Regression analysis of the measured data yielded the following relationship for the system MDTD,

$$\ln(MDTD) = -3.83 + 6.08\sqrt{\nu} \quad (5)$$

For a particular ship's configuration, substitution of Eq. 2 into Eq. 5 allows the MDTD as a function of range to be determined. The range at which the atmosphere degrades the target-background temperature difference down to the system's MDTD is the *maximum detectable range* (MDR), i.e.,

$$MDTD = (T_s - T_{bkg})\tau(MDR) \quad (6)$$

To calculate the MDR for the ship in Fig. 6, the sea background temperature, T_{bkg}, was determined using the sea radiance model in the modified LOWTRAN 7 code which incorporates the Zeisse-Cox-Munk [2] wave slope probability distribution to determine the contributions to the total sea radiance from thermal emissions and sky reflections from the individual wave facets at zero-range. For the scenes in Fig. 6, a value of T_{bkg}=16.3 °C was calculated using the measured meteorological profiles and ICSTL=1 and VIS=19 km. The reliability of the LOWTRAN code to extrapolate the zero-range sea backgrounds to the AGEMA range (for the moderate surface winds encountered here) is evidenced in Fig.7. In this figure, the average sea background temperatures measured from different altitudes (18 m to 120 m) as the plane circled the ship are compared with the modeled temperatures at the AGEMA range which were extrapolated from the zero-range. The agreement is better than 1 °C in most instances.

The average zero-range ship temperatures for different graybody emissivities were 29.2°C (ε=0.85), 25.5°C (ε=0.9), 22.1°C (ε=0.95) and 19.0°C (ε=1.0). To calculate the AGEMA performance range, these temperatures were used with the average zero-range sea temperature to determine the zero-range temperature difference for each emissivity. The ship and background temperature differences were then degraded by the atmospheric transmittance (determined using ICSTL=1 and VIS = 19 km) for each of the emissivities. The temperature differences are plotted versus range and compared with the MDTD of the AGEMA in Fig. 8. The observed MDR (□) at 18.6 km is also shown.

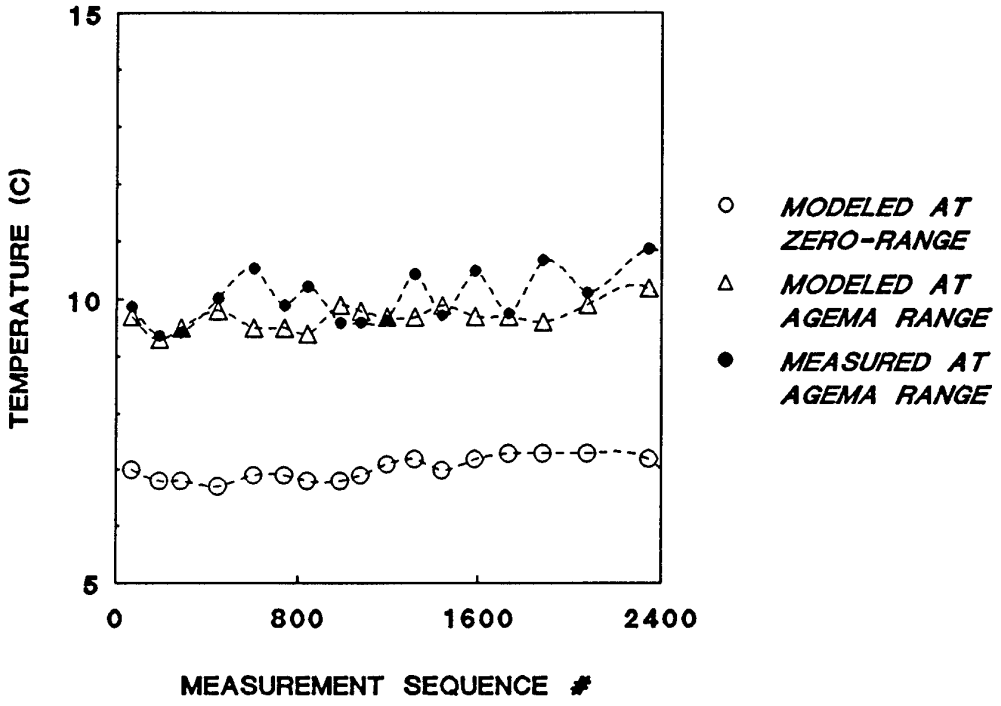


Figure 7. Comparison of measured and modeled sea background temperatures.

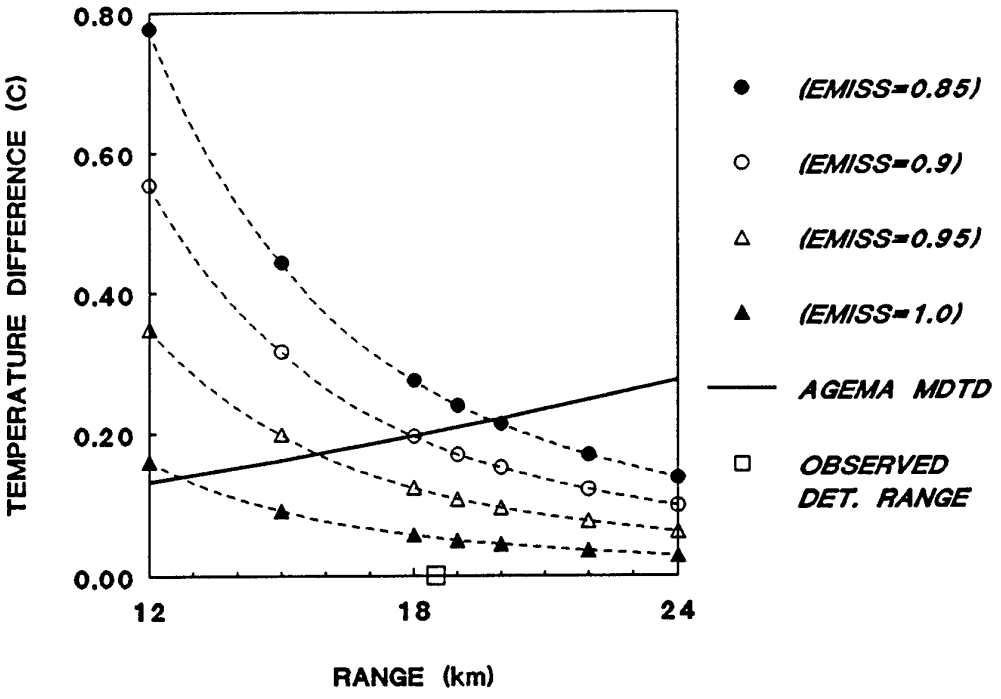


Figure 8. Modeled ship and background temperature differences and AGEMA system MDTD versus range.

7. CONCLUSIONS

In Fig. 8, there is good agreement between the predicted and observed detection ranges for the AGEMA long wave infrared system for emissivities between 0.85 and 0.9. However, a 15% uncertainty in emissivity (i.e., $0.85 \leq \varepsilon \leq 1.0$) caused the performance range predictions to differ from the observed value by approximately +11% and -32%. Larger differences can be expected if the aerosol parameters (ICSTL and VIS) are not accurately known. Future measurements should include these parameters. In their absence (as demonstrated in this case study), ship signatures at close and distant ranges should be obtained to insure the ship temperatures retrieved from different ranges are the same and result in good agreement between predicted and observed performance ranges.

The LOWTRAN 7 code needs to be modified to accept selective radiators, in particular, those for the SW infrared band. For this wavelength band, however, the effects of solar reflections on the retrieved ship temperatures will have to be included in the algorithm. Future efforts should also address the effects of clouds.

ACKNOWLEDGMENTS

This work was supported by the Naval Command, Control and Ocean Surveillance Center, Research, Development, Test and Evaluation Division 88, under Contract No. N66001-94-0064, Delivery Order 004. Dr. Juergen H. Richter was responsible for providing funding for the project under the Office of Naval Research Exploratory Development Program.

REFERENCES

1. Kneizys, F. X., E. P. Shettle, L. W. Abreu, J. J. H. Chetwynd, Jr., G. P. Anderson, W. O. Gallery, J. E. A. Selby and S. A. Clough, "Users Guide to LOWTRAN 7," Air Force Geophysic Laboratory Technical Report No. 88-0177 (1988).

2. Zeisse, C. R., "Radiance of the ocean horizon," J. Opt. Soc. Am. A, Vol. 12, No. 9, September 1995, pp 2022-2030.

3. Suits, G. H., "Natural Sources". In *The Infrared Handbook*, 4th Printing, W. L. Wolfe and G. J. Zissis (Eds.), Environmental Research Institute of Michigan, Ann Arbor, Michigan, Ch. 3, 1-154, 1993.

Paper 36

N. J. Farsaris (GR)

What about combined IR-UV imaging?

Author's reply:

I cannot answer. We have no experience nor made any calculations at UV wavelengths.

Some results from AGARD Study AAS-42: Potential Use of Commercial Satellite Systems for Military Applications.

Magdeleine Dinguirard, ONERA/CERT, 31055 Toulouse - France
Terje Wahl, FFI, 2007 Kjeller - Norway
James Murray McDonald, ONR, Arlington VA 22217-5660 - USA
Gerard Carrin, Matra Marconi Space, 31077 Toulouse - France
David Levy, CNES, 31055 Toulouse - France
Ull Erich, Dornier GMBH, 88039 Friedrichshafen - Germany
Karl Heins Zeller, DLR, 82230 Wessling - Germany
Xavier Briottet, ONERA/CERT, 31055 Toulouse - France

1. ABSTRACT

The commercialization of space is rapidly making satellite communication, navigation and remote sensing accessible to many users, both civilian and military. In 1994 it was decided that there was a need for an AGARD study of the consequences to NATO of this new trend. The AGARD AAS-42 Study Team started its work in January 1995, with government and/or industry representatives from eight countries. The study covered satellite communication, satellite navigation and satellite remote sensing, focusing both on possible NATO use and the threats of adversarial use of commercial satellites against NATO. Final presentation of the study was held in November 1995. As part of the study, data on most known or planned non-military remote sensing satellites have been collected and analysed, and some of the results from this analysis will be reported here. Among the new elements that have been considered, are commercial satellites with resolution better than 10 meters, and near real time processing and delivery of SAR images from radar satellites. It is clear that well-organized use of non-military remote sensing satellites among NATO countries can be a valuable supplement to current NATO surveillance data sources. Also, in certain kinds of crises or conflicts, fast access to commercial remote sensing images can have a high value for a potential NATO adversary.

2. INTRODUCTION

An information revolution is in progress, resulting from the capabilities of a growing number of satellite-based commercial systems to both acquire information of various kinds from any point on the earth and to rapidly transmit that information to any other place on the earth. By commercial, this study refers to "privately or nationally owned space/ground assets that can provide service in exchange of economic value."

A study team (AAS-42) was formed to evaluate the potential military applications of commercial satellite services, both by NATO and its allies and by possible adversaries. In this frame, three kinds of service have been identified: communications, remote sensing and navigation.

This paper will deal only with the remote sensing aspect of the AAS-42 study.

By 1998, over thirty commercial remote sensing satellite systems are planned to be in operation. While each system will include one or more sensors useful for one or more applications, the systems can be broadly grouped into two categories: **Land Remote Sensing Systems (LRSS)** designed to look at natural and man-made features at the land surface: and **environmental sensing systems**, which are further sub-divided into ocean/ice-observing systems and meteorological systems. Data from these systems can be displayed in various ways-- since the data from land-observing and ocean/ice-observing systems are often (but not always) presented in the form of two-dimensional "pictures" or "images" of the land or ocean surface, they are frequently referred to as imaging systems. Purely meteorological data (atmospheric sounding) are three-dimensional in nature and are usually of low horizontal resolution. Accordingly, most meteorological systems may be correctly described as "non-imaging" though the best known METEOSAT and AVHRR are nevertheless imaging sensors.

The instruments aboard imaging satellites sense electromagnetic radiation, and they are usually categorized according to the part of the electromagnetic spectrum (that is, the frequency/wavelength of the radiation) to which they are sensitive. There are three basic categories of imagers:

* **Optical** imaging devices are sensitive to "visible light" just as the human eye is. Optical sensors do not acquire images at night, or through smoke or clouds.
* **Infrared** (IR) imagers are sensitive to radiation of slightly longer wavelength (and slightly higher frequency) than visible light. IR can be subdivisded in : near-IR or short-wavelength-IR (NIR, SWIR) and middle- or mid-wave-IR (MIR, MWIR) which are often used as a complement to visible sensors; long-wavelength-IR is also known as thermal IR (LWIR, TIR), and images a scene by the temperature differences present. TIR sensors which are looking at the emission of the landsape and not at its reflection, have some ability to image through smoke and clouds and can image at night.
* While the two formers were "passive" imagers and needed external light source, **Radar** systems are "active" and illuminate the scene they are sensing. They usually operate as "side-looking radars" (SLRs) whose data are processed by a computer to attain spatial resolution which would otherwise require much larger antennas. SLRs whose data are processed in this way are also known as "synthetic aperture radars" (SARs). Radars can image at night and through smoke and cloud.

Planned commercial imaging systems, both land-observing and ocean/ice-observing, will make accessible to anyone in the world a level of information previously available only to military users of the major industrial nations : frequent, high spatial resolution (1m or greater) observations of several different kinds at any desired location on the globe. Images have no intrinsic value--their value arises from the kinds of information which can be extracted from them. Since several NATO member nations (and some others) have infrastructures to extract military intelligence from high-resolution aerial and satellite imagery, the commercial marketing of high-quality satellite imagery is for NATO primarily a vehicle to supplement intelligence it can already obtain.

Paper presented at the AGARD SPP Symposium on "Remote Sensing: A Valuable Source of Information", held in Toulouse, France from 22-25 April 1996, and published in CP-582.

However, it also provides many potential adversaries with a source of intelligence about NATO nations and forces which was previously unavailable to them.

Because the constant motion of each part ultimately affects the entire atmosphere, meteorological data are useful only if they are taken frequently, cover a large portion of the earth and are incorporated into supercomputer-based forecast schemes. Even national governents, including those of the NATO member nations, lack the resources to accomplish the entire task and this has led to international agreements for sharing of satellite and non-satellite meteorological data among all the nations of the world. Accordingly, all these data are and should continue to be available to NATO and its member nations, and the commercial meteorological satellite systems offer no new opportunities or threats to NATO.

3. LAND REMOTE SENSING SATELLITE SYSTEM CAPABILITIES

Quality (spatial resolution, good image quality, good coverage, etc...) and timeliness are the key considerations for military use of remotely-sensed data. During peace time, quality is the most important factor but, during crisis or war, all surveillance information is useful, and rapid delivery of the data becomes the pre-eminent consideration. In wartime even poor resolution images can be useful, as has been shown during Desert Storm.

3.1. Passive imaging systems

Several passive imaging systems are already on orbit or on under development. They are essentially electro optical cameras operating in the Visible and Near IR part of the spectrum. Their main application is terrain mapping (some have the stereo capability allowing to derive digital terrain elevation models) and if they have spectral bands, resource evaluation (crop and mining identification). They can also be used to assess man or climate damages (fires, floods, explosions). Because of their spectral range, they are only usuable by daylight and under good weather conditions (clear sky). Figure 1 illustrates the different LRSS, existing and being shortly planned in the world, and their evolution with time; on this figure, the relative width of the systems are proportional to their higher available ground resolution.

A more complete descriptions of these systems can be found in vol2 of the AAS-42 Study (classified) or in reference [1]

One can see on Figure 1 the good service continuity of the LANDSAT and SPOT systems which also have both a world wide distribution network allowing a quickest sale of the images. For more rapid and more flexible operation, mobile SPOT ground receiving and processing stations like Eagle Vision may be available.

Figure 1 and table 1 show also the improved performances of the announced US new commercial systems. Several US private firms are actively engaged in developing systems that are expected to provide high-quality high-resolution images in the near future. If successful, these efforts will make inexpensive, high-resolution imagery available world-wide.

One must take into consideration that there are inherent risks involved in centring a military strategy around commercial systems, specially the new ones with no guaranty of service continuity. The planned imagery may not be available as planned due to technical difficulties, cancellation of the program, or policy issues.

Systems	spectral bands	Swath Width	Ground Res.	company
Early Bird	Pan : 0.45-0.8μm MS : 0.50-0.59μm 0.61-0.68μm 0.79- .89μm	6km 30Km	3m 15 m	**Earth Watch** (Ball, Hitachi, CTA, Nuova..)
Quick Bird	Pan : 0.45-0.9μm MS : 0.45-0.52μm 0.53-0.59μm 0.63-0.69μm 0.77- .90μm	36km 36Km	1m 4 m	
Orb -View	Pan : 0.45-0.9μm MS : 0.45-0.52μm 0.52-0.60μm 0.63-0.69μm 0.76- .90μm	8km 8km	1m 8 m	**Orbital Sciences** (Orbital, EIRAD, Fairchild, MDA ..)
SIS1	Pan : 0.45-0.9μm MS : 0.45-0.52μm 0.52-0.60μm 0.63-0.69μm 0.76- .90μm	11km 11km	1m 4 m	**Space Imaging Systems** (Locheed, ESystems, Mitsubishi)

Table .2: New commercial land remote sensing satellites characteristics [2].

3.2. Radar systems

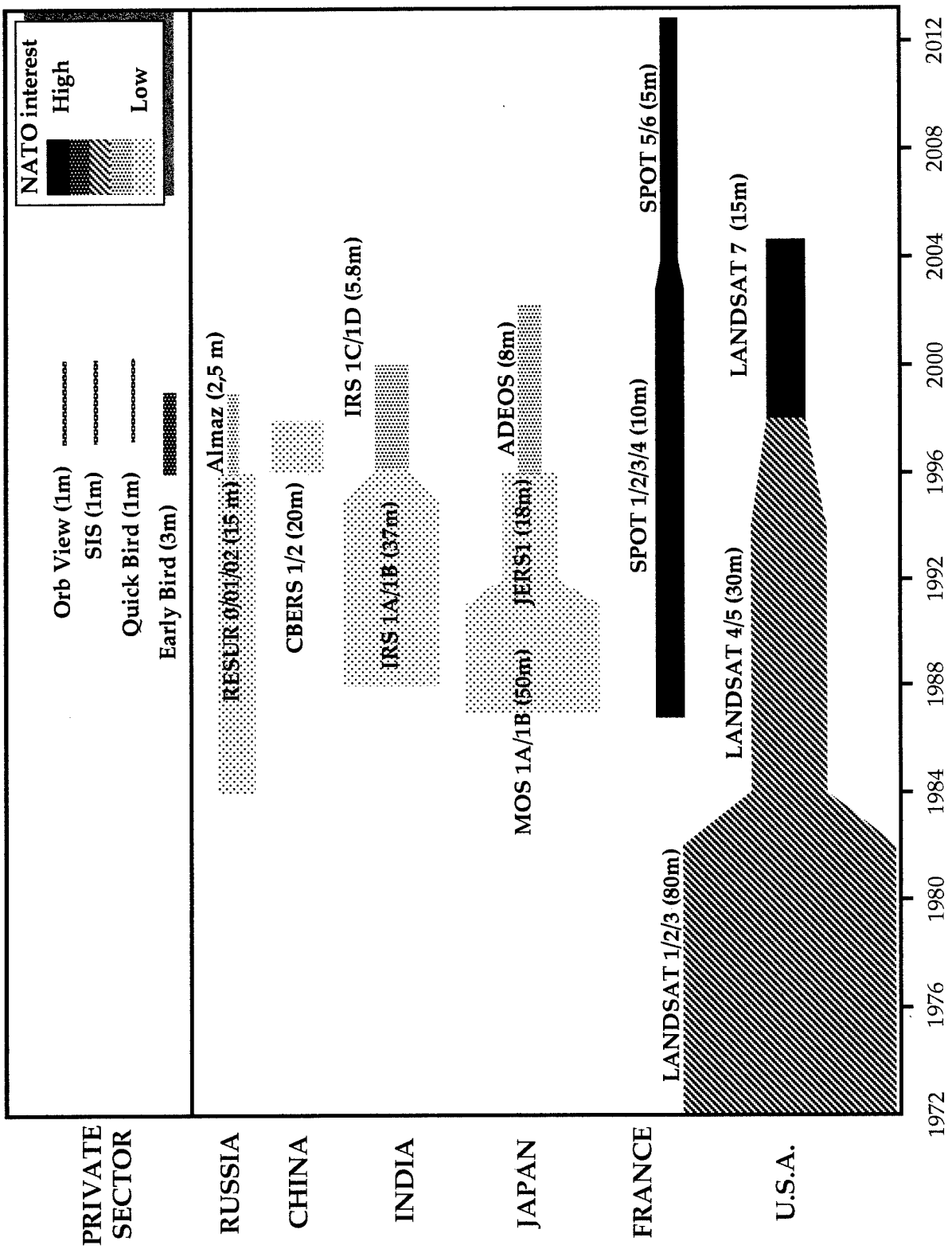
Synthetic Aperture Radar systems have the capability to provide medium-resolution imagery without being affected by adverse weather or night conditions. This indicates potential military application of available imagery, but SAR technology is fairly immature, and potential applications are still being explored. The implementation of SAR to produce imagery over land has several inherent characteristics which makes obtaining SAR imagery over land more difficult that optical systems. Assuming these inherent difficulties can be overcome, there are several applications where SAR imagery can be of use in a military context. Some of these applications are:

- providing cueing for optical sensors;
- detection of targets that cannot be detected by optical sensors;
- collection of radar backscatter properties to enable tactical weapons systems which utilise radar for guidance to perform more effectively;
- mapping of terrain using SAR Interferometry;
- detection of changes in the target area which can provide data for applications such as battle damage assessment or detection of stealth targets.

Several countries have developed SAR imaging capabilities which may be of potential use for NATOmilitary applications. Table 3 describes their main characteristics.

When considering civilian radar imaging systems for military use, it must be noted that the resolution of these systems are typically too poor to provide more than limited capability in military applications. Existing systems can be utilised to supplement other intelligence data, and radar satellite technology is advancing in such a way that it can be expected that future satellite systems may be of higher military utility. The most promising systems are the planned RADARSAT-3 satellite.

Figure 1 : Existing and shortly planned Optical Land Remote Sensing Systems.



3.3. Military use of land civilian systems

Table 2 summarises the applicability of some commercial systems to military applications ranked with + or - accorting they are of interest (+) of or poor interest (-) :

Optical Systems	LANDSAT SPOT	ADEOS IRS ALMAZ	SIS QUICK-BIRD EYEGLASS
Resolution	+	+	++
Weather+day/ night	-	-	-
Radiometric quality.	+	+	?
World wide access	++	+/-	+
Revisit time	+	+	+
Guaranteeof service	+	-	-
Denial	-	-	-
Radar Systems	ERS, ENVISAT	ALMAZ	RADARSAT
Resolution	-	+/-	+
Weather+day/ night	+	+	+
Radiometric quality	+	?	+
World wide access	+	+	+
Revisit time	+	?	+
Guarantee of service	+	-	+
Denial	-	-	-

Table 2 : Military interest of conmmercial land remote sensing satellite systems.

No one commercial system can satisfy all of the military needs but each one is able to meet some requirements and thus can be considered as a possible supplement to already existing surveillance sources. Besides that, commercial systems are likely to be more useful to NATO potential adversaries than to NATO itself and they are vulnerable.

A good dual use of commercial satellite must go through the following main recommendations :

- Commercial remote sensing satellite systems should be used as a complement to national military dedicated remote sensing systems.
- A strategy for the acquisition and fusion of data from operational commercial systems should be analysed. This strategy should repose on good performance, good service continuity systems.
- Methods to deny adversarial access to commercial remote sensing satellite systems should be developped. Countermeasures considering deception and denial must be studied.
- The design of newcoming systems should incorporate encryption along with security and robustness against jamming.

4. OCEAN/AIR REMOTE SENSING SATELLITE SYSTEM CAPABILITIES

Weather conditions have strongly influenced the outcome of warfare throughout the centuries, and even some of today's most advanced weapons can only be used in good weather. Modern meteorology was born from research work conducted after British and French naval units reported severe problems with unexpected storms during the Crimean War nearly 150 years ago. Today, numerous meteorological satellites, both military and civilian, are orbiting the Earth in geostationary or polar orbit, and active microwave satellites are steadily revealing new information about the ocean.

4.1. Air

Most imaging weather satellites provide images with resolution 1 km or poorer, and are usually not considered for direct observation of military forces. In certain cases, ships may leave atmospheric trails in AVHRR images, but generally these satellites are considered for their primary meteorological role only, not for intelligence or targeting.

One single observation of the cloud pattern is of limited value because of the dynamical nature of the atmosphere. Therefore, weather satellites are primarely used together with (and partly as input to) numerical forecast models. This means that the part with the best models can also make the most efficient use of these satellites. An indication of this is that civilian weather satellites were not switched off during the Gulf War.

Thermal infrared imaging is a much favoured method for detection of rocket launches. One has raised the question whether geostationary weather satellites can contribute in ballistic missile warning. A closer look at the parameters shows that the new GOES and METEOSAT systems could have some rudimentary capability in this respect, but that the scan rate and processing time are not adequate for such use.

4.2. Ocean

The launch of ERS-1 in 1991 started the era of continuous presence of civilian radar satellites in orbit. At the time of writing, four such satellites (ERS-1, ERS-2, JERS-1, RADARSAT-1) are in operation. However, differences in the radar parameters among these satellites make them suitable for quite different tasks. Table 3 shows the basic radar parameters for these satellites, including ESA's planned next satellite ENVISAT (to be launched in 1999).

Satellite	**Freq**	**Pol**	**Swath Width**	**Res**	**Incidence angle**
ERS-1, 2	C	VV	100 km	25 m	23°
JERS-1	L	HH	75 km	18 m	35°
RADARSAT	C	HH	60-500 km	9-100 m	20°-50°
ENVISAT	C	Many	60-405 km	30-150 m	15°-45°

Table .3: Existing or planned commercial radar satellites main characteristics.

The ERS-1 and ERS-2 satellites have fixed SAR antennas looking quite steep. These are well suited for oceanographic use, e g wave height or current shears. The satellites are already being used operationally for pollution monitoring [3]. The ERS satellites do have a certain ship detection capability (see Figure 2) but small ships may easily be lost in the wind sea clutter because of the steep incidence angle [4], [5].

JERS-1 has experienced technical problems and is operated on reduced power. Still, because of the (L, HH, 35°) sensor parameters, this SAR has a much higher detection capability against hard targets (ships) than the technologically more advanced ERS satellites).

RADARSAT will clearly be the SAR satellite technically best suited for detection of ships. ESA plans to introduce a mixed polarization mode on ENVISAT, where both VV and VH images can be obtained simultaneously. This is very interesting for detection of both a ship (using VH) and its wake (using VV). However, this can only be done in swaths of width 100 km or less.



Figure 2 : ERS-1 low resolution SAR image showing warships in a Norwegian fjord during a NATO exercise. Such images can be easily transferred on the Internet.. (c) ESA/TSS.

It follows from the above that commercial radar satellites can have a considerable ship detection capability. Coverage will still be poor, though (except at very high latitudes), and this will reduce the value for NATO use, as one would anyhow have to rely on other observation assets. For a potential adversary, however, access in near real time to one single SAR observation may be of high tactical value. It could be used for breaking a blocade or, in extreme cases, for firing modern anti-ship missiles against NATO ships. Equipment for fast processing of radar satellite images is commercially available today, and mobile receiving stations are already a reality.

5. CONCLUSIONS

Commercial satellite systems can clearly be of value to NATO as a complement to airborne sensors and national military remote sensing systems.Our discussion of non-military remote sensing satellites have revealed some interesting differences:

- Meteorological satellites and oceanographic observations from environmental satellites are more useful to the technologically most advanced part in a conflict.

- Near real time use of commercial satellites for detection of adversarial military units (on ground or at sea) is likely to be more valuable to the weaker part in a conflict, especially if he has been deprived use of other major observation assets (aircraft, radar stations).

It follows that nations and alliances are likely to consider methods for denying the adversary fast access to certain commercial satellite images during conflict.

REFERENCES:

[1] H.J Kramer : "Observation of the Earth and its Environment - Survey of Missions and Sensors" 2nd edition Springer-Verlag 1994.

[2] L.W. Fritz : "The Era of Commercial Earth Observation Satellites" PE&RS Jan 96

[3] Wahl T, Skøelv Å, Pedersen J P, Seljelv L G, Andersen J H, Follum O A, Anderssen T, Strøm G D, Bern T I, Espedal H, Hamnes H, Solberg R: Radar satellites: A new tool for pollution monitoring in coastal waters. To appear in Coastal Management, 1996.

[4] Eldhuset K: An Automatic Ship and Ship Wake Detection System for Spaceborne SAR Images in Coastal Waters. To appear in IEEE Tr Geoscience and Remote Sensing, 1996.

[5] Wahl T: Are radar satellites cost-effective for maritime surveillance? Paper No IAAA-95-IAA.1.3.04, 46th IAF Congress, Oslo, October 1995.

Paper 37

J-P. Salmon (BE)

Have you been able in the frame of the study, to analyse if NATO, as an organization, can reach MOU's with the NATO nations owning civilian satellite systems about the access conditions to raw image data in case of crisis?

Author's reply:

It was not in the scope of this study to look at policy aspects but only at scientific and technical ones. But we recommended that NATO should work out contingency agreements with the most promising systems.

F. Lanzl (GE)

1) What are the criteria you use for definition of NATO interests?

2) In particular, why did you rate the Indian IRS-Satellites as being of only low NATO interest?

Author's reply:

1) A mixture of ground resolution, revisit time, world wide coverage, guarantee of access.

2) Because India is not a NATO nation and we may have access to the data non guaranteed in case of crisis, but of course the satellite has very good capabilities.

Paper 37

R. B. Gomez (US)

You presented an excellent review of current and future data gathering remote sensing systems. An important problem is how do we go from data (observation) to information for decision making. Did you look at the status of our current and future capabilities in going from observation to information? Did you review the current software and tools available to help us deal with this important problem.

Author's reply:

AAS-42 did not study image analysis tools or methods as such. Also, within NATO intelligence is typically done nationally, and the nations develop their own tools. If NATO acquires a mobile receiving station, this question will have to be addressed, possibly a subject of another AGARD study.

J. Shi (US)

Is 1 m resolution approaching practical sufficiency of military use?

Since much remote sensing high technology is commercialised and thus available to potential adversaries, where should NATO put its efforts to maintain its technological superiority?

Author's reply:

1 m is not a physical limit so NATO can do better!

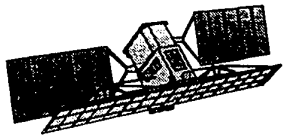
C. Wash (US)

Please comment on the ability of civilian satellite systems to use high resolution infra-red techniques to monitor missile launches.

Author's reply:

We have looked into the problem. For this application one would like to have continued presence, i.e. geostationary orbit. The new generation of sensors on METEOSAT and GOES have some rudimentary capability in terms of the radiometry, but the imaging rate is not good enough. This dual use function will have to be built into the design of future geostationary weather satellites.

RADARSAT: A Dual Use Earth Observation System



Maj M. McKean
RADARSAT Program Systems Engineer
Canadian Forces Detachment
Canadian Space Agency
6767, route de L'Aéroport
St-Hubert, Quebec, Canada J3Y 8Y9

M. Rey
Radar Data Exploitation Scientist
Defence Research Establishment Ottawa
Department of National Defence
3701 Carling Avenue
Ottawa, Canada K1A 0Z4

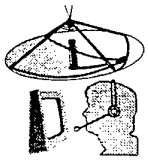


H. Edel
Remote Sensing Coordinator
Fisheries and Oceans Science Directorate
Department of Fisheries and Oceans
200 Kent Street
Ottawa, Canada K1A 0E6



C. Bjerkelund
Oceans Applications
Canada Centre for Remote Sensing
Natural Resources Canada
588 Booth Street
Ottawa, Canada K1A 0Y7





LCol J.L.P. Dionne
Space Liaison and Program Support Officer
Commanding Officer Canadian Forces Detachment
Canadian Space Agency
6767, route de L'Aéroport
St-Hubert, Quebec, Canada J3Y 8Y9

D. Nazarenko
RADARSAT International Inc.
Manager Application Solutions
3851 Shell Road, Suite 200
Richmond, British Columbia
Canada V6X 2W2

1. SUMMARY

Space is emerging as an important component of the global security environment, and the need for all-weather/day-and-night space-derived information is expected to grow in the future. In view of recent events, and anticipated technological developments, there is a growing recognition that commercial Earth observation systems can contribute in maintaining international peace and security in a cost-effective way.

This paper provides an overview of the resource management and surveillance requirements which led to the initial approval of the RADARSAT project, reviews the key elements of the operational system, and describes some of the capabilities that could assist the UN and NATO in promoting international peace and security. Recent research and ongoing trials aimed at demonstrating the ability of the RADARSAT system to satisfy the operational requirements associated with some of these peace and security applications is also addressed.

The ability of RADARSAT to support global multilateral operations is discussed using a generic scenario which addresses the overlapping information requirements that would be associated with a three phase operation involving: predeployment, deployment, and postdeployment. Surveillance sensor integration and data fusion are discussed from a surveillance architecture perspective, and the need for multi-agency cooperation is examined in the context of ongoing work aimed at using RADARSAT data to cost-effectively enhance Canada's recognized maritime picture.

The results presented summarize an important area of ongoing multi-agency cooperation associated with the use of RADARSAT to promote Canada's interests, and as a means of promoting international cooperation and understanding. User acceptance and risk management issues identified as part of ongoing efforts to harmonize Canadian civil and military space-related activities are reviewed in the context of multilateral security cooperation, dual use technology imperatives, and fiscal restraints.

Paper presented at the AGARD SPP Symposium on "Remote Sensing: A Valuable Source of Information", held in Toulouse, France from 22-25 April 1996, and published in CP-582.

2. INTRODUCTION

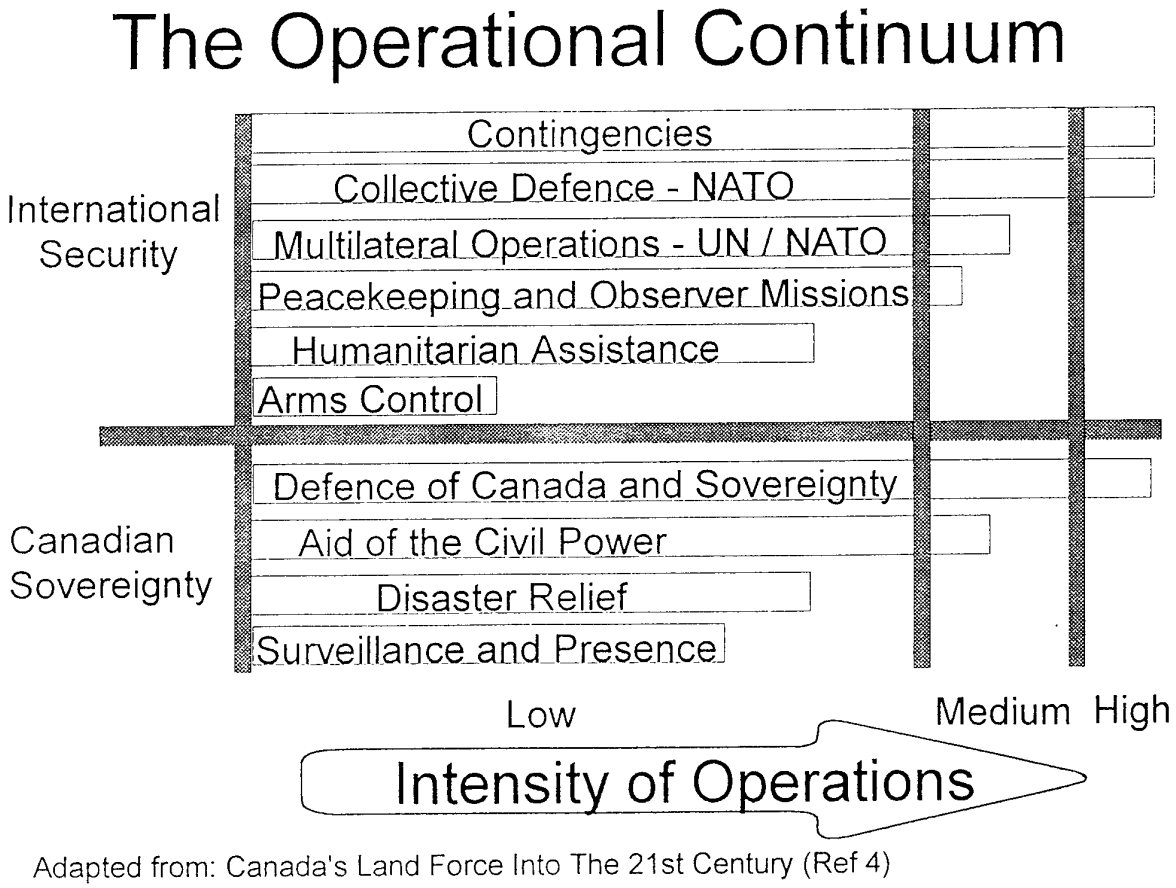
The Cold War is over, and change on an unprecedented scale is sweeping away old ways of thinking (Ref 1). Today's business and defence decisionmakers are confronted with a complex and uncertain world, with technologically advanced dual use systems providing hitherto unimaginable capabilities and threats, in which reduced resources and risk management trade-offs are becoming the norm.

Although RADARSAT was not designed to support specific defence requirements, ongoing cooperation between the Canadian Space Agency (CSA) and the Department of National Defence (DND) indicates that it has the capabilities required to contribute to the effectiveness and efficiency of several sovereignty and security tasks (Ref 2). Multiple Synthetic Aperture Radar (SAR) modes, with relatively high resolution and variable swath/resolution/incidence angles offer: global reach, all-weather/day-and-night wide-area coverage, independent verification, and resource optimization opportunities.

2.1 The Evolving Operational Continuum

The international environment has evolved dramatically since 1989, when the Berlin Wall gave way to what has widely, and inappropriately, been referred to as a "new world order". Significant advances in arms control, conflict resolution and openness have been made, and this has greatly reduced the threat of global war. However, localized pockets of chaos and instability have also emerged, and the proliferation of advanced technologies and weapon systems will increasingly challenge the ability of existing collective security organizations to cope with an unpredictable and fragmented world (Ref 3).

Within the developed world, especially the G7 countries, concerns about an uncertain and unstable world abroad have not generally overcome pressing domestic socio-economic considerations that have generated significant defence cuts. In fact, economic and political imperatives are clearly shifting the focus away from traditional military mid/high intensity operations to multi-role capabilities aimed at the entire operational continuum, as outlined in Figure 1.



Adapted from: Canada's Land Force Into The 21st Century (Ref 4)

Figure 1

2.2 Information Requirements

Today's decisionmakers face an increasingly complex and uncertain world in which information is widely believed to hold the key to most (if not all) of the problems faced by humanity (Ref 1). This situation has been exacerbated by technological advances, which are permitting the integration of vast data archives with an ever increasing amount of space-based and other surveillance data, the disdain which many individuals have for the complexities associated with the operational use of information and technology, the perishable nature of most operational information, and ongoing fiscal restraint.

Ideally, especially when resource constraints can be ignored, it should be possible to acquire and appropriately use all of the information required to solve a given problem. Unfortunately, operationally useful information can be incomplete and/or inconsistent and the management of this uncertainty inevitably involves trade-offs which are driven by the realization that perfect information is worthless if it is not available when it is needed (Ref 5).

From an operational surveillance perspective, the observation of environmental phenomena and human activities can most effectively and efficiently be monitored by a multi-stage process that involves: **detection, location, classification, identification,** and **inspection** (Ref 6). Each of these stages has different requirements, spanning the spectrum from the relatively coarse information needed for wide-area detection to the very focused and precise information associated with inspection. The overall surveillance function can best be accomplished by a surveillance architecture which contains an appropriate mix of platforms and sensors.

2.3 Surveillance Architectures

Global or regional operations require a wide range of data and information, encompassing both archive and near real-time inputs (atmospheric, surface and sub-surface), from an assortment of platforms and sensors to observe changes and to assist in forecasting likely future events. Although a wide range of commercial and defence applications and scenarios are frequently associated with specific surveillance platforms/sensors, multi-purpose surveillance architectures which incorporate multiple data sources provide the most accepted means of providing: guaranteed information; uncertainty management; and adequate risk management (Ref 6).

In Canada's case, a vast and diverse geography and global interests (including UN and NATO responsibilities) combine to create very demanding sovereignty and security surveillance requirements. For a variety of reasons, Canada has opted to focus on the maintenance of multi-purpose capabilities that can contribute to a variety of domestic and international objectives (Ref 3). This approch, which is clearly gaining favour in the Post-Cold War era, is designed to provide government with the broadest range of options while leveraging benefits from related resource expenditures and appropriate dual use opportunities.

The benefits of a multi-purpose surveillance architecture, Figure 2, are especially evident when considering the surveillance of Canada's remote Arctic and maritime approaches. With a 1,200 + Km area of interest radius, stretching past the North Pole, cued classification/identification/inspectionofuncooperative objects can most cost-effectively be accomplished using wide-area detection and location information from a satellite like RADARSAT (Refs 6, 7, 8).

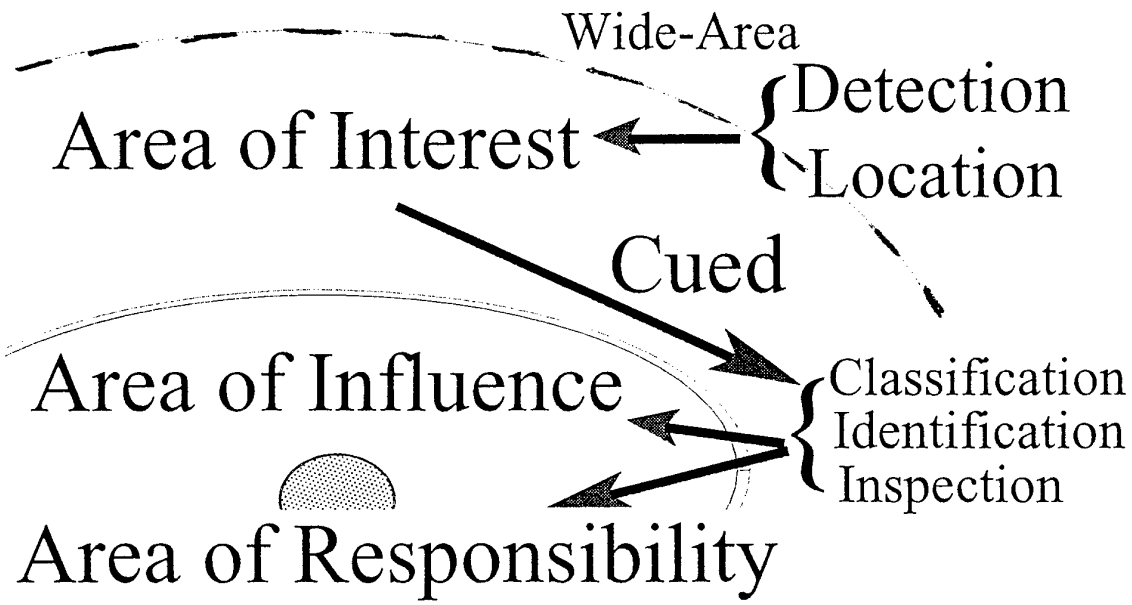


Figure 2

3. THE RADARSAT SYSTEM

RADARSAT is a Canadian-led project that involves several provinces, a number of government departments, the private sector, and the United States. Although the SAR satellite features prominently in all project related information, it is important to note that RADARSAT is an end-to-end system (command, control, reception, and exploitation), which is designed to be the world's first operationally-oriented commercial radar satellite. Numerous RADARSAT related documents exist, spanning the spectrum from detailed technical references to glossy marketing material. Individuals wishing more than the basic information presented here are encouraged to refer to the technical overview contained in the Canadian Journal of Remote Sensing Special Issue RADARSAT (Ref 9), contact the CSA, and consult with RADARSAT International (User Guide to Products and Services - Ref 10).

3.1 The RADARSAT Mission

RADARSAT is Canada's commitment to space technology for better resource management and better understanding of our environment, for the benefit of Canadians and humanity (Ref 7). The RADARSAT era achieved an important milestone with a successful launch on 4 November 1995 by the United States National Aeronautics Space Administration (NASA), and current plans call for data continuity through the launch of a second satellite in 20001.

As outlined in the 1977 "Satellites and Sovereignty" report of the interdepartmental task force on surveillance satellites (Ref 6), the requirement for a Canadian SAR surveillance satellite encompasses both civil and defence applications. Specific emphasis is placed on addressing the sovereignty and economic surveillance challenges associated with Canada's vast and diverse territory, as depicted in Figure 3.

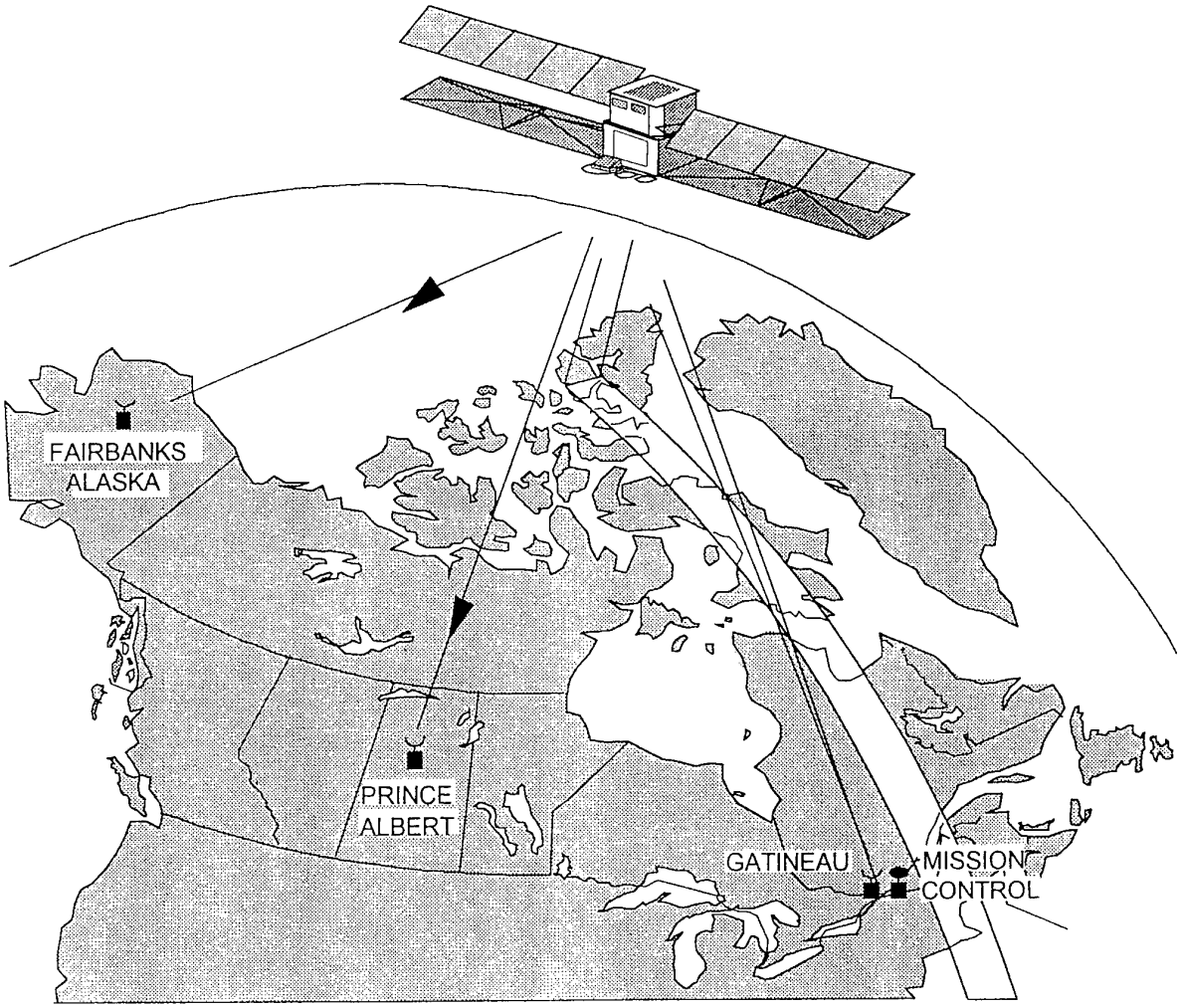


Figure 3

3.2 The Radar Imperative

Radar is an essential component of any Canadian surveillance system because of its ability to permit observations independent of cloud, fog, or darkness (Ref 6). The use of imaging SAR sensors, which permit the detection of relatively small targets and physical features from satellites, is especially attractive for three very different reasons. First, SAR provides unique information for many non-surveillance operational and non-operational applications (especially when multiple modes can be acquired, as shown in Figure 4). Secondly, and of significant importance, as an imaging sensor SAR provides contextual information which can assist in scenario assessment. Third, and perhaps most importantly, the technological skills associated with the processing, handling, analysis, integration, and dissemination of large amounts of digital SAR data encompass many of those required by modern knowledge-based economies and defence organizations (see Table 1). In this context, Earth observation expenditures can be viewed as potentially important engines for social and economic progress into the 21st century while providing immediate commercial and sovereignty/security benefits (Ref 2).

Table 1

RADARSAT SAR Characteristics (Ref 7)

Frequency	5.3 GHz / C Band
RF Band width	11.6, 17.3 or 30.0 Mhz
Data Rate (max)	105 Mb/s (R/T) 85 Mb/s (recorded)
Polarization	HH
Aspect	Right looking (North) Left looking (Antarctica)
SAR on-time	28 minutes / orbit (see Table 2 and Figure 5)
Modes	Multiple (25 pre-defined)
Tape recorders	2 X 10 minutes capacity

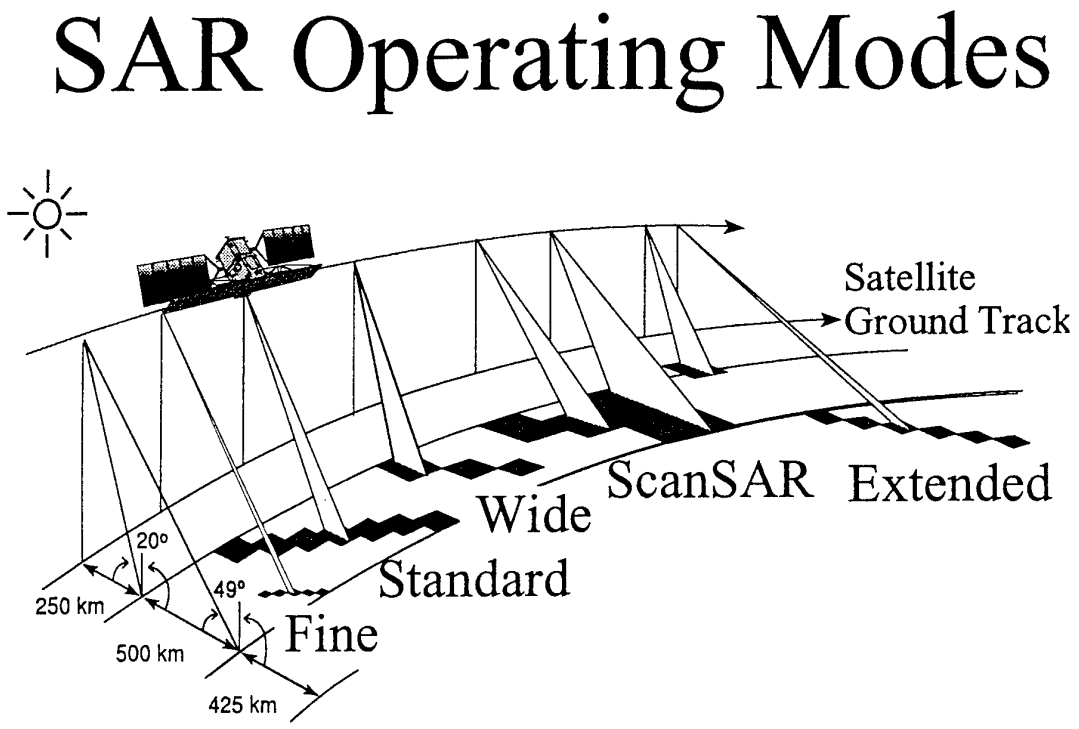


Figure 4

Table 2

RADARSAT Orbit (Ref 7)

Geometry	Circular, sun-synchronous (dawn-dusk)
Altitude	798 km
Inclination	98.6°
Period	100.7 minutes
Orbits Per Day	14 (Descending Equatorial Crossing at 06:00hrs)
Repeat Cycle	24 days
Coverage	Global 3 - 5 days Canada 2 - 3 days (varies with SAR mode)
Power	28 minutes/orbit SAR imagery (rarely in eclipse - Figure 5)

3.3 RADARSAT Application Areas

The application areas which can benefit from the information that RADARSAT will provide are limited only by user requirements and inventiveness. General guidelines for nine top level and 36 specific applications are provided in the RADARSAT user guide (Ref 10).

In the surveillance or environmental monitoring context, RADARSAT has significant potential as both a sensor of opportunity and as a dedicated wide-area surveillance asset. From disaster response (e.g., earthquake and oil spill mitigation) to systematic ice/maritime reconnaissance, RADARSAT can provide cost-effective change detection for large areas of interest.

In the defence and humanitarian assistance context, RADARSAT is expected to provide varying degrees of utility in four "Peace and Security" application areas: maritime surveillance; geographic support; intelligence support; and research and development. The ability of RADARSAT and other medium to low resolution satellites to satisfy the operational requirements associated with each of these application areas needs to be demonstrated, which is the subject of an ongoing multi-agency evaluation, and feasible solutions for the operational issues noted in Table 3 need to be identified.

Table 3

Operational Issues for Peace and Security Applications	
Most or All Applications	Target Characteristics Data Fusion Options Data Reception Processing Facilities Mode vs Revisit Information Integrity Data Policy
Maritime Surveillance	Near Real-Time Needs Sea State and Wind Multiple Users Cued Response Time
Geographic Support	Seasonal Variations Scale and Format Coastal Areas 2D vs 3D Projections
Other	User Specific

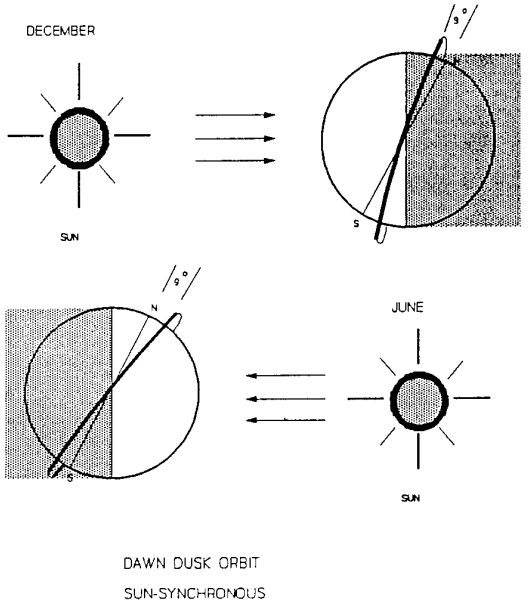


Figure 5

4. PROMOTING PEACE AND SECURITY

Earth observation information can be used in many different ways to satisfy the information needs of a broad spectrum of users. The importance of this information has been recognized by the UN, and several countries have supported calls for an international satellite monitoring agency to promote peace and security in all its forms - political, military, economic and environmental (Ref 11). Recent developments, especially those associated with the Persian Gulf and Bosnia conflicts, clearly suggest that non-classified commercial Earth observation imagery offer can cost-effectively enhance some UN and NATO operational capabilities (Refs 12, 13, 14).

The precise impact that the availability of non-classified all-weather/day-and-night information over a wide range of resolutions and incidence angles will have is still being evaluated, as shown in Table 4. However, a quick examination of the information needed to support multilateral operations, some observations on how RADARSAT capabilities could be used to support a generic multilateral operations scenario, and an overview of related sovereignty surveillance issues provides some food for thought.

4.1 Multilateral Operations and Information

Traditionally, Division and Brigade level information needs have been driven by the immediate operational concerns associated with troops in contact. However, the trend is towards UN and NATO global multilateral operations with large areas of responsibility, and recent experince clearly indicates that better use of space systems is required to acquire and disseminate the information needed to support these operations (Ref 12).

Although the scope and objectives of each mission will vary, the need for accurate and timely information must invariably be addressed during each operational phase: **Predeployment**, **Deployment**, and **Postdeployment** (Ref 13). This is especially important in the context of the new international environment and ongoing and anticipated UN/NATO humanitarian missions involving short notice and potentially significant resource commitments. Unless adequate predeployment and deployment information can be obtained and updated systematically, with space-based military and/or commercial Earth observation systems offering most of the feasible and cost-effective options, the conduct and success of these missions can be very problematic (e.g., Somalia).

TABLE 4

Anticipated RADARSAT Utility for Typical Point Targets (Ref 8)

TARGET	PERFORMANCE
Buildings	Highly Detectable, Recognizable
Airports, Runways	Highly Detectable, Recognizable
Installations	Individual Components
Large Aircraft (parked)	Detectable, Recognizable
Small Aircraft (parked)	Detectable
Large Ships	Detectable, Classification to Vessel Type
Small Ships	Detectable
Ship Wakes	Detectable
Mobile Land Targets (Tanks, APCs etc.)	Large Formations Detectable
Missiles (on ground), Support Structures	Detectable

Note: A statistically significant number of RADARSAT images need to be evaluated before performance is assessed.

4.2 Multilateral Operations and RADARSAT

To effectively and efficiently support multilateral operations, increasingly better use of space systems will be required to provide accurate and timely information for strategic and tactical commanders (Refs 12, 13). Information is required for *predeployment* planning, before significant resources are committed, during *deployment* as circumstances and available resources change rapidly, and throughout the *postdeployment* phase.

Exact information requirements must be tailored to each mission. However, generic planning scenarios typically begin with the acquisition and independent verification of baseline geospatial and humanitarian information for the entire deployment area and immediate surroundings. Access and staging points need to be identified (airports, ports, etc.), terrain trafficability/hydrology analysis and land use details must be combined with an assessment of available infrastructure, and potentially significant humanitarian issues need to be identified early. In addition, modelling of deployment/postdeployment options should support both national and multilateral decisions to commit, defer, or reject each potential UN and NATO mission.

After a multilateral operation has been approved the operational commander should have sufficient responsibility and resources to finalize planning and preparations, execute the deployment and begin tactical operations. In practice, however, the overlapping nature of the information sources and processes associated with independent national decision making, and resource constraints inevitably require national and/or multilateral resources to supplement operationally assigned capabilities.

Security considerations, which vary with each multilateral operation, have historically created information transfer problems. As collecting/processing/controlling information continues to grow in importance, it is unlikely that *sensitive national sources* will readily or consistently be available for unconstrained release. Innovative solutions, like providing deployed force commanders with transportable facilities for the direct reception and exploitation of multiple Earth observation satellites (Ref 13), need to be accepted at all levels if operational effectiveness is to be maintained or enhanced as multilateral operations evolve to encompass a broader range of military activity (Ref 11). In this context, especially when data fusion opportunities are considered, the vital contribution that commercially available non-classified Earth observation data can have in promoting international peace and security should be clear (see Figure 6 for a preliminary assessment of RADARSAT applicability).

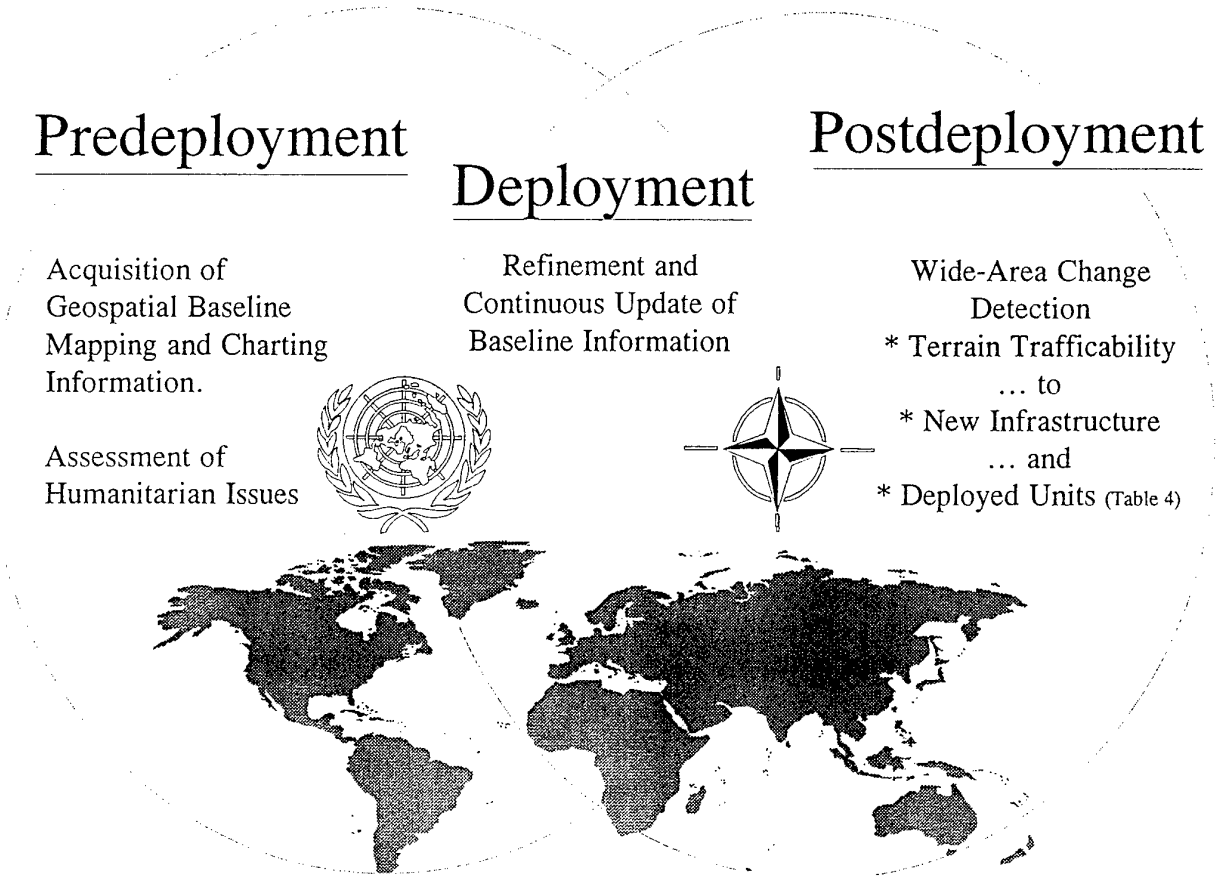


Figure 6

4.3 Sovereignty Surveillance

In the Canadian context, sovereignty control requires the surveillance of an immense territory and extensive ocean exclusive economic zones. To effectively and efficiently maintain Canadian sovereignty, an integrated surveillance architecture composed of an appropriate mix of sensors and systems - including surveillance satellites - is required (Ref 6). Some of the components of a possible surveillance architecture are depicted in Figure 7.

Clear parallels can be drawn between a number of sovereignty surveillance tasks, and elements of several recent or ongoing UN and NATO missions. Although these similarities are not currently pervasive, recent calls for the UN to take a more active role in environmental surveillance and related activities could provide the required operational push for countries like Canada to more fully integrate their sovereignty and international surveillance capabilities (Ref 11).

Several comprehensive studies and reports are currently being pursued on most of the sovereignty control functions and tasks noted in Table 5, and results are expected in the near future (Ref 15). In the interim, based on preliminary analysis of RADARSAT commissioning data, it is clear that a wide range of sovereignty surveillance information can be extracted from RADARSAT data (Ref 16).

Table 5

Sovereignty Surveillance and
Suggested RADARSAT SAR Imaging Modes

Renewable Resource Management (e.g., Fisheries)
- SCNfar, W2, W3, S4-S7, F1-F5, EH1-EH6

Non-Renewable Resource Management (e.g., Oil Rigs)
- W2-W3, S3-S7, F1-F5, EH1-EH6

Marine Environment (e.g., Oil Spills)
- SCNnear, W1, S1-S3, F1-F2

Navigation Control (e.g., Ice Routing)
- SCNwide, SCNnear, SCNfar, W2-W3, S3-S7

Defence (e.g., Base Map Production)
- W1-W3, S1-S7, F1-F5

International Ocean Management (e.g., Weather)
- SCNnear, W1, S1-S3

Ocean Service (e.g., Search and Rescue)
- SCNnear, W1-W2, S1-S7, F1-F5

Note: Image resolution varies with swath coverage
F = Fine, S = Standard, W = Wide, EH = Extended High, SCN near/far/wide are Multiple Beam Products (Wide-Area to Focused Coverage - See Ref 15)

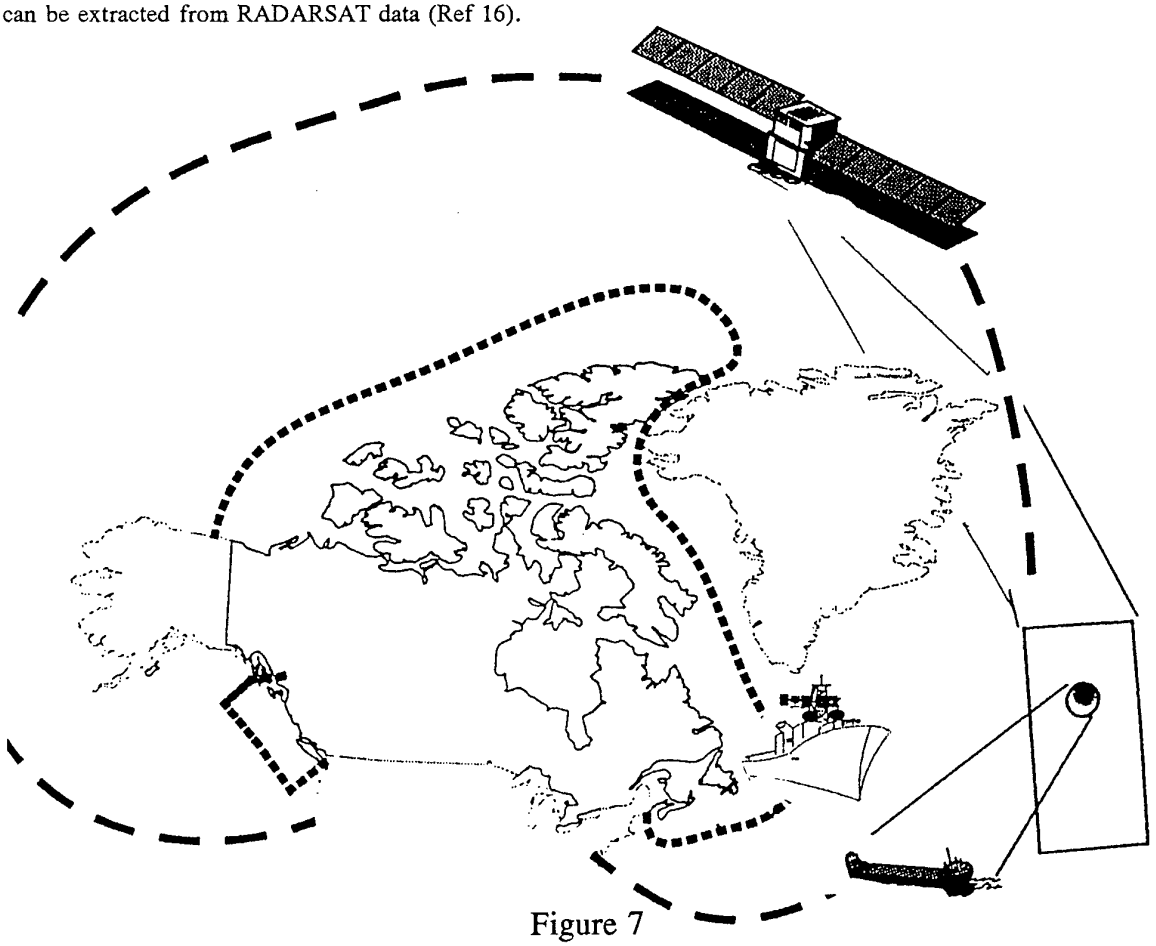


Figure 7

5. THE DUAL USE BUSINESS CASE

To effectively use RADARSAT each potential user needs to understand and anticipate how they can address the user acceptance and risk management issues that are associated with integrating commercial space-derived information into their operations.

5.1 User Acceptance

Until recently, satellite imagery and derived information in the 1 to 9-meter resolution range was closely associated with the classified programs of a limited number of countries and their close allies. While high resolution imagery, less than 1-meter for optical and less than 3-meter for SAR, was generally assumed to offer the greatest military utility (despite tight security and wide-area coverage restrictions).

In 1990/91 perceptions that low and medium resolution systems offered little operational capability were severely tested in the Persian Gulf. Fuelled by the successful exploitation of non-classified 10-meter SPOT and 30-meter Landsat imagery for defence and humanitarian purposes, a new higher resolution commercial Earth observation era is beginning to emerge (Ref 14). However, not everyone has become an overnight dual use believer, and proponents of tightly controlled classified systems still need to be convinced of the dual use benefits.

5.2 Cost-Effectiveness

In view of the technological advances that have occured and are anticipated in the near-term, it is reasonable to assume that the tools and techniques required to disseminate and exploit all-source imagery/information will increasingly be dual use in nature. In this environment, which is already upon us, advanced economies and modern defence organizations are all becoming more dependent on data fusion and the rapid flow of digital information to remain competitive. As stovepipe structures and systems continue to be replaced by integrated systems it is very likely that the tools and techniques required to use the entire spectrum of information, including medium resolution satellite imagery, will become essential to keep pace operationally in an environment where short decision cycles are becoming the norm.

With little or no operating cost differential expected between most future single and dual use information support systems (Ref 1), commercial vs classified imagery cost-effectiveness should depend on data acquisition costs. For commercial off-the-shelf satellite imagery users this depends on the amount of data used, as outlined in Table 6.

Table 6

Area Coverage for Potential Crisis Support

		Images @ 30 Days (RADARSAT Modes)	Sq. Km @ 30 Days
Bosnia			
	Fine Mode	90	180,000
	Standard	90	900,000
Haiti			
	Fine Mode	30	60,000
	Standard	34	340,000

Note: Numbers based on conservative simulations (use of multiple satellites provides more frequent coverage, and multiple spectral bands and polarizations provide additional information content)

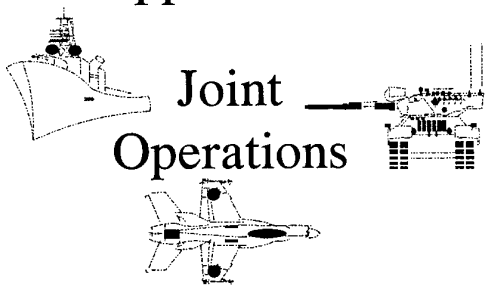
5.3 Resource Optimization

When all is said and done (from the analysis of costs to the impact of system performance on operations), a hybrid or integrated surveillance solution is likely to offer the most acceptable resource optimization opportunities for the widest number of applications. Where appropriate, this builds on the reliability and performance of classified programs, and leverages available dual use commercial systems and capabilities.

It is the opinion of the authors that ongoing efforts to address the identified operational issues associated with the use of commercial Earth observation systems for peace and security applications need to continue, as outlined in Figure 8, and should be accelerated in some areas. Specifically, non-aggressive civil and military dual use should be supported, opportunities to influence the design and acquisition of future capabilities must be carefully examined, and appropriate measures to prevent unauthorized exploitation of current and future capabilities need to be pursued.

Operational Requirements

Identification of Potential Applications



Joint Operations

Doctrine and Policy Considerations

Evaluation and Demonstration

Potential Enhancements

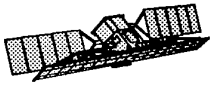


RADARSAT-1

Data Policy

Rapid Access to Data

* User Specific Processing
* Additional Beam Modes
* Transportable Station



RADARSAT-2 and Successors

Increased Information Content

* Resolution
* Multiple Polarization
* Dual Frequency

Information Security

* Encryption
* Access to Archive

Dual Use Benefits

Sophisticated Off-The-Shelf Capabilities

Jobs and Economic Development

International Market Niches



Figure 8

6. REFERENCES

1. Toffler A., and Toffler H., "War and Anti-War", Toronto, CAN, Little, Brown and Company, 1993.

2. Dionne J.L.P. LCol (Ed), "Canadian Space Agency (CSA) Long Term Space Plan and Department of National Defence (DND) Space Plan Synergies Process Report", January 1995.

3. Department of National Defence (DND), "Defence White Paper", 1994.

4. Department of National Defence (DND), "Canada's Land Force into the 21st Century", 1994.

5. McKean M.C. Maj, "Management of Uncertainty in Knowledge Based Systems: A Survey ", Unpublished M.Sc. Thesis, RMC, CAN, 1992.

6. Government of Canada, "Satellites and Sovereignty", Report of the Interdepartmental Task Force on Surveillance Satellites, August 1977.

7. Canadian Space Agency (CSA), "RADARSAT Mission Requirements", RS-CSA-SP0001-NC, September 1991.

8. Charpentier R., and Rey M., "Space-Based Surveillance: Technology and System Options Summary Report", CRAD Space Studies Report No. 5, October 1995.

9. Canadian Journal of Remote Sensing, "Special Issue RADARSAT", Volume 19, No. 4, November-December 1993.

10. Radarsat International (RSI), "RADARSAT Illuminated", RSI User Guide, Preliminary Version, July 1995.

11. United Nations (UN), "International Cooperation in Space Activities for Enhancing Security in the Post-Cold War Era", Report of the Secretary-General, 1994.

12. Department of Defence (DoD), "Conduct of the Persian Gulf War Final Report to Congress", 1992.

13. National Research Council, "Strategic Technologies for the Army of the Twenty-First Century: Special Technologies and Systems", National Academy Press, Washington, D.C., 1993.

14. Nanz T., "Bosnian Victory for Remote Sensing", Space News, February 19-25, 1996, pp 19-20.

15. Vachon P., and Olsen R., "RADARSAT SAR Mode Selection for Marine Applications", Backscatter, October 1995, pp 3-18.

16. Bjerkelund C., "News from Ottawa", Backscatter, February 1996, p 4.

Paper 38

M. McKean (CA)

Quelle est la politique de commercialisation des images Radar Sat?

[Does a policy exist relating to the commercialization of Radar Sat images?]

Author's reply:

En resumé: contrôle gouvernemental.

[There is government control].

N.N.

What are the commercial arrangements associated with RADARSAT, and how dos this impact on data policy?

Author's reply:

The RADARSAT programme was conceived as a government-industry partnership, with the United States providing the launch in exchange for access to a percentage of data, and RADARSAT International has international distribution rights. Access to data is controlled within a framework that respects the United Nations treaty on the exploitation of space and related remote sensing principles. In the event of potential sensitivity, associated with international peace and security, the Canadian government will make a decision - on a case by case basis.

Automatic change detection in space borne SAR imagery.

D G Corr
Space Department,
R16 Building,
DRA Farnborough,
Hants., GU14 6TD, UK

S W Whitehouse
Data Sciences (UK) Limited.
Meudon House,
Meudon Avenue,
Farnborough, Hants., GU14 7NB, UK

1. SUMMARY

This paper describes new techniques for the automatic detection of change within synthetic aperture radar (SAR) images produced from satellite data. The interpretation of this type of imagery is difficult due to the combined effect of speckle, low resolution and the complexity of the radar signatures. The change detection techniques that have been developed overcome these problems by automatically measuring the degree of change between two images. Changes within a time sequence of images (for example due to demolition or construction; damage; collection or dispersal of mobile assets) can then be determined by logic.

The principle behind the technique used is that when satellite repeat orbits are at almost the same position in space (i.e. within a few hundred metres) then unless the scene has changed the speckle pattern in the image will be unchanged. Comparison of images therefore reveals genuine change, not change due to fluctuating speckle patterns.

The method has been validated using local information on building changes due to construction or demolition, and changes in vehicles within car parking areas. With data from ERS-1 the minimum area over which change can be detected is approximately 50 m by 50 m.

2 INTRODUCTION

The extraction of information of military utility from satellite SAR images is difficult due to several factors:

a. The non intuitive types of signatures exhibited by man made objects.
b. The speckle noise inherent within the imagery.
c. The relatively low resolution (e.g. 25 m) of data available from civil satellite SAR systems.

A new technique has been developed that is suitable for change detection, and which overcomes the problems associated with signature recognition and speckle. This technique has become practicable with data from ERS-1 (Earth Resources Satellite 1) as a result of the high stability of the satellite's orbits. The orbit stability is such that data can be acquired on separate passes from positions which differ by only a few hundred metres across track. At a slant range of about 850 km this means that repeated viewing takes place from almost the same place in the sky. Under these observation conditions the radar signatures of stable objects are invariant (i.e. the speckle pattern is frozen) and an automatic method of change detection is possible.

The change detection technique that has been developed is suitable for three military applications:

a. The identification of places where there has been a change in the disposition of mobile assets.
b. Damage assessment
c. Construction or demolition

This paper describes the change detection techniques developed and their validation.

3 PRINCIPLE OF CHANGE DETECTION

3.1 Image speckle

SAR imagery has a characteristically speckled appearance. This random variation in pixel amplitude has the effect that an area of uniform characteristics (e.g. a grass field) does not appear with a uniform tone within a SAR image. A SAR image of part of Farnborough, obtained by ERS-1, is shown in the upper part of Figure 1. This shows a 4 look 25 m resolution (C band) SAR image. An idea of the potential information content within the scene can be judged from the corresponding optical image, obtained from the Russian DD5 system (resolution 2 m), shown in the lower part of Figure 1. The effect of speckle is immediately apparent. The SAR image is largely uninterpretable.

The speckled appearance of SAR imagery arises because the value of each pixel is the sum of all the signals received from within the resolution cell of the SAR system (25m by 25 m for the image shown). These signals originate from the often large number of the scattering elements which constitute a particular object. It is important to realise that if the object is unchanged between successive observations and the observation point is at the same position then the result will be unchanged. The random speckle pattern observed in an image would be unchanged i.e. it would be frozen. Under these conditions any changes to an object result in an observable change in the speckle pattern.

The repeat orbits of ERS-1 are almost parallel. The constraint on the distance between the orbit tracks necessary to satisfy the requirements for frozen speckle is that they are within a few hundred metres. A large number (typically 20%) of ERS-1 repeat orbits satisfy this constraint.

© British Crown copyright 1996/DERA
Published with the permission of the controller of Her Britannic Majesty's Stationery Office

Paper presented at the AGARD SPP Symposium on "Remote Sensing: A Valuable Source of Information", held in Toulouse, France from 22-25 April 1996, and published in CP-582.

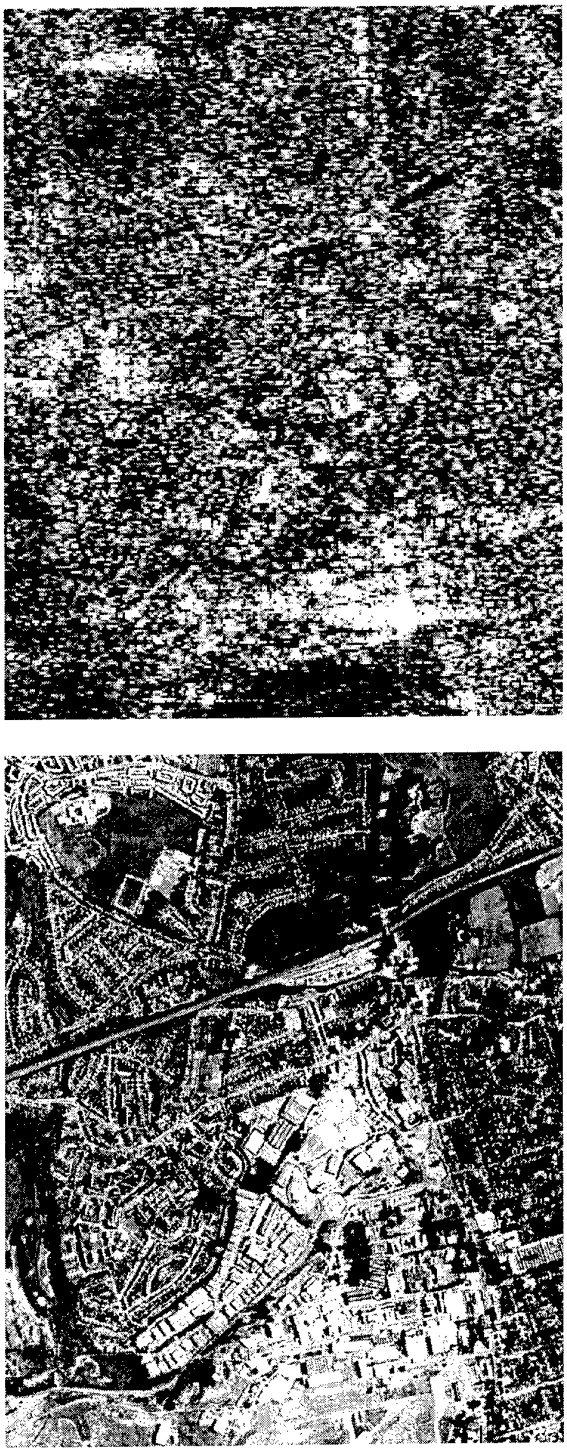


Figure 1 ERS-1 4 look SAR image of Farnborough (above), DD5 image (below) (2.7 km by 2.2 km).

3.2 Scene characteristics

The objects within a scene can be divided into those which are man made and those which are natural, as illustrated below:

Man made

a. Buildings - offices, factories, warehouses, houses etc.
b. Roads
c. Railways
d. Runways
e. Structures - bridges, industrial plant, storage tanks etc.
f. Mobile systems - vehicles, aircraft etc.

Natural

a. Cultivated land i.e. farm land
b. Uncultivated land i.e. open land
c. Forest
d. Rocks
e. Water
f. Parks and gardens
g. Animals and people

The changeability of these objects, together with an estimate of the typical time over which a change takes place is described below:

Unchanging targets

a. Fixed or static man made structures
b. Stable natural structures (e.g. rocks)

Changing targets

a. Unstable natural targets (e.g. lakes, rivers) - seconds, minutes
b. Changing natural targets (e.g. crops, trees etc.) - hours, days, weeks, months
c. Animate targets (i.e. animals, people) - seconds, minutes, hours
d. Changing man made targets (e.g. car parks, roads with traffic) - minutes, hours
e. Changes to man made structures (e.g. construction, demolition) - seconds, minutes, days, weeks.

It is evident that changes in natural targets will depend on the time of year (growth state and growth rate) and meteorological factors such as wind.

The next section describes strategies for detecting changes associated with man made targets.

3.3 Change detection strategies

If two scenes are compared, the similarity of the scenes depends on the time interval between observations and the scene content, as described in the preceding section. (The technique by which similarity is measured will be described in Section 4.4.) By consideration of the times for object change, it is evident that two change detection strategies are possible; one in which a comparison is made between images obtained over a relatively short time interval, and one in which the time interval between observations is large. In the first case it would be expected that the images would be highly similar, except for natural regions which are known to be rapidly changing (e.g. water). Relatively sudden man made changes (e.g. due to vehicle movements or demolition) would be directly apparent in an otherwise unchanging scene. In the second strategy, when the time interval between observation is relatively large, a comparison between two images will define the unchanging areas. These areas are due to man made objects and inert natural structures such as rocks. Changes in man made structures can then be determined by monitoring

changes to the areas defined to be unchanging within successive pairs of images.

In practice comparison between images is made difficult since man made objects generally coexist with natural objects (e.g. trees next to a building). This results in the properties of pixels becoming mixed; they are partly stable and partly changing. The likelihood of this occurring depends primarily on the radar resolution and the objects' physical size. There is also another mechanism which will also introduce mixed properties into the signatures from man made targets, this occurs when there are components of the target return which involve multiple scattering with natural targets (e.g. scattering from buildings onto trees or vice versa).

4 PRACTICAL REQUIREMENTS

The practical requirements for change detection are methods and tools for the following functions:

a. Data selection
b. Data co-registration
c. Pre-processing
d. Measurement of change between images
e. Change detection

Each of these functions are considered in turn below.

4.1 Data selection

The region chosen for validation of the change detection techniques developed was Farnborough UK.

Selection of data for analysis then depended on the following factors:

a. That the satellite overflies the site of interest and that the radar system was operating.

That pairs of images can be obtained from orbits which are similar in position.

The operational characteristics of ERS-1 depend on the scientific objectives defined by the European Space Agency. Several phases of operation have been defined. These are listed up to September 1994 in Table 1.

Phase	Phase class	Date range	Repeat cycle (days)
A	Commissioning	Jul 91 - Dec 91	3
B	1st Ice	Dec 91 - Mar 92	3
C	Multi-disciplinary	Apr 92 - Dec 93	35
D	2nd Ice	Dec 93 - Mar 94	3
E	Geodetic	Apr 94 - Sep 94	168

Table 1 ERS-1 operating phases

The orbit characteristics for each class of phase are different, and consequently data from different types of phase cannot be used together (e.g. data from Phases B and D can be used together, but data from Phase B cannot be used with Phase C).

Only Phases C and D provided useful and recent coverage of Farnborough. Phase C spanned over a year, potentially allowing all seasonal effects to be studied. Phase D data allowed change over small time increments to be studied, but was only available for the winter months January to March.

Recently with the launch of ERS-2, pairs of images 1 day apart can be obtained during "tandem" operations of the satellites ERS-1 and ERS-2. Each satellite is however in a 35 day repeat orbit.

The precise orbital characteristics of ERS-1 vary as a consequence of air drag and the gravitational effects of the sun and the moon. The severity of air drag increases during the time of solar flares. The operational performance of ERS-1 has been controlled so that the satellite is within 1 km of a nominal reference orbit. The European Space Agency has tabulated (Ref. 1) the actual orbit separations between all pairs of orbits for which data has been acquired for which the separation is less than approximately 1.1 km. This data was used to select viable image pairs for change analysis. It was found that approximately 20% of the all the data pairs available for Farnborough during Phases C and D have separations less than 300 m.

4.2 Image co-registration

Precise co-registration between pairs of complex full resolution SAR images is required. Without this, apparent changes between two images would be perceived when a comparison is made, simply because there is not a like-with-like comparison.

The registration procedure used operates by cross correlating the amplitudes of two complex images. The position of the peak in the cross correlation function determines the image shift required to co-register the images. Cross correlation is performed not between the whole images, but on a number of smaller regions sampled from the image. The reason for this is that the properties of the images are not homogeneous. At some places there is a high degree of correlation between the images, but at other places it may be very low. The shift necessary for co-registration was derived from the data from the regions which had good correlation. The resulting co-registration accuracy was approximately 1/8 of the pixel dimensions. A specific SAR image was designated to be the master image and all of the other SAR images used were registered to it. One master image was selected for Phase C and one for Phase D.

4.3 Data pre-processing

In Section 4.4 it will be explained that the measurement of change is based upon the computation of the complex product $I_1 I_2^*$ between two images I_1 and I_2. This measurement process will be degraded unless the following pre-processing steps are performed

a. Range spectra noise reduction
b. Interferometric phase removal

If it were possible to make repeated observations from exactly the same point in the sky then both of these processes would not be required. In practice there is a small difference in the incidence angles for the observations at the surface. In order to understand the effect of this, consider the spectrum of the image data taken in the across track (i.e. ground range) direction. It can be shown (Ref. 2,3) that there is a relative frequency shift between the observations of Δf in the ground spectra. Consequently different parts of the ground spectra are admitted within the receiver bandwidth, as shown in Figure 2 (W is the transmitter chirp pulse bandwidth (15.5 MHz)).

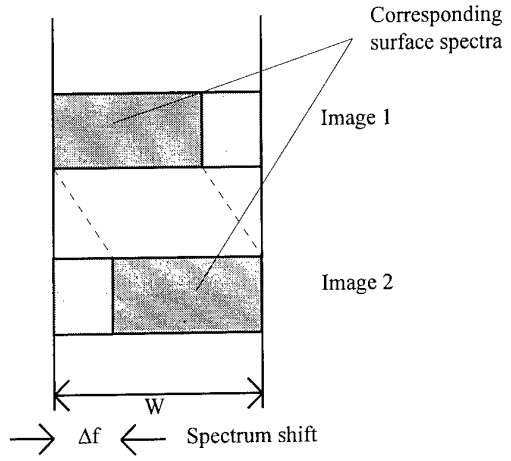


Figure 2 Image spectra

The relative shift between the image spectra Δf is given by

$$\Delta f = \frac{-f_0 \Delta\theta}{\mathrm{Tan}(\theta - \alpha)} \qquad 1$$

θ = surface incidence angle
Δθ = difference in incidence angles between observations
α = surface slope (to horizontal)

The process $I_1 I_2^*$ is equivalent to the convolution of the spectra of the two images. Therefore the parts of the spectra which are not related by a frequency shift (shown in lighter shading) will introduce noise into this process. This noise can be removed by filtering.

A potential difficulty in the implementation of the noise reduction filter arises because the ground is generally not flat (i.e. $\alpha \neq 0$). Nevertheless the implementation of a simple constant bandwidth filter is still beneficial for a first order reduction in noise in the process $I_1 I_2^*$.

The interferometric phase removal process is now described. Since the corresponding parts of the image range spectra are shifted in relative frequency by Δf then convolution in the frequency domain will result in an output which is modulated at a frequency of Δf. It is necessary to remove this modulation prior to measurement for similarity. This can be achieved by cancelling the following across track phase progression from the product $I_1 I_2^*$

$$\phi(x) = \frac{4\pi}{\lambda} \Delta\theta \mathrm{x} \mathrm{Cos}\theta \qquad 2$$

where x is measured in the across track direction (ground range), and the radar wavelength is λ.

Typical values of the rate of change of φ (in degrees per m) with across track position x are given in Table 2 for several observational baselines, and for a slant range of 850 km. (The phase φ is termed the interferometric phase.)

Baseline (m)	Horizontalrate of change of φ (deg/m)
100	1.4
300	4.1
500	6.9

Table 2 Rates of change of φ

This phase correction assumes that the imaged surface is flat. A digital elevation model would be required if the local height variation within a scene was significant; however a simple correction assuming a flat earth removes the first order effects.

4.4 Measurement of change

In practice because of mixed pixel properties and some inevitable small differences in the viewing position there are always likely to be some changes between images, even for nominally unchanged objects. It is therefore necessary to measure the degree to which pixels are similar, rather than to attempt a direct yes/no type of decision on pixel similarity. This can be performed by the computation of the coherence (Ref 1,2) C between two registered complex (full resolution) SAR images:

$$C = \frac{\left|\left\langle I_1 I_2^* \right\rangle\right|}{\sqrt{\left\langle |I_1|^2 \right\rangle \left\langle |I_2|^2 \right\rangle}} \qquad 3$$

here I_1 and I_2 represent the pixel values I_{ij} in image 1 and 2 respectively.

$\langle\ \rangle$ denotes average over a small area

The coherence function is computed over a small area. The size of this area is determined by a balance between the need for a reduction in the statistical scatter of the measurement and the spatial discrimination required. For ERS-1 data this required the summation to be performed over a region of approximately 50 m by 50 m.

The coherence is a normalised function in the range 0 to 1. This corresponds to the range of comparison states from entirely different to exactly the same. It should be noted that the normalisation of the function is such that (in the absence of noise) the result is not dependent on target scattering cross section; i.e. it applies equally to both strong and weak targets.

Consideration of the form of the coherence function shows that it is in effect a normalised matched filter process. That is a detection process which attempts to detect a specific signal (in this case a two dimensional signal) with the characteristics defined by I_2 within the signal I_1. A maximal filter response occurs when the two signal are the same.

It is now clear that any systematic phase progression between the images must be removed before the coherence is calculated. Without phase cancellation the integration would result in a meaningless measurement.

4.5 Change detection

The measurement of coherence provides a description of the differences between two images. Change detection involves decision logic based upon coherence measurements at a particular.

Interpretation of the coherence between pairs of images a short time apart (e.g. 3 days) requires regions of low coherence to be defined. These will indicate places which undergo continuous change or where a sudden change has taken place.

The interpretation of the coherence between images of 35 days or greater apart is more involved. At its simplest it can

be performed using three images. If these images are in chronological order 1, 2,and 3, the coherence is calculated between the pairs 1 and 2, and 2 and 3. Change detection requires the determination of places where significant changes in coherence have occurred (i.e. from high to low or vice versa). If the coherence between images 1 and 2, and between images 2 and 3 is denoted as C_{12} and C_{23} respectively. Then those positions are sought where:

$C_{12} < L$ AND $C_{23} > H$ implying construction 4

$C_{12} > H$ AND $C_{23} < L$ implying demolition 5

A major part of the research undertaken was directed towards measuring the level of coherence exhibited by different types of objects, in order to support decision logic. These issues are described in the next section.

5 INVESTIGATION

5.1 Factors for investigation

There are a number of factors which can potentially affect coherence that require investigation. These are:

a. Object type
b. Mixed pixel properties (size of objects)
c. Time of year
d. Interval between observations
e. Separation between repeat viewing positions
f. Meteorological conditions

The strategy by which these factors were investigated is described below.

It is impracticable to identify large numbers of specific objects of one type in order to investigate their properties. Areas of predominantly one type of object were identified. The following classes of object type were defined:

a. Large scale industrial (as typified by parts of the DRA Farnborough site)
b. Small scale industrial i.e. factory units
c. Airfield
d. High density residential i.e. blocks of flats, roads with few trees
e. Low density residential i.e. mixture of trees, houses, gardens and roads
f. Trees
g. Uncultivated countryside
h. Cultivated countryside

The likelihood of mixed properties within the coherence measurement window depends on the relative size of man made object and the measurement window (50 m by 50 m). Large scale industrial buildings are likely to occupy most of the measurement window, although there can be some cars and trees present. In contrast, in low density residential areas a high proportion of the measurement window is occupied by gardens.

The time of year can be investigated by using data from phase C. Various time intervals between pairs of images are available in multiples of 35 days. Time intervals up to 1 year were studied. Data from phase D allows shorter time intervals between images to be studied (in multiples of 3 days), but only for the period January to March.

The spatial separation between orbits is not a freely variable parameter; the separations used varied from 14 m to 420 m..

5.2 Coherence data analysis

The characteristics of coherence for each of the target types were analysed by histograms.

Selection from Phase C data provided 5 image pairs with a 35 day interval between them between September 1992 and September 1993. Image pairs for autumn and winter, autumn and spring, and autumn to autumn (i.e. 1 year apart) were also selected. On analysis of the coherence data it was found that in general the time of year and the time interval between images had no particular effect. It was concluded that 35 days is sufficiently long for the degree of dissimilarity between images to reach a stable characteristic level.

Using data from Phase D, coherence was investigated for time intervals from 3 to 69 days. Inspection of the coherence histogram results for 3 day time intervals showed that very high coherence was generally present for all classes, except for those which were inherently unstable (e.g. the tree regions). As the time interval between image pairs increased the coherence fell to levels similar to those measured from the phase C data. Data from one tandem pass (1 day interval) was also analysed and the results were found to be similar to the data for 3 day intervals.

The mean values of coherence for all the 35 day data and an example of 3 day coherence data are summarised in Table 3. These illustrate the chief results of the analysis, namely that there is high coherence of man made objects over long periods of time and that there are generally high levels of coherence throughout the data for 3 days or less except for targets with inherent instability.

Target Class	Mean coherence 35 days or greater	Mean coherence 3 day
Large scale industrial	0.7	0.9
Small scale industrial	0.5	0.8
Airfield	0.4	0.9
High density residential	0.4	0.9
Low density residential	0.3	0.7
Trees	0.3	0.5
Cultivated land	0.3	0.9
Uncultivated land	0.3	0.6

Table 3 Summary of coherence data

The values in the table are mean values and some exceptions were noted in uncultivated land. For example some areas of short grass (used for sheep grazing) could exhibit high coherence over 35 days.

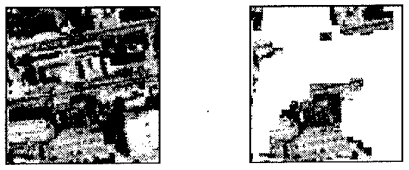
It was interesting to note that the coherence levels were similar for both coniferous and deciduous woodland in both summer and winter.

The spread in the coherence histograms, between the 10% and 90% occupancy levels, was also measured for each target class. The mean value was about 0.5, however higher values

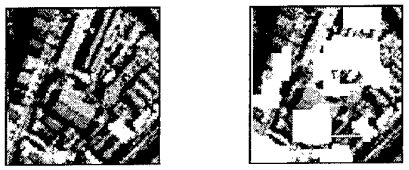
were measured for the large scale industrial class (0.6) and low values for the trees class (0.3). It is thought that these results indicate more about the homogeneity of the regions selected for analysis, than intrinsic target properties (i.e. the tree regions were entirely trees, but even the large scale industrial region had some trees present; there were numerous cars present also).

To illustrate how the coherence varies over a scene, Figure 3 shows data for 35 days time interval (June 1993 with July 1993) for 4 of the target classes investigated. On the left hand side of the figure the DD5 optical data is given for the four specific region types; on the right hand side grey denotes the places where the measured coherence has exceeded 0.7.

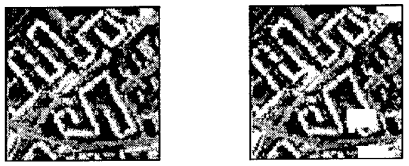
LARGE SCALE INDUSTRIAL



SMALL SCALE INDUSTRIAL



HIGH-DENSITY RESIDENTIAL



LOW-DENSITY RESIDENTIAL

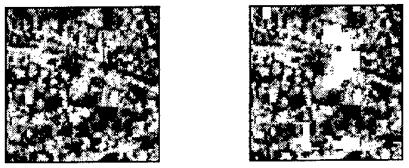


Figure 3 Optical image (left), grey regions superimposed where coherence> 0.7 (right).

When the orbit separation between image pairs was considered, it was consistently found that wide separations resulted in a lower general level of coherence for man made structures, particularly for large scale industrial buildings. This was in spite of the noise improvement afforded by the noise reduction filter pre-process. There are two possible reasons for this. The pre-processing noise reduction filtering process reduces the range resolution within the image data because it reduces the range spectrum bandwidth. For example, for an orbital separation of 400 m the bandwidth is reduced by a third. The range resolution therefore increases by 50%. Consequently mixed pixel characteristics become more likely. Secondly, it is suggested that there may be an uncancelled interferometric phase component in the signatures of tall buildings, this would become more significant for wider baselines.

Meteorological data for the analysis periods was obtained, however no direct effect of weather conditions could be determined. It was unfortunate that extreme differences of conditions between images (e.g. when one image was obtained during rain) did not occur within the data available.

In order to determine the effect on coherence of changes in the disposition of cars within car parks, 5 large busy car parks were specifically delimited for measurement. The mean coherence was found to be 0.5 for both 3 day and 1 day intervals between imaging.

5.3 Change detection

Following the analysis of the coherence data for 35 day or longer time intervals it was clear that changes could only be reliably determined for large structures, when the coherence changed from a high to a low value or vice versa. Logic of the type described by Equations 4 and 5 was employed. The decision thresholds were set manually at conservative levels which would minimise the likelihood of false alarms. It was not difficult to determine these levels. Typical values are of the order 0.8 for H and 0.3 for L.

6 VALIDATION

6.1 Detection of vehicle movement

An important part in interpreting coherence measurements over a short time interval was the method of presentation of the coherence data. A high resolution (2 m) optical image, obtained by the DD5 system, was used to provide background information through an intensity - hue - saturation display. In this display the optical image provided the intensity level and the coherence changed the hue (e.g. from dark to light blue with increasing coherence) at a fixed saturation level.

Given the general context to the imagery provided by the DD5 data it was immediately apparent that the images were highly coherent throughout except for regions of woodland and water (which would be known). Within urban areas regions of anomalously low coherence appeared. On examination of the display these areas were identified to be trees, and the large car parking areas associated with offices, shops, railway station etc.

Since the presence of trees within the scene could be known beforehand, this result indicates that vehicle activity can be monitored; at assembly points for example.

6.2 Construction and demolition

Validation of the use of coherence data from the 35 day repeat cycle for the detection of changes due to construction or demolition was based on the use of records maintained by DRA and the local government authorities. A fundamental problem with this data was however that it was seldom precise. For example events would be recorded by the date of planning application or contract issue and the date when site inspection confirmed that work had been completed.

The detection process described in Section 5.3 was employed using several selected data sets. The positions where change was predicted were overlaid onto the high resolution DD5 optical data. All places at which a change event was predicted were noted and the sources of ground truth checked for confirmatory information. It was found that a large proportion

of the changes detected could be confirmed, and several of the places for which there was no information were at locations where activity was likely. A few examples of the change events detected were the construction of a supermarket; the demolition of an army barracks; the modification of a hospital into an officer's mess; the clearance of houses in preparation for a new road and the removal of portable office accommodation.

These results are considered to show that the technique can be operated with low false alarms. Missed change events appear to be a function of object size and shape compared to the coherence measurement area, although both the orientation and the materials of construction may be additional factors.

The decision logic employed is suited to changes associated with large constructions. It is predicted that increased radar system resolution would reduce the effect of mixed pixel properties and allow changes to be detected in smaller buildings.

6.3 Advanced decision technique

A more elaborate decision strategy, based on fuzzy logic, has also been investigated (Ref. 5) for the detection of demolition or construction. In this method it was possible to use the coherence histogram data to form a continuous decision surface which represented the likelihood of construction or demolition. While this technique was judged to give the best results, it was only a marginal improvement over the simple threshold technique described here. It was also unclear how helpful likelihood information really is in interpreting change. The principal benefit of the advanced decision technique was that its use has shown that the simple decision techniques employed are justifiable.

7 CONCLUSIONS

Imagery of complex scenes obtained from low resolution spaceborne SAR systems is difficult to interpret. When the repeat orbits exhibit high stability SAR images are obtained from almost the same position in space. Under these conditions it has been shown that the coherence between two SAR images is an important feature with which to characterise the objects within a scene. It provides a measure of the degree of change between two observations. For monofrequency monopolarisation SAR systems (such as ERS-1) the availability of a new characteristic feature adds considerably to the potential for exploitation of data from the system.

Coherence has been studied using ERS-1 SAR imagery of Farnborough (Hampshire, UK) and countryside nearby. It has been shown that the signatures of fixed man made structures exhibit stability. Natural targets generally have unstable signatures. Trees exhibit low stability even over time intervals as short as 3 days. Agricultural scenes exhibit a range of stability dependent on the time of year and the crop-type. Generally all vegetation exhibits low stability after 35 days. Motor vehicles provide a significant changing component within scenes.

It was found necessary to use a coherence measure calculated over a region of approximately 50 m by 50 m. This is large compared to many man made objects within the scene, for example, domestic housing. Measurements made over areas such as these, therefore include objects with different coherence (e.g. house, cars and garden). These measurements consequently had mixed properties partly stable, partly unstable. In contrast, the DRA site for example is mostly characterised by large buildings and the measurements have distinctive high coherence properties. Change detection could therefore only be reliably performed for objects which are large enough to fill an area of approximately 50 m by 50 m. For higher resolution sensors than ERS-1 this area would be proportionally reduced.

Data obtained at intervals of 35 days (or longer) has been used to show that changes in the coherence properties between image pairs can be used to detect change. A high proportion of the detected changes could be validated against historical data. Some changes were missed but it is thought that this is due to the size of the objects being to small compared to the measurement window.

By the use of data obtained 3 days apart it was shown that areas with active vehicle movements (e.g. car parks) could be identified. This type of data should also be capable of revealing sudden changes to man made structures. Unfortunately historical records concerning redevelopment do not define the precise time at which demolition events occurred. These records were not intended for the purpose required by this study. It has therefore not been able to validate the detection of sudden changes.

The techniques developed could either be used within an automated processing system or they can be used for manual interpretation. For manual interpretation it was found that a display technique which superimposed the data onto optical imagery greatly increased its interpretability. Structures within the coherence data which are inconsistent with the optical imagery can lead to the identification of change (e.g. forest clearance, construction).

The two types of measurements, one based on short term and the other on long term coherence, complement each other. One is suited to rapidly changing operational activities (damage assessment, vehicle activity), and the other to longer term surveillance (construction, modification) for strategic or verification purposes. In both cases the information is likely to provide an important cueing mechanism for other higher resolution sensors.

8 REFERENCES

1. G A Solaas, 'SAR interferometry orbit listing', Vols1,2,3 ESRIN/ERS Missions section
2. C Prati and F Rocca, 'Limits to the resolution of elevation maps from stereo SAR images', Int. J Remote Sensing, 1990 Vol11, No 12, pp 2215-2235
3. F Gatelli et al., 'The wavenumber shift in SAR interferometry', IEEE Trans on Geoscience and Remote Sensing, Vol. 32, No. 4, July 1994, pp 855-865
4. H Zebker and J Villasenor, 'Decorrelation in interferometric radar echoes', IEEE Trans on Geoscience and Remote Sensing, Vol. 30, No. 5, September 1992, pp 950-959
5. D Corr, S Whitehouse, D Mott and J Baldwin, 'Automatic change detection in space borne SAR imagery', Proceedings 2757 of SPIE AeroSense '96.

Paper 39

M. McKean (CA)

Comment on the point regarding the use of complex imagery (coherence). I believe that it is important to use the contextual information that the speckle provides (wide-area multi look and lower resolution can provide an overview of roads and other features which provide additional information to that provided by complex data and products).

In the RADARSAT context processing of fine beam data to approximately 8 m and approximately 6 m single look complex can provide different contexts.

G. Brown (US)

Were you dealing with the amplitude and phase in your change detection work?

Author's reply:

Sorry, I should have said that the data used was complex.

La spatiocarte numérique Défense

M.N. SCLAFER, Centre Géographique Interarmées
16bis av Prieur de la Côte d'Or - 94114 ARCUEIL cdx - France
J.P. CANTOU, Ingénieur chargé des relations avec la Défense, IGN Espace
Parc technologique du Canal - 24 rue Hermès
31527 RAMONVILLE SAINT AGNE Cdx - France
Tél (33) 05 62 19 18 18 Fax (33) 05 61 75 03 17

SOMMAIRE

1. La spatiocarte à vocation civile

2. La filière spatiocarte

2.1. Les moyens

2.2 La chaîne de production d'IGN Espace

3. La spatiocarte numérique défense

3.1. Les principes

3.2. Les produits

3.3. Le processus de production

1. La spatiocarte à vocation civile

Dans le domaine civil la spatiocarte est née dans les années 80 du besoin qu'avaient les aménageurs de disposer pour leurs projets à l'export d'une cartographie à jour et bon marché, qualités inaccessibles par des moyens classiques de production de cartes.

Constituée d'un fond d'image rectifié dans la géométrie des cartes traditionnelles et complété de surcharges cartographiques, la spatiocarte a d'abord été éditée sous forme analogique et a constitué ainsi un des principaux vecteurs de promotion de l'imagerie acquise par les satellites SPOT. Leur résolution au sol jusqu'à aujourd'hui inégalée (10 m) convient parfaitement pour obtenir en quelques semaines une couverture aux échelles du 1:50 000 et 1:100 000 de vastes territoires mal cartographiés.

La demande croissant auprès des opérateurs commerciaux, Spot Image demande à l'IGN de créer en 1989 à Toulouse une unité de production industrielle de spatiocartes, sur la base de compétences et d'expérimentations développées à partir des données des satellites Landsat et SPOT. La production annuelle d'IGN Espace passe de 250 spatiocartes en 1989 à 1300 en 1993 et dans le même temps les produits de la cartographie spatiale se diversifient : confection de modèles numériques de terrain (MNT) à partir de couples stéréoscopiques SPOT, de documents images destinés à des relevés GPS, de séries superposables d'images multitemporelles pour le suivi et le contrôle des terres agricoles et enfin de spatiocartes numériques, en tant que source de données pour les systèmes d'information géographique (SIG).

L'USGS, organisme chargé de l'élaboration des cartes topographiques sur le territoire américain, utilise en routine l'imagerie SPOT pour mettre à jour les feuilles les moins urbanisées. Spot Image a défini en collaboration avec IGN Espace les spécifications du format GIS-Geospot sous lequel sont édités les produits SPOTView (spatiocartes numériques standard).

L'éclatement de la guerre du Golfe donne lieu à la première utilisation des spatiocartes dans un contexte militaire opérationnel : le Centre Géographique Interarmées fait réaliser par IGN Espace une série de 45 spatiocartes SPOT destinées aux forces alliées. Depuis lors le CGI s'est doté de capacités de production en interne et a défini avec l'IGN des spécifications décrivant un produit plus élaboré : la "spatiocarte numérique défense".

2. La filière spatiocarte

2.1. Les moyens

L'ensemble des opérations décrites en 2.2. s'effectue par ordinateur sur des données numériques (images raster, objets géographiques vecteurs ou textes).

La modélisation géométrique des images s'effectue à l'aide de points d'appui connus en coordonnées terrain : ces points proviennent soit de relevés terrain de type GPS, soit d'images anciennes de géométrie connue, soit plus couramment de cartes topographiques sous forme papier. La disponibilité d'une *cartothèque* et d'une *base de données de systèmes géodésiques* à vocation mondiale constitue donc un atout considérable pour couvrir à la demande n'importe quelle zone géographique.

Une *architecture informatique* de type client serveur en réseau local apparaît la plus appropriée pour répondre aux exigences de productions suivantes :

- forte capacité de production
- gestion multi-chantiers
- partage des données images entre plusieurs ateliers
- fonctions de calcul lourd (traitement d'image)

Paper presented at the AGARD SPP Symposium on "Remote Sensing: A Valuable Source of Information", held in Toulouse, France from 22-25 April 1996, and published in CP-582.

• fonction graphiques de visualisation, retouches, interprétation, dessin, édition nécessitant une interface opérateur ergonomique et dédiée
• moyens d'acquisition et d'édition de document papier grand format.Pour cela, IGN Espace s'est doté d'un ordinateur central de type DEC-Alpha assurant les tâches de calcul lourd et celles de serveur de données (50 Gigaoctets de disque) et développe ou acquiert sur ses 15 stations de travail de type MacIntosh des logiciels graphiques adaptés aux besoins de chaque atelier.

Enfin un scanneur et un restituteur grand format (A0) permettent en interne l'acquisition de données et l'édition des produits finis selon des standard de qualité cartographique.

Si les outils informatiques jouent un grand rôle dans l'obtention d'une forte automatisation de la production, la qualité des produits obtenus dépend avant tout de la *compétence d'opérateurs polyvalents* et aguerris en matière de cartographie (connaissance de la qualité des cartes, du modelé de terrain, ...), de géométrie des images, de photo-interprétation, de maniement des logiciels, autant de techniques nécessitant une formation initiale poussée et un entretien continu des connaissances au fil des évolutions technologiques (nouveaux capteurs, informatique).

2.2 La chaîne de production d'IGN Espace (cf. schéma joint)

Elle est constituée de 5 ateliers :

• un atelier de modélisation géométrique permettant le recalage de différentes imageries (SPOT, ERS et bientôt Radarsat, IRS 1C)

• un atelier d'acquisition du relief par corrélation automatique de couples stéréo SPOT

• un atelier d'assemblage, de découpage géographique et d'amé-lioration du fond d'image géocodé

• un atelier d'interprétation du fond image et de rédaction des surcharges cartographiques

• un atelier d'édition des produits sous forme graphique et numérique.

Comptant sur un effectif de 25 opérateurs répartis sur ces ateliers, IGN Espace dispose d'une capacité annuelle de production d'environ 1000 spatiocartes simples (image géocodée surchargée de toponymes et points cotés), 200 spatiocartes interprétées (avec réseaux, hydrographie, zones urbaines, etc) et 200 modèles numériques de terrain issus de la corrélation de couples stéréo SPOT.

3. La spatiocarte numérique défense

Le Centre Géographique Interarmées a souhaité se doter de cartographies numériques sur fond d'images spatiales de façon à répondre aux besoins des armées en cas de conflit sur des territoires mal cartographiées : visualisation du théâtre d'opérations, mesures de distances, d'altitudes, la préparation des missions est grandement facilitée par l'affichage sur écran des données cartographiques.

3.1. Les principes

L'équipement en spatiocartes de vastes zones géographiques suppose la disponibilité d'images acquises sur une période limitée. Il est rare de pouvoir trouver en archive une telle homogénéité. Seul le *système SPOT* permet le recours à la programmation de la prise de vue, qui présente de plus l'avantage d'offrir des segments d'images acquis en un seul passage du satellite. En effet la programmation en segments permet de réduire considérablement la quantité de points d'appui nécessaires, tout en rigidifiant la géométrie de l'ensemble. Il faut se rappeler que le contexte n'autorise *aucune opération de terrain*, ni détermination de points GPS, ni recueil d'informations : seules les meilleures cartes topographiques existantes sont utilisées, à la fois pour caler les images et comme source d'informations géographiques.

L'élaboration des spatiocartes est réalisée en 2 grandes étapes :

1) la constitution du fond d'image géocodé,
2) le report d'informations fournies par les cartes (JOG notamment) sur le fond d'image.
Cette deuxième étape est délicate : elle suppose la définition d'une légende, une sélection sur les cartes des objets correspondants, en respectant une certaine homogénéité sur la zone à couvrir, puis la rédaction sur fond d'écran des objets choisis.

3.2. Les produits

Deux types de produits sont confectionnés :

• des spatiocartes à l'échelle du 1:100 000 sur fond d'images SPOT panchromatiques (noir et blanc à résolution 10 m), couvrant l'ensemble de la zone demandée
• des spatiocartes à l'échelle du 1:50 000 sur fond d'images SPOT "P + XS" (couleur à résolution 10 m), couvrant les principales villes de la zone.

Ces spatiocartes sont éditées à la fois sous forme *graphique* (documents papier obtenus par impression offset en quadrichromie) en vue d'un usage sur le terrain et sous forme *numérique* (base de données image et vecteurs) comme support à la préparation de mission.

En numérique, le fond d'image est livré en format raster GIS-GeoSpot et les données vecteur organisées par couche d'information (textes, légende, routes, hydrographie, etc) sont livrées en format EDIGEO, en vue d'une intégration au format VMAP par le CGI.

La légende de ces spatiocartes est extraite de la nomenclature DIGEST de l'OTAN ; elle correspond essentiellement aux thèmes figurant sur les cartes JOG au 1:250 000, qui constituent la principale source d'information cartographique, et inclut une composante mise à jour résultant de l'interprétation d'objets nouveaux sur l'image mais inexistants lors de la fabrication de la carte de référence.

La précision du fond d'image, qui tient lieu de précision pour l'ensemble du jeu de données (raster et vecteur) en tant que support de dessin de ces données, dépend de la précision des cartes ayant fourni les points d'appui ainsi que de la disponibilité de MNT sur la zone. A titre indicatif, l'utilisation d'un jeu de cartes au 1:50 000 couvrant même partiellement une zone par ailleurs bien couverte par des segments de prise de vue et par un MNT de type DTED-Niveau 1 permet d'atteindre une précision de localisation de 20 m r.m.s..

3.3. Le processus de production

La *confection du fond d'image* repose sur les techniques suivantes :

- la spatiotriangulation de blocs d'images SPOT
- la corrélation automatique de couples SPOT pour générer si besoin des MNT plus précis que les MNT-DTED Niveau 1
- le mosaïquage d'images géocodées
- l'application de traitement d'amélioration de la lisibilité des fonds d'image.

Le *report d'informations cartographiques* sur ce fond suppose la définition préalable de règles d'interprétation et de rédaction par thème, à l'usage de l'équipe d'opérateurs (Ex : continuité des réseaux, y compris entre feuilles ; objets mis à jour ou non). Un travail de préparation sur papier avec les cartes et les images permet de sélectionner les objets cartographiques afin d'assurer une homogénéité du travail au sein de l'équipe. Le travail de rédaction proprement dit peut alors démarrer sur fond d'écran MacIntosh à l'aide d'un logiciel de dessin du commerce permettant la gestion d'images en pleine résolution.

Electronic Countermeasures Against Synthetic Aperture Radars

G. Krämer
FGAN - FHP
Neuenahrer Straße 20
D - 53343 Wachtberg
Germany

1. Summary

Airborne and spaceborne Synthetic Aperture Radars (SAR) are in use for remote sensing. To hide military targets from being detected by these systems, they might be attacked by Electronic Countermeasures. Since SAR systems are characterised by long-time coherent processing of echo signals with integration gains typically in the order of 80dB, coherent jammers using Digital Radio Frequency Memories (DRFM) can generate false targets and camouflage patterns with only moderate jamming powers. The influence of a single coherent deception target, largely exceeding the normal radar echo power received from one single resolution cell, and the generation of a deceptive scene, are studied with a simulation model of a generic airborne SAR.

2. Simulation model of generic airborne SAR

A computer simulation model of a generic airborne SAR has been developed, which models range resolution by a linear FM chirp impulse and performs range migration compensation. The radar is assumed to operate in the side-looking mode at the X-band frequency f_0 = 10GHz and is assumed to be carried on a platform with velocity 150m/s (540km/h or 292knots) on a straight and level course at an altitude of 5,000m.

A SAR image section consisting of 128×128 picture elements (pixels) is calculated from a corner reflector arrangement and the coherent jamming signal. Thermal noise and terrain returns are not included in the simulation. A list of the design parameters of the radar which are relevant to the simulation is given in Tab.1

Wavelength	λ	3cm
Range resolution	δ_r	1m
Azimuth resolution	δ_{az}	1m
minimum range	R_{min}	15km
maximum range	R_{max}	30km
Altitude	H	5,000m
Platform velocity	v_R	150m/s

Tab.1 Design parameters of generic airborne SAR

2.1 Antenna pattern

The antenna is modelled as a uniformly illuminated rectangular aperture. From the required azimuth resolution δ_{az} of the system it follows for the aperture length L in the azimuth direction /1,2,3,4/

$$L = 2 \cdot \delta_{az} \tag{2-1}$$

which gives L = 2m. The SAR antenna width W in elevation is determined by the radar wavelength λ and the elevation beamwidth θ_{el} as

$$W = \frac{\lambda}{\theta_{el}} \tag{2-2}$$

From the illuminated swath according to Fig.1, the elevation beamwidth follows

$$\theta_{el} = \arccos\frac{H}{R_{\max}} - \arccos\frac{H}{R_{\min}} \tag{2-3}$$

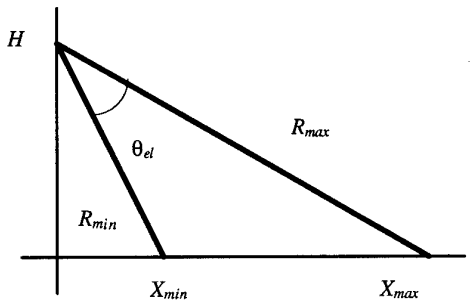


Fig.1 Illuminated swath

The gain G of the antenna with untapered rectangular aperture $A = L \times W$ is given by

$$G = 4\pi\frac{LW}{\lambda^2} \tag{2-4}$$

The antenna has a sinc-pattern, corresponding to the signal received from a point source in the far field at azimuth φ_{az} and elevation φ_{el}, of the shape

$$s(\varphi_{az},\varphi_{el}) = \sqrt{G} \cdot \mathrm{si}(\pi\frac{L}{\lambda}\cos\varphi_{el}\sin\varphi_{az})\,\mathrm{si}(\pi\frac{L}{\lambda}\sin\varphi_{el}) \tag{2-5}$$

where si(x) denotes the function sin(x)/x. If the look direction in azimuth and elevation is indicated by the unit vector

Paper presented at the AGARD SPP Symposium on "Remote Sensing: A Valuable Source of Information", held in Toulouse, France from 22-25 April 1996, and published in CP-582.

$$\underline{x}_{sE} = \begin{pmatrix} x_{sE} \\ y_{sE} \\ z_{sE} \end{pmatrix} = \begin{pmatrix} \cos\varphi_{el}\cos\varphi_{az} \\ \cos\varphi_{el}\sin\varphi_{az} \\ \sin\varphi_{el} \end{pmatrix} \quad (2\text{-}6)$$

the antenna pattern is given by

$$s(\varphi_{az},\varphi_{el}) = \sqrt{G}\,\mathrm{si}(\pi\tfrac{L}{\lambda}y_{sE})\mathrm{si}(\pi\tfrac{L}{\lambda}z_{sE}) \quad (2\text{-}7)$$

2.2 Range compression

The radar is assumed to use an FM chirp signal with base-band representation

$$s_{FM}(t) = \mathrm{rect}(\frac{t}{T})\exp(j2\pi\frac{f_H t^2}{2T}) \quad (2\text{-}8)$$

In Eq.(2-8), T indicates the pulse duration, rect(x) is the rectangular function with

$$\mathrm{rect}(x) = \begin{cases} 1 & \text{if } -\frac{1}{2} < x < \frac{1}{2} \\ 0 & \text{elsewhere} \end{cases} \quad (2\text{-}9)$$

and f_H denotes the frequency swing. The instantaneous frequency of the FM impulse, with φ denoting its phase angle, is defined as

$$f_{inst} = \frac{1}{2\pi}\cdot\frac{d\varphi}{dt} = f_H\cdot\frac{t}{T} \quad (2\text{-}10)$$

and is depicted in Fig.2.

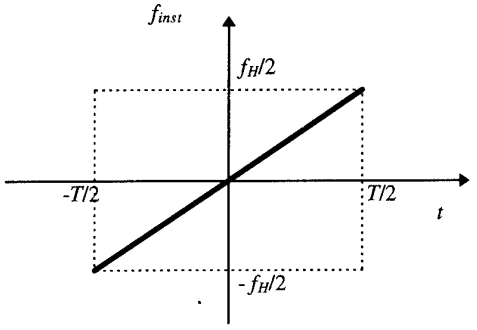


Fig.2 Instantaneous frequency of FM impulse

For the case $f_H T >> 1$, the range resolution δ_r of the FM impulse is given by /5/

$$\delta_r = \frac{c}{2f_H} = \frac{c}{2B} \quad (2\text{-}11)$$

with c denoting the velocity of light and $B = f_H$ its bandwidth. Since the required range resolution of the generic airborne SAR is δ_r = 1m according to Tab.1, from Eq.(2-11) the necessary chirp bandwidth follows as B = 150MHz.

The quantity

$$B_1 = B\cdot T \quad (2\text{-}12)$$

is known as the time-bandwidth-product of the FM impulse. Inserting this into Eq.(2-8), the FM chirp can be written as

$$s_{FM}(t) = \mathrm{rect}(\frac{t}{T})\exp(j\pi\cdot B_1(\frac{t}{T})^2) \quad (2\text{-}13)$$

As it is approximately a low-pass signal with cut-off frequency $f_H/2 = B/2$, it can be represented by its samples taken at time distance $\Delta t = 1/B$. The number of samples representing the signal in the interval T, where it is not equal to zero, then becomes equal to the time-bandwidth-product B_1. For the simulation it has been chosen B_1 = 512. Denoting by fix(x) the integer part of x, the digital representation of the FM impulse follows from Eq.(2-13)

$$s_{FM}(k) = \begin{cases} \exp(j\pi\frac{k^2}{B_1}) \\ \quad \text{if } -\mathrm{fix}(\frac{B_1}{2}) \le k \le -\mathrm{fix}(\frac{B_1}{2}) + B_1 - 1 \\ 0 \quad \text{elsewhere} \end{cases} \quad (2\text{-}14)$$

2.3 Derived SAR parameters and raw data generation

With the formulas presented above, all parameters of the generic airborne SAR relevant to the simulation may be calculated from the design parameters listed in Tab.1. The derived parameters are listed in Tab.2.

Antenna azimuth dimension	L	2m
Antenna elevation dimension	W	17.4cm
Antenna gain	G	37dB
Chirp bandwidth	B	150MHz
Sampling interval	$\Delta t = 2\delta_r/c$	6.67ns
Time-bandwidth-product	B_1	$2^9 = 512$
uncompressed pulse duration	$T = B_1/B$	3.41µs

Tab.2 Derived parameters of generic airborne SAR

Fig.3 indicates the co-ordinate system on which the calculations are based. The radar is assumed to move in the y,z-plane on a straight line parallel to the y-axis at an altitude H.

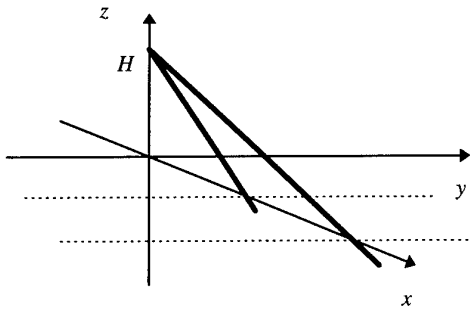


Fig.3 Side-looking airborne SAR

The SAR antenna is assumed to be tilted in elevation such that its boresight is directed to the centre of the illuminated swath, i.e. to the point $(x_{max} + x_{min})/2$ on the abscissa in Fig.1. All test reflectors and the deception jammer are located in the

x,y-plane and the radar is assumed to perform strip mapping in the side-looking mode.

To reduce the amount of data to be processed, the radar is modelled as operating at the minimum pulse repetition frequency necessary for a SAR, which corresponds to the Nyquist rate for the maximum Doppler frequency of the scattering centres scanned by its side-looking antenna beam and is given by /2/

$$f_{PRF\min} = \frac{v_R}{\delta_{az}} \tag{2-15}$$

For this case, the distance Δy travelled by the radar between successive pulses becomes

$$\Delta y = \frac{v_R}{f_{PRF\min}} = \delta_{az} \tag{2-16}$$

Samples are calculated in the simulation program at aircraft positions with y-co-ordinates increasing in steps Δy, the platform motion between pulses is not modelled.

In a practical radar, a higher pulse repetition frequency might be used, the SAR image resolution, however, would not be different from the simulated SAR, since resolution is determined by pulse bandwidth and synthetic aperture length.

For the collection of raw data, an array for complex samples is provided of size $(2\times B_1)\times(2\times N_1)$ with $N_1 = B_1 = 512$. The received radar signal is assumed to originate from a triangular arrangement of corner reflectors, extending over 25 azimuth resolution cells and 25 range resolution cells, and the signal of a coherent jammer. All signal sources are located in the x,y-plane. A SAR image section of $N_B\times N_B$ picture elements (pixels) with $N_B = 128$ is calculated.

With R_0 denoting the range corresponding to the centre of the SAR image section when it is on boresight, the azimuth extent of the antenna beam footprint is given by

$$B_K = R_0 \cdot \theta_{az} = R_0 \cdot \frac{\lambda}{2\delta_{az}} \tag{2-17}$$

Representing the integer part of x by fix(x), the selected number N_{int} of radar returns coherently integrated by the SAR processor for that range is

$$N_{\text{int}} = \text{fix}(\frac{B_K}{\Delta y}) \tag{2-18}$$

with Δy according to Eq.(2-16). As the SAR image section calculated is very small, one single reference function with N_{int} samples in the azimuth direction is sufficient for the correlation with the raw data.

To generate the raw data, the arrangement of signal sources in the x,y-plane is successively displaced in the y-direction in N_B+N_{int} steps of size Δy. At each step the contributions of all sources are calculated under consideration of the antenna pattern and their distances to the antenna phase centre. The reflected signals are weighted by the two-way antenna pattern, for the jamming signals only the one-way pattern is applied. From the source distances z_s, the corresponding carrier phase angles are determined by $(4\pi/\lambda)\cdot z_s$ and the time delays of the FM chirp signals are calculated. For the jamming signal, modulation and delay inside the jammer, as described below, are also taken into account.

Each line of the raw data array contains samples taken in range direction, consecutive lines correspond to the progressing displacement in azimuth.

2.4 SAR image generation

The approach chosen for the generation of the SAR image is the correlation of the raw data in the array of size $(2\times B_1)\times(2\times N_1)$ with a reference function including carrier phase history and range migration.

With the aircraft velocity v_R, the range $R(t)$ of a single scattering centre being at its position of closest approach at time $t = t_0$ at range R_0 is given by /2/

$$R(t) = \sqrt{R_0^2 + (v_R \cdot (t-t_0))^2} \approx R_0 + \frac{v_R^2 (t-t_0)^2}{2R_0} \tag{2-19}$$

Neglecting the platform movement during the signal propagation time, as $v_R << c$ (c denoting the velocity of light), an impulse transmitted at time t is received at time $t+2R(t)/c$. After conversion of the impulses received from the scattering centre into the baseband, this corresponds to a carrier phase shift by $(4\pi/\lambda)\cdot R(t)$ and a signal delay of $\tau = 2R(t)/c$. In the case of a high resolution SAR, range compression must not be performed before azimuth compression and independently of it, as this varying signal delay results in a migration of the echoes from a single scattering centre through the range cells /3/.

In the simulation model, as a first step, an FFT is performed on the individual lines of the raw data array, which contain the range samples at distinct azimuth positions. In the Fourier domain, the time shift of a signal corresponds to the multiplication of its FFT with a phasor term where the phase is linearly increasing with frequency.

Let $z_R(k_{az})$ denote the displacement of a scattering centre with respect to its minimum range, for some distinct azimuth position with index k_{az}, which is derived from the parabola

$$z_R(t) = \frac{v_R^2 (t-t_0)^2}{2R_0} \tag{2-20}$$

of Eq.(2-19). This displacement corresponds to a time shift $\tau_R(k_{az}) = 2z_R(k_{az})/c$ of the received FM impulse. The operation in the Fourier domain equivalent to an inverse time shift, is the multiplication of the $2\times B_1$ Fourier coefficients in the line k_{az}, at the positions k_r, $1\le k_r\le 2B_1$, with the numbers

$$X_1(k_{az},k_r) = \begin{cases} \exp(j2\pi \dfrac{k_r - 1}{2B_1} \cdot \dfrac{z_R(k_{az})}{\delta_r}), \\ \qquad \text{if } 1 \le k_r \le B_1 \\ \exp(j2\pi \dfrac{k_r - 1 - 2B_1}{2B_1} \cdot \dfrac{z_R(k_{az})}{\delta_r}), \\ \qquad \text{if } B_1 + 1 \le k_r \le 2B_1 \end{cases} \tag{2-21}$$

Since the raw data are samples of the radar baseband signal, the carrier phase shift accompanying the range delay must

also be taken into account. Therefore all Fourier coefficients in a line are multiplied with a second phasor

$$X_2(k_{az}) = \exp(j\frac{4\pi}{\lambda} \cdot z_R(k_{az})) \tag{2-22}$$

to compensate for the carrier phase shift.

In the small range extent of the SAR image section calculated, the differences of parabolas according to Eq.(2-20), corresponding to scattering centres with different minimum ranges, are negligible.

When the two multiplications have been performed in N_{int} consecutive lines, the Fourier coefficients corresponding to all scattering centres in one particular azimuth resolution cell are identical in all the lines, the lines can therefore be summed up as the first step of their coherent processing. The sum is then multiplied with the FFT of the reference FM chirp

$$s_{FMr}(t) = \begin{cases} \exp\left(-j\pi \dfrac{(\frac{3}{2}B_1 - 2 - k_r)^2}{B_1}\right), \\ \qquad \text{if } B_1 + 1 \le k_r \le 2B_1 \\ 0 \qquad \text{elsewhere} \end{cases} \tag{2-23}$$

and inverse FFT is performed, which completes the coherent processing of all scattering centres at the particular azimuth resolution cell.

To perform the same coherent processing for the scattering centres in a neighbouring azimuth resolution cell, the template defined by $X_1(k_{az},k_r)\cdot X_2(k_{az})$ has to be shifted by one line and multiplication and summation have to be carried out again as the first step. This is equivalent to a convolution, and this convolution is performed in the simulation program via FFT. After this convolution, each line of the array contains the Fourier transforms of the echoes from all scattering centres in one particular azimuth resolution cell. The final step is the multiplication with the FFT of the reference FM impulse and inverse FFT.

Due to the periodicity of convolutions by FFT, zero padding is performed. The range reference function therefore contains B_1 samples of the chirp and B_1 zeros. In the reference function used for the azimuth convolution, $N_{int} \le N_1$ samples are different from zero, the remaining $2\times N_1$-N_{int} array locations are filled with zeros.

The form of the SAR antenna pattern might be included in the azimuth reference function, which would in a real system lead to a slight increase of the signal-to-noise ratio. In the examples presented in this paper it is not included.

3. Reference scene

The reference scene consists of a triangular array of corner reflectors. Two legs are oriented in the x- and y-direction, respectively, of Fig.3, the third leg is the diagonal. The extent of the triangle in range and azimuth, respectively, is 25 metres. The lower left corner is at the position $x = X_0 = 20$km and $y = 0$. Four corner reflectors are arranged on each leg, in addition reflectors are located at the three corners.

A section of 128×128 pixels of a SAR image is calculated, corresponding to 128 range resolution cells by 128 azimuth resolution cells. The SAR image is displayed in range and azimuth, the usual projection into the x,y-plane has not been performed.

Fig.4 shows the SAR image of the reference scene. The brightness values are quantised with 8 bits, corresponding to 256 gray levels. In contrast to the usual display on a screen, the maximum echo power corresponds to black. The quantising interval is so chosen, that the echo amplitude obtained from one single corner reflector in the centre of a single resolution cell leads to about 70% of the maximum gray level. Apart from their range and azimuth sidelobes, reflectors in the centre of a resolution cell appear in only one pixel, reflectors at other positions generate contributions in neighbouring pixels.

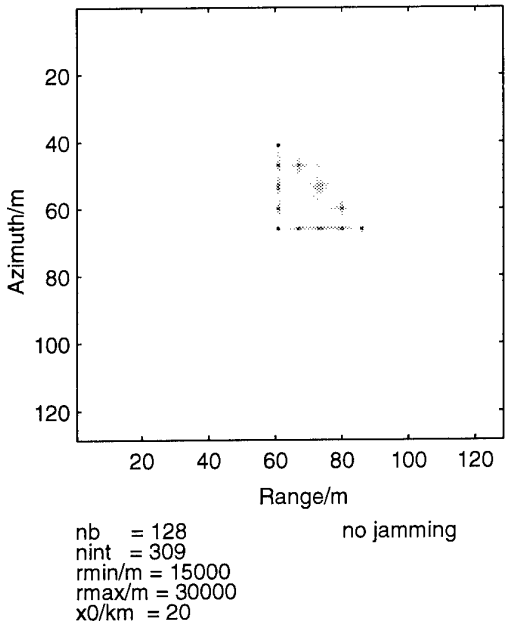


Fig.4 SAR image of reference scene

4. Coherent jamming

A deception jammer containing a Digital Radio Frequency Memory (DRFM) is assumed to receive the SAR impulses and to retransmit them with various time delays and phase modulations. The phase history of a straight-through repeated jammer signal is identical to the phase history of a real target echo from the jammer location, including all phase modulations due to deviations from the ideal flight path, and is therefore processed by the radar with its full processing gain.

By a time delay of the jamming signal, deceptive images can be generated displaced in range with respect to the true jammer location. For very large time delays, a mismatch occurs between the parabolic phase history corresponding to the true jammer location, given by Eq.(2-20) for a closest range R_0 of

the jammer, and the parabola of the reference function corresponding to the deceptive range, different from R_0. This mismatch leads to a minor defocusing in azimuth and may be neglected.

An azimuth displacement of a deceptive pixel with respect to the true jammer location may be generated by a phase shift of the retransmitted SAR impulses increasing from pulse to pulse, which will be investigated below. A jamming signal of this kind is similar to the echo of a moving SAR target with a radial velocity component, and might at least partly be suppressed by moving target processing algorithms.

4.1 High-power coherent jamming

It is assumed that the SAR uses a very high average transmitter power, P_{av}, corresponding to an average power density in the antenna aperture of 500W/m^2. With the antenna aperture of 2m×0.174m from Tab.2, this is equivalent to P_{av} = 174W.

With the SAR antenna gain G it follows from the radar equation that a coherent jammer at distance R to the radar would generate the same signal strength as a point target with radar cross-section σ^2, if its average effective radiated power $ERP_{j,av}$ was

$$ERP_{j,av} = \frac{P_{av} \cdot G}{4\pi R^2} \cdot \sigma^2 \qquad (4\text{-}1)$$

Assuming σ^2 = 0.1m^2, corresponding to a sphere of about 40cm diameter, with G = 37dB from Tab.2, it follows for R = R_{max} = 30km from Eq.(4-1) for the necessary jammer ERP $ERP_{j,av}$ = 7.7μW or -51dBW.

The jammer is assumed to transmit its deception signal with a rectangular horn antenna with an effective aperture A_{eff} = $a \cdot b$, with a and b denoting horizontal and vertical dimension, respectively, of the horn mouth. The gain of this jammer antenna is given by

$$G_j = 4\pi \frac{a \cdot b}{\lambda^2} \qquad (4\text{-}2)$$

The respective beamwidths in azimuth and elevation are given by

$$\theta_{az} = \frac{\lambda}{a} \quad , \quad \theta_{el} = \frac{\lambda}{b} \qquad (4\text{-}3)$$

With λ =3cm and a = b = 20cm, G_j = 27dB and $\theta_{az} = \theta_{el} = 8°$ are obtained.

If the jammer transmits a deception signal generating a false target in one single resolution cell, using an average transmitter power $P_{jam,av}$ = 100mW = -10dBW, this corresponds to an average jammer ERP

$$ERP_{jam,av} = P_{jam,av} \cdot G_j \qquad (4\text{-}4)$$

and it follows $ERP_{jam,av}$ = 17dBW. The received jamming signal would exceed the echo power received from the sphere in the above example by 68dB. With an average transmitter power of 1W, the difference would even be 78dB.

It should be noted that the average jammer transmitter powers correspond to the generation of a single false target in the SAR image. If false targets are to be generated having separations of more than the extent of one resolution cell, either in azimuth or in range, the corresponding jamming signal components are uncorrelated and the respective powers add up.

To show the effect of increasing jammer power, a jammer generating a single deceptive target has been placed at the position x = 20010.25m and y = -5.25m in the x,y-plane of Fig.3, which corresponds to a position below the lower leg of the triangular reference pattern in the SAR image. The jammer is assumed to operate in the straight-through repeater mode generating a single false target at its true location. The quantisation interval for the SAR image is not changed from that used in Fig.4, i.e. it is assumed that each of the corner reflectors used in the reference scene generates a signal in the picture element corresponding to its location of about 70% of the maximum gray level.

In Fig.5 it is assumed that the signal received from the coherent jammer exceeds the echo from one single corner reflector by $L1$ = 10dB. In the succeeding figures Fig.6 ... Fig.8 the jammer power level is increased corresponding to $L1$ = 40dB, 68dB and 78dB, respectively. It can be seen that the reference scene becomes increasingly masked by the range and azimuth sidelobes of the response on the coherent jammer.

It should be noted that, even for the highest jamming power assumed, there would be no saturation in the SAR processing, since either the thermal noise at the receiver input or the total received terrain return power, without presence of jamming signals, are in the order of 70dB greater in a SAR than the power received from the scattering centres in one single resolution cell /6/.

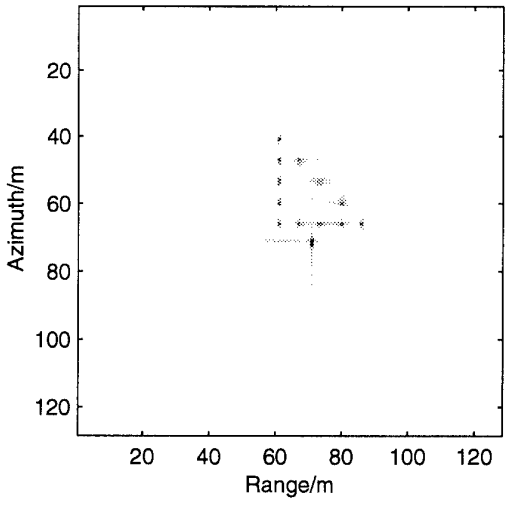
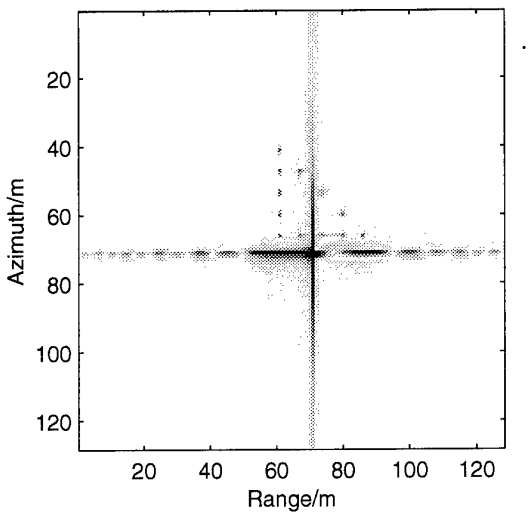
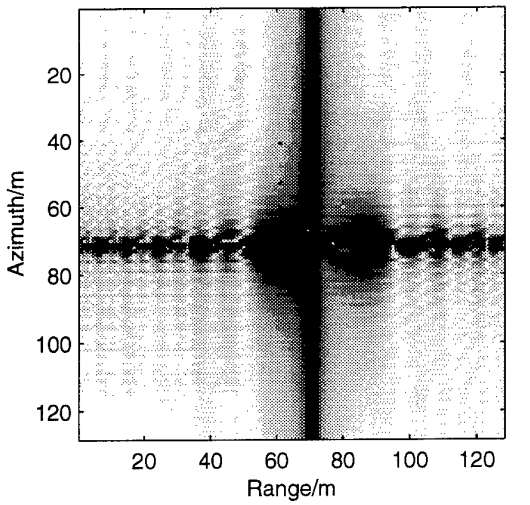


Fig.5 Straight-through repeater, $L1$ = 10dB



nb = 128. x0/km = 20.
nint = 309. L1/dB = 40.
rmin/m = 15000. jammer position:
rmax/m = 30000. [20010.25 , -5.2500]

Fig.6 Straight-through repeater, *L1* = 40dB



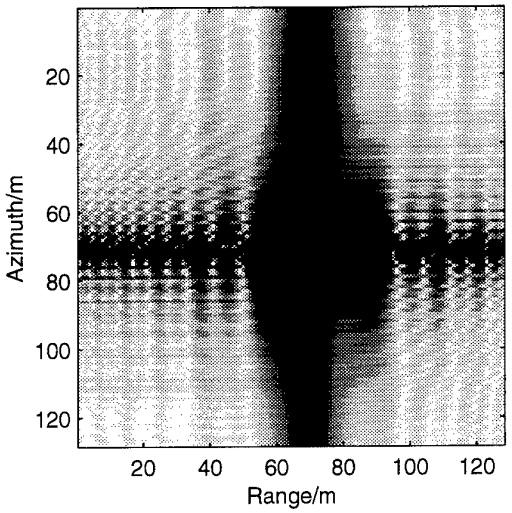
nb = 128. x0/km = 20.
nint = 309. L1/dB = 68.
rmin/m = 15000. jammer position:
rmax/m = 30000. [20010.25 , -5.2500]

Fig.7 Straight-through repeater, *L1* = 68dB

4.2 Deception scene

Deceptive pixels displaced in range from the true jammer location are generated by delayed transmission of the received SAR pulses, deceptive pixels displaced in azimuth may be generated by multiplying the samples of the received SAR pulses with a phasor, whose phase is incremented by a constant step from pulse to pulse.

The azimuth location of a pixel generated by a scattering centre is determined by the location of the apex of its associated range parabola according to Eq.(2-20). Let the phase history of the echo of a scattering centre, being at its position



nb = 128. x0/km = 20.
nint = 309. L1/dB = 78.
rmin/m = 15000. jammer position:
rmax/m = 30000. [20010.25 , -5.2500]

Fig.8 Straight-through repeater, *L1* = 78dB

of closest approach at a range R_0 at time $t = t_0$, be described by

$$\varphi(t) = \varphi_0 + \frac{4\pi}{\lambda} \cdot \frac{v_R^2 (t - t_0)^2}{2R_0} \tag{4-5}$$

The addition of a phase term linearly increasing with time gives

$$\varphi(t) = \varphi_0 + \frac{4\pi}{\lambda} \cdot \frac{v_R^2 (t - t_0)^2}{2R_0} + \alpha \cdot t \tag{4-6}$$

The apex of this parabola appears where its derivative is zero, thus

$$\begin{aligned} \varphi'(t) &= \frac{4\pi}{\lambda} \cdot \frac{v_R^2 (t - t_0)}{R_0} + \alpha \\ \varphi'(t_1) = 0 \quad &\Leftrightarrow \quad t_1 = t_0 - \frac{\lambda}{4\pi} \cdot \frac{R_0}{v_R^2} \cdot \alpha \end{aligned} \tag{4-7}$$

The time displacement

$$\Delta t = -\frac{\lambda}{4\pi} \cdot \frac{R_0}{v_R^2} \cdot \alpha \tag{4-8}$$

from the previous apex location at t_0 corresponds to an azimuth displacement

$$\Delta x = v_R \cdot \Delta t = -\frac{\lambda}{4\pi} \cdot \frac{R_0}{v_R} \cdot \alpha \tag{4-9}$$

Inversely, to generate an azimuth displacement by Δx, the value of α has to be chosen

$$\alpha = -\frac{4\pi}{\lambda} \cdot \frac{v_R}{R_0} \cdot \Delta x \tag{4-10}$$

In the simulation model, the minimum pulse repetition frequency of a SAR is used, according to Eq.(2-15), corresponding to the pulse repetition interval

$$T_{PRI} = \frac{1}{f_{PRF}} = \frac{\delta_{az}}{v_R} \tag{4-11}$$

The phase increment $\Delta\varphi$ between consecutive pulses, generating a deceptive target displaced in azimuth by Δx with respect to the true jammer location, therefore has to be

$$\Delta\varphi = \alpha \cdot T_{PRI} = -\frac{4\pi}{\lambda} \cdot \frac{\delta_{az}}{R_0} \cdot \Delta x \tag{4-12}$$

The deception scene is determined by a matrix $T(k_{az1},k_{r1})$ of size 20×20 shown in Tab.3. The matrix element in line k_{az1} and column k_{r1} indicates the amplitude of a deceptive pulse transmitted out of the DRFM, delayed in range by $(k_{r1}$-1) range resolution cells of size δ_r and phase modulated for displacement in azimuth by $(k_{az1}$-5) azimuth resolution cells of size δ_{az}. The letter Z indicates a random number with $10^{-2} < Z < 10^{-1}$, generated separately for each matrix location where it appears. The element $T(5,1)$, marked in boldface, corresponds to the straight-through repeated SAR pulses. Elements in line 5 correspond to the pulses with delay only, elements in other lines generate deceptive targets with azimuth displacement, for the elements outside of column 1 accompanied by range displacement as well.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z
2	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z
3	1	1	1	1	1	1	1	1	1	1	0	0	0	Z	Z	Z	Z	Z	Z	Z
4	1	1	1	1	1	1	1	1	1	1	0	0	0	Z	Z	Z	Z	Z	Z	Z
5	**1**	1	1	1	1	1	1	1	1	1	0	0	0	Z	Z	Z	Z	Z	Z	Z
6	1	1	1	0	0	0	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z
7	1	1	1	0	0	0	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z
8	1	1	1	0	0	0	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z
9	1	1	1	0	0	0	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z
10	1	1	1	0	0	0	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z
11	1	1	0	0	0	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z
12	1	1	0	0	0	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z
13	1	1	0	0	0	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z
14	1	1	0	0	0	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z
15	1	1	1	1	0	0	0	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z
16	1	1	1	1	0	0	0	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z
17	1	1	1	1	0	0	0	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z
18	1	1	1	1	0	0	0	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z
19	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z
20	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z

Tab.3 Definition of deception scene

The deception scene might represent a building complex with a shadow zone behind it, generated by the zeros in the matrix, surrounded by terrain with lower reflectivity, generated by the random numbers. The simulated SAR image of the deceptive scene, together with the reference scene from the corner reflector arrangement, is shown in Fig.9. In order not to be easily distinguished from real targets, the powers of the deceptive targets, corresponding to a weighting with one in the matrix, exceed the power received from a corner reflector by only $L1$ = 5dB.

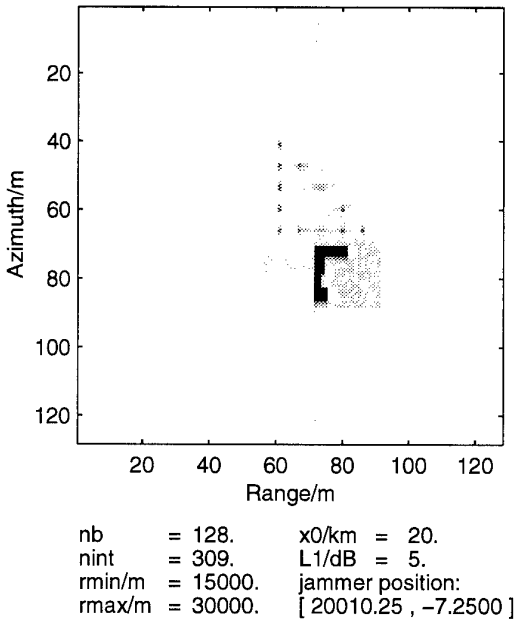


Fig.9 Simulated SAR image with deceptive scene

Whereas a real scene is ground fixed, the flight path of the SAR carrier platform determines the orientation of the deceptive scene. Delayed deception pulses generate false targets with range displacement, phase modulation leads to azimuth displacement. To match a deceptive scene to a real scene therefore requires the exact knowledge of the SAR carrier flightpath by the jammer.

5. Acknowledgement

The author wishes to thank Mrs. Coletta Schumacher for writing the simulation program in MATLAB.

6. References

1. Skolnik, M. I., "Radar Handbook, second edition", McGraw-Hill, Inc. (ISBN 0-07-057913-X).

2. Stimson, G. W., "Introduction to Airborne Radar", Hughes Aircraft Company, El Segundo, California.

3. Wehner, D. R., "High Resolution Radar", Artech House, Inc., (ISBN 0-89006-194-7).

4. Carrara, W. G., Goodman, R. S., Majewski, R. M., "Spotlight Synthetic Aperture Radar", Artech House, Inc. (ISBN 0-89006-728-7).

5. Cook, C. E., Bernfeld, M., "Radar Signals", Academic Press Inc., New York, London, 1967 (Library of Congress Catalog Card Number: 66-30803).

6. Krämer,G., "Influence of Electronic Countermeasures on Synthetic Aperture Radars", paper 14, Proceedings of the AGARD SPP Symposium "Environmental Factors in Electronic Warfare related to Aerospace Systems", Pratica di Mare (Rome), Italy, 8-11 May 1995.

SAR measurements of extended targets at 94 GHz

H. Schimpf, H. Essen

FGAN-Forschungsinstitut für Hochfrequenzphysik
Neuenahrer Str. 20
D-53343 Wachtberg-Werthhoven, Germany

Abstract

Since the beginning of 1995, FGAN–FHP is flying their polarimetric CORA–94 radar in a side-looking SAR configuration. The combination of Doppler processing and high range resolution via frequency agility provides a comparable resolution in cross range and range.

In the first part of the paper, some fundamental issues like depth of focus, range migration and speckle statistics are addressed and compared to results at the more commonly used longer wavelengths. It is found that due to the short wavelength some of the processing requirements are easier to fulfill at 94 GHz than at X–band, for example.

In the second part, several ways are discussed how to use the polarimetric information within each pixel for the detection and discrimination of extended targets. A comparison is made between a polarimetric whitening filter, a selective Stokes vector averaging and other methods that may help to obtain a stable description of the polarimetric state of the signals to be processed subsequently. The main effort is aimed at discriminating extended man-made objects from a natural clutter background. A scheme is developed that combines an amplitude based thresholding prescreener, to select candidate targets, and a feature based discrimination stage which provides the final decision. The example of two bridges crossing a river is used to demonstrate preliminary results.

1. Introduction

In recent years, radar imaging has gained more and more importance for military applications. Typical examples are reconnaissance and surveillance, but also in dual mode seekers, where a radar sensor is combined with an IR sensor, radar images have to be generated to be compared or matched with the IR images. The general task for the sensor is to search a certain area of expectance, which is known via a priori information, to identify autonomously certain extended structures within that area and eventually define an adequate aimpoint on that target.

In former years, radar maps were generated using real beam resolution, in some cases combined with high range resolution. Depending on the scanning geometry and on the available processor power, also the cross range resolution to a certain degree could be improved by applying Doppler beam sharpening techniques. But only with the introduction of synthetic aperture processing it became possible to achieve a pixel resolution that is comparable to IR and even photographic images. Also, this high resolution makes it more promising that the polarimetric information provided by the radar sensor can be used to enhance traditional imaging methods.

Due to its good all weather capability and above all due to the miniaturization requirements in seeker applications, millimeter wave radar is of great importance to the tasks described above. Therefore, it appeared only natural to FGAN–FHP to fly their coherent, polarimetric high range resolution radar CORA-94, operating at a nominal frequency of 94 GHz, in a side looking SAR configuration. Another advantage of millimeter wave frequencies is, as will be shown in the following, that the requirements for range walk and phase compensation are easy to fulfill and facilitate the SAR processing considerably.

In section 2, some general considerations pertaining to millimeter wave SAR will be described. It follows a short description of the CORA-94 radar, the measurement geometry, and the image generation procedures that were used. In section 4, several different methods of preprocessing are outlined that serve to create either amplitude or polarimetric images, or that help to reduce the image speckle. Finally, a short description is given of a possible detection scheme to extract extended targets from their natural clutter background, and a few preliminary results are shown.

2. Millimeter wave SAR

2.1 Range walk

The range walk is defined as the difference δR between the maximum range and the minimum range from the antenna to any point on the ground during the processing time ([2], p.287). If we choose the processing interval to be the time needed by the the aircraft to cover half the diameter of the antenna beam width Θ at the nominal slant range R, namely $R\Theta/2$, then it holds $\delta R = R\Theta^2/8$. The requirement is that $\delta R < \frac{1}{4}\Delta R$ where ΔR is the extent of the range resolution cell. Now we set $\Theta \approx \lambda/D_{antenna}$ and $D_{antenna} \approx 2\delta x$ where δx is the maximum obtainable cross range resolution. If one inserts these

Paper presented at the AGARD SPP Symposium on "Remote Sensing: A Valuable Source of Information", held in Toulouse, France from 22-25 April 1996, and published in CP-582.

expressions in the above inequality one finally gets

$$\left(\frac{\delta x}{\lambda}\right)^2 > \frac{R}{8\Delta R}$$

With typical values of $\delta x = 0.15m$, R=1000m and $\Delta R = 1.00m$ the condition becomes $2060 > 125$ and hence is easily fulfilled at 94 GHz.

2.2 Depth of focus

The depth of focus (DOF) is the accuracy with which a given quadratic phase reference function must be matched to the slant range under consideration ([3], p.1-4). It defines the number of different reference functions required for the complete swath that is processed. The requirement for the DOF is that the maximum phase error allowed at the end of the aperture be less than $\pi/4$ (corresponding to $\lambda/8$).

The linear motion of the antenna gives rise to a quadratic phase function. If we denominate by $L = v \cdot T_p$ the distance travelled during the processing time T_p then the two-way phase change with respect to the nominal slant range R_0 is given by

$$\Delta\Phi_0 = 2\pi\frac{(L/2)^2}{\lambda R_0}$$

Accordingly, the phase change at a slant range larger than R_0 by a distance 'D' is

$$\Delta\Phi_1 = 2\pi\frac{(L/2)^2}{\lambda(R_0 + D)}$$

Now, the largest possible change of slant range 'D' that still fulfills the condition

$$|\Delta\Phi_0 - \Delta\Phi_1| < \frac{\pi}{4}$$

is called the 'depth of focus'. This means that the phases of all point targets contained in a swath between 'R_0' and '$R_0 + D$' can be compensated by means of the same quadratic phase function.

Solving the above condition for 'D' results in

$$D < \frac{\lambda R_0^2}{2L^2}$$

If one substitutes for the cross range resolution $\delta x = \lambda R_0/2L$ then the DOF requirement can be stated as

$$D < \frac{2(\delta x)^2}{\lambda}$$

which becomes $D < 627m$ for $\delta x = 1.00m$. This means that the complete maximum swath width of 480 m can be processed with only one quadratic phase reference function. This is a considerable simplification as compared with longer wavelengths.

2.3 Speckle behaviour

In order to assess the speckle behaviour at 94 GHz, eleven more or less homogeneous subareas representing four different background types, namely forested area, grassland, water and shadow, were defined within the two bridge image. The size of these subareas varied between 250 and 2600 pixels. Three different types of processing were compared, namely nine-look, Stokes averaging and single-look, as described in the following. As a measure to describe the speckle quantitatively, the ratio

$$\gamma = \frac{\sigma}{\mu}$$

was chosen as described by Novak et. al. [4],[6]. In this expression, σ means the standard deviation and μ the mean value of the pixel power intensities (in linear units). For an ideally homogeneous area without any spatial variation one expects a Rayleigh distribution of the return amplitudes, and consequently γ=1 if no integration is applied. This value is found in the following table for "shadow, single-look, LL or LR". In most other cases, one observes rather pronounced deviations that are due to averaging or to a spatial variation of the Rayleigh mean.

		trees	grass	water	shadow
9-look	LL	.61	.44	.44	.45
	LR	.78	.50	.43	.39
	PWF	.60	.375	.33	.32
1-look	LL	1.10	.99	1.02	1.004
	LR	1.28	1.09	.995	1.005
	PWF	.91	.77	.76	.75
'Stokes'	LL	1.03	.97	.85	.87
	LR	1.01	.745	.71	.695
	PWF	.77	.59	.50	.505

Tab.1: measure of speckle σ/μ

In this table, LL means left-hand circular polarization on transmit and receive ('even'), LR means LHC on transmit and RHC on receive ('odd'), PWF means the application of the polarimetric whitening filter. There is a clear reduction of the amount of speckle for nine-look as compared to single-look processing. On the other hand, there is no clear tendency whether the 'odd' or the 'even' channel shows stronger speckle. This seems to depend also on the surface type. If one compares the values of table 1 with those given in [4](table 2) for the Lincoln Laboratory ADTS 35 GHz SAR data, one finds rather good agreement for grass and shadows, lower values for trees. From this one might conclude that the speckle behaviour is rather similar for 35 GHz and 94 GHz. However, one has to be careful with a final judgement, as the ADTS values are for linear (HH and HV) instead of circular polarization. Moreover, one finds that γ varies strongly across the image and depends on the size of the subarea chosen as well as on its homogeneity. This topic deserves more in-depth analysis.

3. Experimental Set-Up

For the measurements discussed here, the coherent, frequency agile 94 GHz-experimental Radar CORA-94 was mounted into a cargo aircraft C-160 in side looking geometry. The radar transmits

a fixed polarization (in this case left hand circular, or LHC), it is equipped with a dual channel receiver to register simultaneously the parallel and orthogonal polarization component of the backscattered signal. A detailed description can be found in [1]. A stepped frequency wave form was used with a bandwidth of 160 MHz resulting in a high range resolution of about 1.0 meter. In principle, the radial resolution could be brought down to 0.40 meter, but the data rate was limited by the data acquisition electronics and the capacity of the Ampex DCRSi recorder. The nominal depression angle was 22°, at a slant range of 800 m. The 3-dB two-way antenna beamwidth was 1° in azimuth and 12° in elevation.

The maximum possible swath width that could be obtained with the parameters given is 480 m. However, as the antenna was not mounted on a stabilized platform and there was no on-line range tracker implemented, only an effective swath width of 360 m (48 range bins with a spacing of 7.50 m) was processed.

3.1 Image generation

In a first step the radar data undergo off-line range track and calibration. The range track is done by means of a low-pass filtered sliding window and is part of the general motion compensation procedure that makes also use of the aircraft pitch, roll and crab angles. The calibration is based on pre- and postflight staring measurements against trihedral and dihedral precision reflectors which are mounted on high wooden poles to avoid multipath effects. After calibration, within each range gate the high range resolution (HRR) profiles are created via inverse 16-point FFT and combined to one large HRR profile that extends over the complete swath width that is created by the 12° elevation beamwidth of the lens antenna.

Next, a time sequence of 64 consecutive HRR profiles is combined for the purpose of Doppler processing. Before the Doppler FFT, however, Hamming weight factors are applied, and the quadratic phase reference compensation based upon a low-pass filtered value of the crab angle (Doppler drift) is performed to focus the image. The so obtained new Doppler profiles are then used to update the crab angle estimate. This procedure plays an important role because crab angles up to 10° have been observed during unfavorable wind conditions.

The effective PRF per frequency step was $2060 s^{-1}$, thus, taking into account the aircraft velocity of v=$65 ms^{-1}$ and the azimuthal diameter of the antenna footprint at the nominal slant range, a total processing time of about 0.25 sec corresponding to 512 pulses, would have been available per Doppler FFT. However, in order to avoid motion compensation within a processing interval, the FFT length was limited to 64, and one 512-point FFT was replaced by 15 64-point FFTs with an overlap of 32 pulses (corresponding to 1 m flight path), each. Out of each Doppler profile, only the middle 9 values belonging to the beam center were retained. The range-Doppler pixels were then projected on a 1 m by 1 m x-y-grid to maintain equal resolution in range and cross range ("square pixels"). The whole procedure described so far is performed simultaneously for both orthogonal receive channels in order to generate radar images for co- and cross-polarisation. Finally , one gets an average number of 9 polarimetric signal values per resolution cell that can either be amplitude averaged to provide a 9-look image or be subjected to some polarimetric processing as described below.

The unambiguous range R_u of any FFT based on the PRF of $2060 s^{-1}$ (for each of the 16 transmit frequencies) is $R_u = \frac{PRF \cdot \lambda R}{2v} \approx 40 m$ which corresponds approximately to the azimuth beam diameter between the first nulls. In this way one avoids that scatterers from outside the 3 dB beamwidth are folded back into the FFT range thus limiting the dynamic range of the image.

4. Image preprocessing

The radar sensor does not only allow the generation of images with different resolution and different degrees of integration, but also provides several channels of independent information, namely two orthogonal amplitude channels plus the two pertinent absolute phases, Therefore it depends on the type of further processing which type of image one needs as input.

Four different ways of image preprocessing were analyzed and are briefly described in the following. The first one is the conventional multilook processing where the amplitudes of all 9 signals falling into one final resolution cell are integrated incoherently ("nine-look"). The result are two amplitude images belonging to the orthogonal receive channels, and containing no polarimetric information besides the splitting into two channels. In the case of circular transmit polarization, though, which was used by CORA 94, this is already a valuable polarimetric information, because the orthogonal receive channels correspond to the "odd-bounce" (flat plate, trihedral corner) and the "even-bounce" (dihedral corner) scattering mechanism, respectively.

If one needs the full polarimetric information for the purpose of discrimination and classification, however, then a different kind of preprocessing is required. The simplest way is to ascribe a weighting factor to each of the 9 signals within each resolution cell and to retain only that signal which originates closest to the boresight position of the antenna. One gets a fully polarimetric single-look image with the disadvantage of having given away the integration gain that would result from reduced resolution. This type of image was considered mostly for comparison.

The third method is a compromise between the first two methods. For each signal, its Stokes' vector (position on the Poincaré-sphere) $\vec{S}$=Q,U,V is computed which in a circular polarization basis can be written as

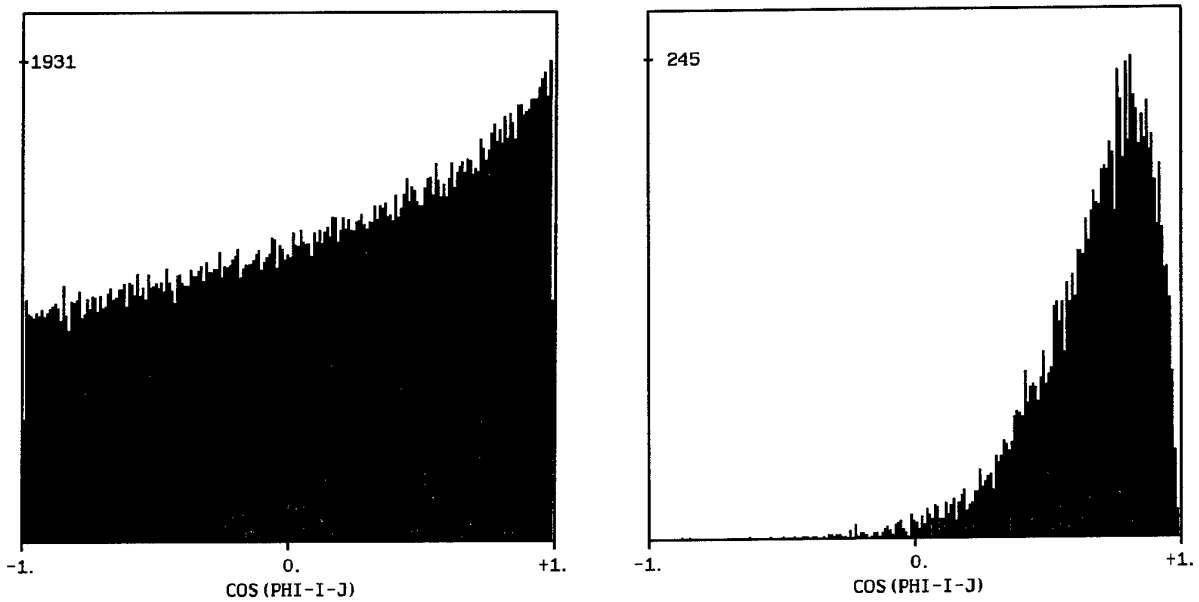


Fig. 1. Cosines of relative angles between Stokes' vectors, left: all 9 vectors, right: 3 closest vectors.

$$\vec{S} = \begin{pmatrix} Q \\ U \\ V \end{pmatrix} = \begin{pmatrix} 2E_L E_R \cos\delta' \\ 2E_L E_R \sin\delta' \\ E_L^2 - E_R^2 \end{pmatrix}$$

(in this definition we follow [12]), corresponding to a backscattered wave vector $\vec{E}$:

$$\vec{E} = \begin{pmatrix} E_r \\ E_l \end{pmatrix} = \begin{pmatrix} E_R e^{i\omega t} \\ E_L e^{i(\omega t + \delta')} \end{pmatrix}$$

Next, a number of 'n' (here, n=3 was chosen) such vectors is selected such that the relative angles between them are minimal, i. e. one selects those n out of 9 Stokes' vectors that lie closest to each other in their vector space. The Q-, U- and V-components of these n vectors are now averaged and, if necessary, transformed back to the quadrature components 'I' and 'Q'. The effect of this selection can be seen in fig. 1. It shows histograms of the cosines of the relative angles $\cos\phi_{ij} = \vec{S_i} \cdot \vec{S_j} / |\vec{S_i}||\vec{S_j}|$. The left histogram is for all nine vectors within each resolution cell whereas the right one pertains to the 3 closest Stokes' vectors. One recognizes that the distribution changes from near uniform to one with a pronounced peak near 35 degrees. The result is an image with a certain integration gain of between 3 dB and 5 dB which nevertheless contains some meaningful polarimetric information that can be used either for target/clutter contrast enhancement or for the construction of discrimination features. Fig. 2 shows a comparison between 9-look, single-look and Stokes averaging for the two bridge scene (LL channel). One sees that the Stokes averaged image shows rather good speckle reduction and resembles visually more the 9-look than the single-look case.

Another preprocessing method that can be combined with either one of the aforementioned methods is the so called "Polarimetric Whitening Filter" for the reduction of image speckle. As the data gathered with CORA 94 do not comprise the complete scattering matrix, the expressions derived in [4], [5],[7] had to be simplified. The form actually used for the "whitened" pixel intensity is:

$$y = |LR|^2 + \epsilon |LL|^2$$

where $|LR|$ and $|LL|$, respectively, are the amplitudes in the two preprocessed orthogonal channels, and $\epsilon = \frac{E(|LR|^2)}{E(|LL|^2)}$ is the ratio between the main diagonal elements of the clutter covariance matrix. It turned out that for forested areas as well as for grassy land ϵ was close to 2.5, therefore this value was used throughout.

5. Target detection

The radar based identification of extended man-made structures embedded in a natural clutter background consists of several steps. First, the radar image layout has to be created by determining the pixel (resolution cell) size and performing the projection to a ground map. Next, the desired information contents of each resolution cell has to be chosen and the preprocessing done accordingly. The third step would be a prescreener,

i. e. some sort of a thresholding procedure which narrows down the possible candidate target sites and thus reduces the computational burden on the subsequent processing stages. Based upon the output of the prescreener, feature vectors would then be constructed that help to discriminate between man-made targets and natural background.

For the prescreener and the discriminator, only single range profiles or a small number of range profiles must be used. This takes into account the fact, that in realistic systems, the incoming radar backscatter signals form a time sequence. That means that each detection algorithm has to start its work while the radar image is still being built up. The final product of the overall radar processing is an image that contains within each pixel the decision result "target yes/no". This image can subsequently be used, probably in connection with an original radar amplitude image, for further image processing techniques like pattern matching, comparison with photographical or IR images etc.

5.1 Prescreener

The prescreener used here as an example works on single range profiles with an extent of 280 cells. First, the power (odd, even, odd+even, with or without PWF) is summed over all 280 cells: P_{280}. Next, an area of interest is defined, the size of which is adapted to the target under consideration. Here, a size of 20 cells (i. e. 20 m range extent) was chosen which seems to be adequate in the case of a bridge that runs perpendicular to the line of sight. This area of interest in the form of a sliding window is now moved along the range profile, and the total power within each window is determined: P_{20}. The area is declared a potential target site when the threshold condition

$$P_{20} \geq \kappa \cdot \frac{20}{280} P_{280}$$

is fulfilled, i. e. when return power within the window exceeds the average expected power by a certain factor. Here, $\kappa = 1.5$ was chosen.

5.2 Discrimination

For discrimination between man-made structures and natural clutter background, four features were used. Only those cells that had passed the prescreener were subjected to the feature test. The features are:

$$F_1 = \frac{Std.dev.}{mean}(odd + even\,power)_{w=20}$$

$$F_2 = \frac{\sum_{3\,strongest\,lines}(odd + even\,power)_{w=20}}{\sum_{i=1}^{w}}$$

$$F_3 = Std.dev.(even\,power|dB)_{w=20}$$

$$F_4 = \frac{\ln(20) - \ln(M_2)}{\ln(2)} \cdot (20\ strongest\ even\ power\ pixels)_{16\times 20}$$

F_1 and F_2 were already used in [10] with good success, F_3 is explained in [8],[9], all three are calculated on windows with a size of w=20 range cells by 1 Doppler cell. F_4 also was introduced by Novak et al ([8], it is the "fractal dimension" of the area formed by the 20 strongest cells within a field under test (20 range cells by 16 Doppler cells), M_2 is the number of 2x2-boxes that are needed to cover all 20 high even power cells. The idea is, that in the case of clutter, these cells tend to be more isolated (i. e. fractal dimension closer to zero), whereas for artificial structures, they should show a stronger tendency to be connected to each other (fractal dimension between 1 and 2). To compute F_4, 16 consecutive range profiles have to be grouped together, which is no problem for real-time processing, as 16 usually is less than one Doppler-FFT block.

For the feature test, one needs references for clutter and for man-made targets, respectively. These were obtained from an independent scene containing the FGAN establishment (various buildings, antennas, cars etc.) surrounded by fields and wooded areas. It is clear that this reference for man-made objects is only a very rough approach because it does not even try to come close to the specific structures of the two bridges. Nevertheless it seemed interesting to analyze how far one can get with such weak assumptions.

The references are stored in the form of mean value and standard deviation of the respective feature, based on the feature distributions for the reference areas. One then computes as described in [11] the distances of the feature vector under test to the target reference and to the clutter reference:

$$D_T = \sqrt{\sum \frac{(F_i - F_{Ti})^2}{\sigma_{Ti}^2}}$$

$$D_C = \sqrt{\sum \frac{(F_i - F_{Ci})^2}{\sigma_{Ci}^2}}$$

where F_{Ti} and F_{Ci} are the mean values and σ_{Ti} and σ_{Ci} are the standard deviations of feature F_i (i=1,2,3,4) from the target and clutter reference distributions. The sum sign in these expressions can mean all four features or any subset therefrom.

Now, any pixel under test is said to belong to the class of man-made structures if the condition

$$D_T < \gamma \cdot D_C$$

is fulfilled where the factor $\gamma < 1$ has to be chosen so as to minimize the number of false alarms while not affecting the detection of the real target.

6. Results

As an example for the performance of the proposed detection scheme a scene was selected which contains two bridges crossing a river (fig. 2). The lower bridge (as seen in the image) is a modern highway

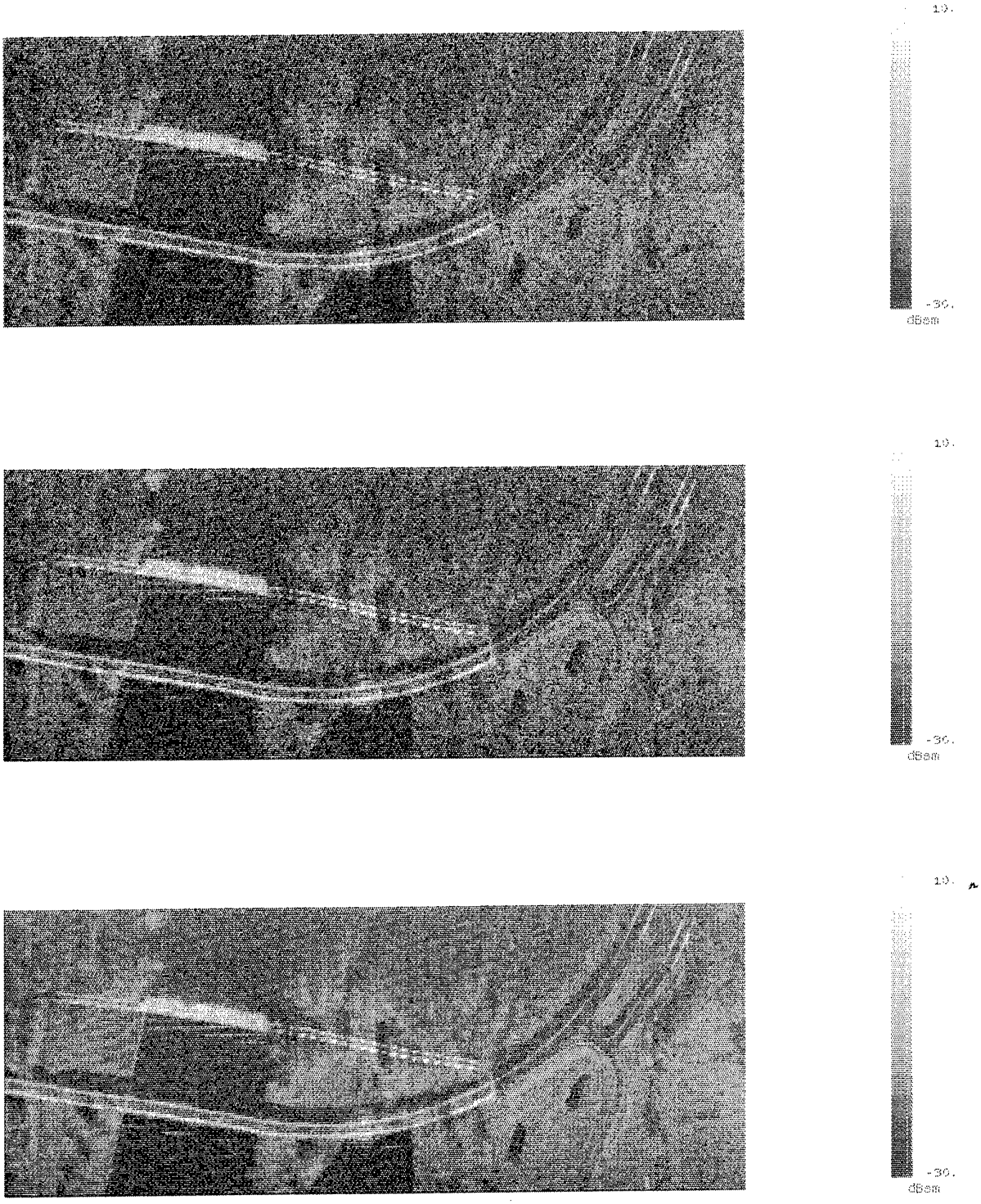


Fig. 2. Comparison between single-look (top), selective Stokes averaging (center), and 9-look image (bottom), polarization LL.

bridge made of concrete, with metal handrails accompanying the pedestrian walkways on both sides. After crossing the water the bridge continues for several hundred meters crossing some marshland before the terrain rises to the level of the road. The upper (old) bridge has only half the width of the

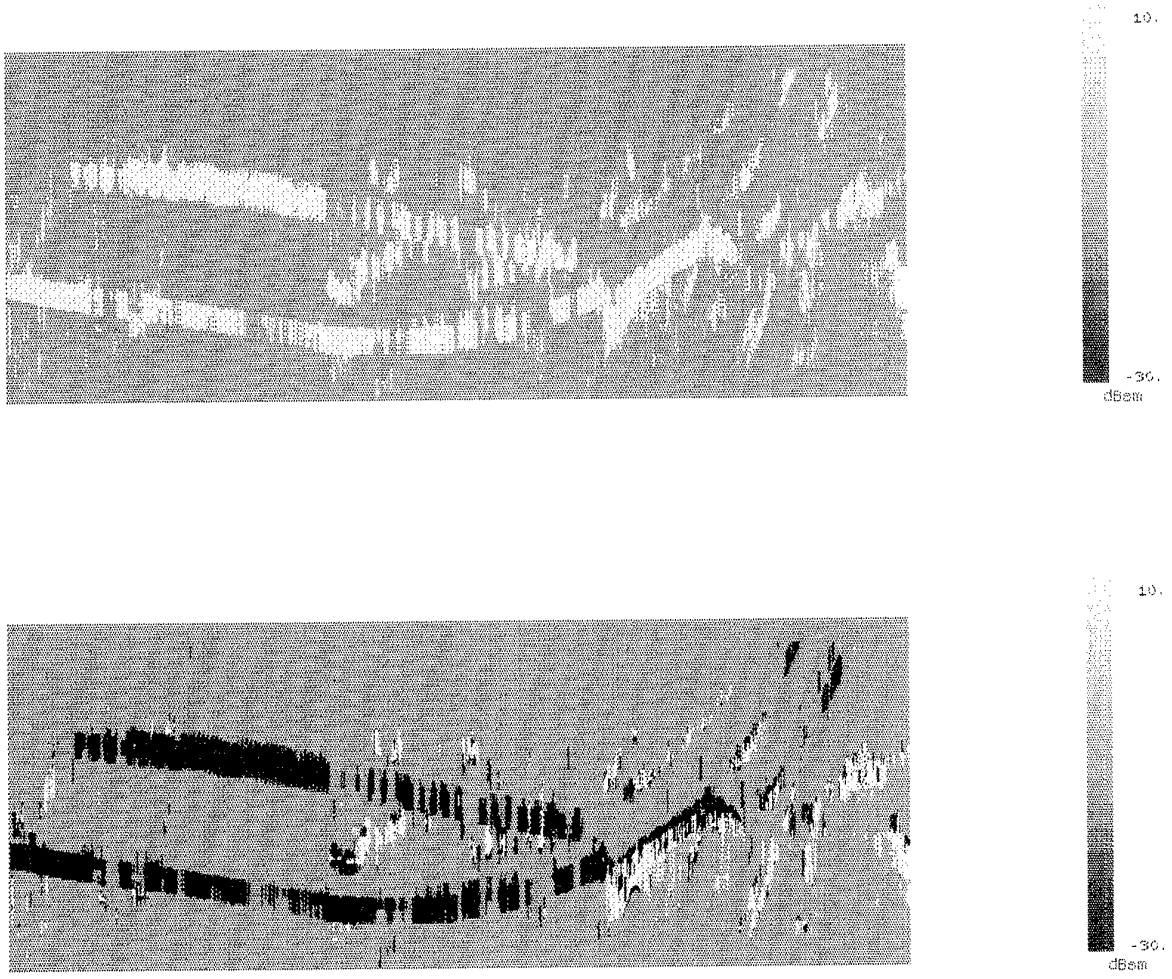


Fig. 3. Result of prescreener (top) and subsequent discrimination with feature F_3 (bottom) for Stokes averaged image without PWF.

modern one, it consists of a complicated framework of cast iron on both sides of the road surface, the height of the iron structure is about 4 meters. The old bridge ends where its former course is cut by the more recent bridge construction.

The prescreener was tried on three different total power (odd+even) images, namely the 9-look averaged image, the selective Stokes averaged image and the single-look beam-weighted image, and on the pertinent PWF versions of these images. Fig. 3a shows one example of the prescreener result for the Stokes averaged case without PWF. All pixels that exceeded the threshold were set to +10, all others to -10 for good visualization. One finds that both bridges passed the prescreener, but that there is a lot of potential false alarms in the right part of the scene. As one expects, the single look image, with and without PWF, shows more gaps in the bridge pattern than the images with higher integration gain. The PWF does not differ visually from the total power image. The 9-look image prescreener result resembles closely the one with the Stokes averaged image.

In the next step, based upon the prescreener output, the feature test for discrimination is performed. Here, one has the choice between 15 different combinations of 1 to 4 features, moreover the distance factor γ can be varied. As there are six different inputs from the prescreener as described above, one sees that only a few out of the high number of possibilities can be analyzed. The criterion with which one judges the performance is partly qualitative by deciding visually whether the bridge pattern is recognizable well enough to warrant a precise and unambiguous pattern match in the subsequent image processing stage. And it is partly quantitative by determining the fraction f_1

		single-look		'Stokes'		9-look	
	γ	w/o PWF	w/ PWF	w/o PWF	w/ PWF	w/o PWF	w/ PWF
1	0.8	0.54	0.489	0.594	0.536	0.516	0.507
2	0.8	0.584	0.517	0.618	0.576	0.595	0.530
3	0.6					0.379	
	0.8	[0.097]	[0.099]	0.493	0.532	0.403	0.480
4	0.8	0.895	0.819	0.646	[0.667]	0.748	0.757
1,2	0.8	0.571	0.510	0.611	0.556	0.527	0.518
	1.0	0.587	0.555				
1,2,3	0.5			0.440		0.433	
	0.8	[0.395]	[0.393]	0.626	0.562	0.506	0.517
1,2,3,4	0.8	0.697	0.724	0.747	0.724	0.728	0.733

Tab.2: Fraction f_2 of false alarms in the test area

of pixels that passed the prescreener, and the fraction f_2 of these pixels that afterwards passed the feature test. For the calculation of f_1 and f_2, an area at the right end of the image was chosen where there can only be unwanted, i. e. false targets. The size of this area is 68,600 m^2 which constitutes a third of the complete scene.

Fig. 3b shows the discrimination result of the prescreener output in fig. 3a. Here, only faeture F_3 was used, with $\gamma = 0.8$ as distance coefficient. All pixels that passed the feature test were deliberately set to -30dB for good visualization. One sees that the two bridges passed the discriminator almost "unharmed". There are still quite a few false alarms in the right part of the scene (f_2=0.493, i. e. about 50% of the potential target pixels were eliminated). One cluster of these belongs to the road that leaves the lower bridge and obviously has some metal rail along the sides. The other cluster in the upper right belongs to the case where two parallel roads in the SAR image show bright edges. Although there does not exist any ground truth for this region, it seems as if there might be a shallow ditch that is crossed via bridges. Thus, this might actually not be a false target. All the other false alarm pixels are randomly scattered over the scene and should not deteriorate the performance of any pattern matching algorithm.

The results of the other detection runs are summarized in table 2. The prescreener rate f_1 lies between 8.5% and 9.4% in the false alarm test area for all six image types.

When a number appears in square brackets it means that the bridges were more or less eliminated by the discriminator. In all other cases the structure of both bridges was well preserved. One sees that F_3 fails completely on the single-look image (obviously the standard deviation of the clutter is almost as high as that of man-made structures due to the speckle), whereas it performs reasonably well on the 'Stokes' and the 9-look image. F_4 is the least successful of all features, the relatively best result is achieved on the 'Stokes' image, but here the effect of the PWF is desastrous. The reason for the bad performance of F_4 is not clear, perhaps a different choice of parameters could help. The combination of several features does in no case result in better performance than the use of individual features. The PWF shows its strongest effect on the images with the least integration (single-look and 'Stokes'), in the case of the 9-look image (F_3, F_4) it is even counterproductive. On the average, a reduction of the number of false alarms by 40% to 60% is achieved, and it seems plausible that by a better tuning of the feature parameters and by a more specific target reference, this result can still be improved.

7. Conclusions

It has been shown that SAR imaging at 94 GHz is a valuable tool also for military applications. One of the major advantages besides the good all-weather capability is the straightforward processing of focussing and motion compensation as compared to longer wavelengths. It has been demonstrated that for image generation it is possible to obtain a certain degree of integration gain while retaining useful polarimetric information that can be used for discrimination and classification purposes. A detection scheme for extended man-made targets has been outlined consisting of a threshold based prescreener and a feature based discrimination stage.

Four different discrimination features, partly polarimetric, partly geometric, were analyzed in different combinations. It was shown that they per-

form almost equally well on multilook images and on Stokes averaged images, whereas single look images cause a degradation. Whether or not the application of a polarimetric whitening filter leads to an improvement, depends on the type of feature. The example of two different bridges crossing a river is used to demonstrate the overall performance of the detection scheme. Although the number of false alarms is still rather high, they tend to be distributed randomly so that the bridges are the only characteristic structure in the image. Therefore, the output of the radar detection stage should be a valuable input to a subsequent pattern matching scheme. As the features that were tested are rather simple, and the target reference was a rather crude choice, a further improvement in the suppression of false alarms appears realistic.

References

[1] Essen, H.; Schimpf, H.: Remote sensing with an airborne 94 GHz synthetic aperture radar. European Conference on Synthetic Aperture Radar, Königswinter, March 1996.

[2] Wehner, D. R.: High-Resolution Radar, 2nd edition, 1995

[3] Keydel, W.: Basic principles of SAR. AGARD Lecture series 182, Bad Neuenahr 1992

[4] Novak, L. M. et al: Optimal Processing of polarimetric SAR imagery. Lincoln Lab. Journal Vol.3,2 (1990)

[5] Novak, L. M. et al: Studies of target detection algorithms that use polarimetric radar data. IEEE AES-25,2, p.150-165, 1989

[6] Novak, L. M. et al: Optimal speckle reduction in polarimetric SAR imagery. IEEE AES-26,2, p.293-305, 1990

[7] Novak, L. M. et al: Optimal polarimetric processing for enhanced target detection. IEEE AES-29,1, p.234-243, 1993

[8] Novak, L. M. et al: Performance of a high-resolution polarimetric SAR automatic target recognition system. Lincoln Lab. Journal Vol.6,1 (1993), p.25-52

[9] Kreithen, D. E. et al: Discriminating targets from clutter. Lincoln Lab. Journal Vol.6,1 (1993), p.53-76

[10] Schimpf, Hartmut: Polarimetric target/clutter discrimination. 3rd Int'l Workshop on Radar Polarimetry at IRESTE, Nantes, 1995, p.402-412

[11] Schimpf, H., Essen, H.: Evaluation of camouflage means on land targets at 94 GHz. AGARD Conf. Proc.542 (Supplement), Palma de Mallorca, Spain, 1993, p.19-1 to 19-10

[12] Kraus, John D.: Radio Astronomy, ch.4, McGraw Hill 1966.

Acknowledgement

This work was sponsored under contract of the German Ministery of Defence. We acknowledge the development of the SAR processing software and the generation of the SAR imagery by Alfred Wahlen.

Paper 42

U. Lammers (US)

You mentioned the potential use of 94 GHz SAR in connection with dual mode seekers. Are you envisioning a fly by before the actual attack?

Author's reply:

The most realistic application for 94GHz SAR is on board drones for surveillance and reconnaissance purposes.

For dual mode seekers (e.g. in connection with IR) a forward squinted SAR is a good possibility. Other schemes (seeker circling above the target area: spot-SAR combined with side-looking IR) have to be explored.

Airborne System Detection and Localization of Radio Interference Sources

B. Audone, F. Bresciani

G.F. Sistemi Avionici, 10072 Caselle Torinese, Torino, Italy

F. Novarese

Politecnico di Torino, Italy

Abstract - The theory of high resolution techniques used for direction finding, is rather well established for fixed antennas. The theory of a system based upon the rotational invariance techniques (ESPRIT) using a single pair of antennas in motion, such as it happens in airborne installation, is described. The results of a simulated receiver system are also presented.

1. Introduction

Radio Direction Finding has many applications, such as aircraft navigation , military reconnaissance and localization of disturbing electromagnetic field sources. Most of the existing techniques use methods, which assume the presence of a single radiation source. With multipath propagation caused by reflections or by the presence of more sources at the same frequency, these methods fail. High resolution techniques, allow to overcome this limitation.

High resolution is the ability of a Direction Finding system to resolve two or more signals that are separated by less then the beamwidth of an antenna array. The classical *Rayleigh Resolution Criterion* states that two signals are resolved if their two peaks in the spectrum are distinguishable.High-resolution techniques can overcome the *Rayleigh Resolution* limit, because they assume the underlying model of the received signal. The assumed model can include the number of received signals, the antenna array, the noise statistics, and even the signal modulation. This additional information, supplied by the more detailed signal model, makes better resolution obtainable.

In the next chapter a high resolution algorithm is applied to the samples measured by two antennas mounted on an aircraft. Then, an experimental direction finding system applying these methods is described.

2. Array Doublet in motion

We consider an array composed of two sensors separated of a distance ξ.

The elements of the doublet are assumed to have identical, but not necessarily isotropic, patterns (Fig. 1).

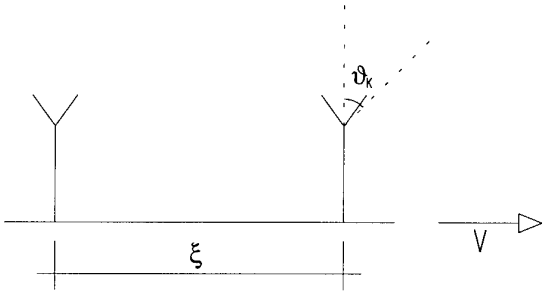


Fig. 1 Doublet in motion and signal arrival geometry.

During the observation interval, each of the sensors changes its position in an arbitrary but known way. To simplify the presentation, we assume that the array elements and signal sources are co-planar. Let d narrow band plane waves at a known frequency ω_0 impinge on the array from directions $\{\theta_1,\ldots\ldots, \theta_d\}$. Using complex signal representation, the signal received on the first element, can be written as:

$$x_1(t)=\sum_{k=1}^{d} s_k(t)Exp[j\phi_{1k}]+n_1(t)$$

where:

$$s_k(t)=A_kExp[-j\omega_0(1+(v/c)sin\theta_k)t]$$

A_k is the complex amplitude, ϕ_{1k} is the carrier phase, and c is the velocity of propagation. At the second element, the corresponding signal is:

$$x_2(t)=\sum_{k=1}^{d} s_k(t)Exp[j\phi_{1k}]\ \Delta_k+n_2(t)$$

where:

$$\Delta_k=Exp[-j\omega_0(\xi/c)sin\theta_k]$$

is the additional propagation time to the second element of the doublet.

Paper presented at the AGARD SPP Symposium on "Remote Sensing: A Valuable Source of Information", held in Toulouse, France from 22-25 April 1996, and published in CP-582.

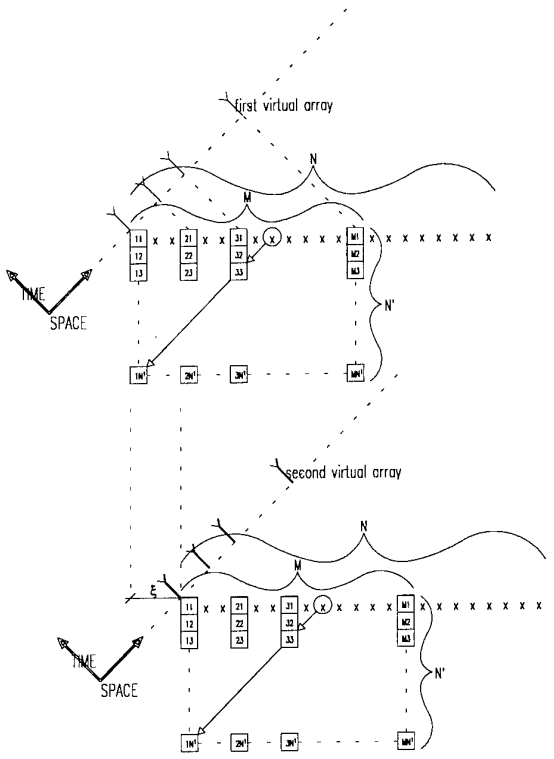


Fig. 2 Overview of virtual arrays in space-time representation

Now we take N samples of the signal $x_1(t)$ and N samples of the signal $x_2(t)$, $(N>>d)$ at a constant sampling rate r. The time Δt between two samples is $1/r$; the total observation time T over which data are taken is $T=(N-1)\Delta t$.

If we assume the doublet moves with a constant velocity v parallel to the displacement ξ, it is possible to build a matrix with M elements and N' snapshots. We build the matrix dividing the N data samples of x_1 and x_2 into M $(N>>M>>d)$ equilength overlapping column-vectors of N' elements. The time delay between two adjacent column-vectors is $\tau = n\Delta t$ corresponding to a geometrical distance $\xi=\tau v$ (Fig. 2).

The parameter n is given by the following relation

$$n=(N-N')/(M-1)$$

In our case we have selected $N=8000$ samples, $N'=1000$ and $M=15$ virtual sensors. It follows that the mutual distance between the two antennas of the synthetic array is $500\ \Delta t$.

Analytically the two matrices of Fig. 2, corresponding to the doublet are

$$X_1=A_1S+N_1$$
$$X_2=A_1\Phi S+N_2$$

where A_1 is an Mxd Vandermonde matrix

$$\begin{bmatrix} 1 & 1 & . & . & 1 \\ Exp[-j\phi_1]n\Delta t & Exp[-j\phi_2]n\Delta t & . & . & Exp[-j\phi_d]n\Delta t \\ Exp[-j\phi_1 2]n\Delta t & Exp[-j\phi_2]n\Delta t & . & . & Exp[-j\phi_d 2]n\Delta t \\ . & . & . & . & . \\ . & . & . & . & . \\ Exp[-j\phi_1(M-1)]n\Delta t & Exp[-j\phi_2(M-1)]n\Delta t & . & . & Exp[-j\phi_d(M-1)]n\Delta t \end{bmatrix}$$

where $\phi_k=\omega_0(v/c)sin\theta_k$

S is the dxN' signal matrix

$$\begin{bmatrix} s_{11} & s_{12}Exp[j\omega_0\Delta t] & s_{13}Exp[j\omega_0\Delta t2] & . & . & s_{1N'}Exp[j\omega_0\Delta t(N'-1)] \\ s_{21} & s_{22}Exp[j\omega_0\Delta t] & s_{23}Exp[j\omega_0\Delta t2] & . & . & s_{2N'}Exp[j\omega_0\Delta t(N'-1)] \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ s_{d1} & s_{d2}Exp[j\omega_0\Delta t] & s_{d3}Exp[j\omega_0\Delta t2] & . & . & s_{dN'}Exp[j\omega_0\Delta t(N'-1)] \end{bmatrix}$$

and

$$\Phi=Diag[Exp[j\omega_0(\xi/c)sin\theta_1],Exp[j\omega_0(\xi/c)sin\theta_2],...,Exp[j\omega_0(\xi/c)sin\theta_d]]$$

Combining X_1 and X_2, , the following expression is obtained

$$X=\begin{bmatrix} X_1 \\ X_2 \end{bmatrix}=AS+N$$

where

$$A=\begin{bmatrix} A_1 \\ A_1\phi \end{bmatrix};N=\begin{bmatrix} N_1 \\ N_2 \end{bmatrix}$$

The objective is to estimate the arrival angles θ_k from the matrix X. The total least squares (TLS) ESPRIT method is used [1]. The basic idea is to exploit the rotational invariance of the underlying signal subspace induced by the traslational invariance of the virtual sensor arrays. The relevant signal subspace is the one that contains the outputs from the two virtual subarrays X_1 and X_2, described in Fig 2.

The algorithm can be summarised as follows.

1. Obtain an estimate of R_{xx} (correlation matrix), from X
2. Compute the eigen-decomposition of R_{xx}
   $R_{xx}E=E\Lambda$
   where $\Lambda=diag\{\lambda_1,.....,\lambda_M\},\lambda_1\geq...\geq\lambda_M$ and
   $E=\{e_1,.....,e_M\}$
3. Estimate the biggest d eigenvalues of Λ
4. Obtain the vector matrix $Es=\{e_1,.....,e_d\}$ and decompose it in E_0 and E_1 using the selection matrix J virtual subarrays,

$$JE_s=\begin{bmatrix} E_0 \\ E_1 \end{bmatrix}$$

5. Compute the eigendecomposition($\lambda_1 > \ldots > \lambda_{2d}$)

$$\begin{bmatrix} E^*_0 \\ E^*_1 \end{bmatrix} \begin{bmatrix} E_0 & E_1 \end{bmatrix} = H \Lambda H^*$$

6.

$$H = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix}$$

where * indicates complex conjugate

7. Calculate the eigenvalues ϕ_k of $\Psi = -V_{12} V_{22}^{-1}$
8. Estimate $\theta_k = sin^{-1}\{c \arg(\phi_k)/(\omega_0 \xi)\}$

The maximum number of angles that can be resolved with ESPRIT algorithm is M; in the complex matrix M shall be greater than the number of sources d. It is also possible to use a single element antenna; in this case the samples of $x_1(t)$ and $x_2(t)$ are obtained from two time-shifted subsets of the same samples. The estimated angles would be affected by sampling time jitters or velocity errors.
Several parameters introduced in the above algorithm, must be chosen properly if we want to find the desired directions of the sources. One of these parameters is the time displacement τ between adjacent subvectors. If τ is too large, one cannot be certain that the matrix A will have full rank because two columns of A may be identical for certain arrival angles. Two distinct θ_k will always give two different values of
$Exp[-j\omega_0(v/c)sin\theta_k \tau]$ if $\tau \leq c/2vf_0$.
Next, we consider the total observation time T. In a practical situation, the maximum value of T is limited by the validity of the assumption that the signal sources are in the far field of the doublet. In practice, however, with a source at a finite distance, the signal arrival angle at the doublet changes as the doublet moves. For a source at distance R_0, the arrival angle could change as much as $\varepsilon = Tv/R_0$. Thus the maximum value of T is limited by the maximum value of ε that can be tolerated.

3. Simulation system

In order to validate the previous theory in a more realistic situation a simulation of the receiver has been carried out using Omnisys a CAE tool developed by HP EEsof.
The Direction Finding (DF) system consists of a receiver connected to an array doublet in motion in the presence of two radiating sources operating at the frequency of 2000 MHz. The general view of the DF system is shown in Fig 3.

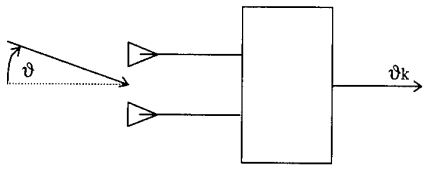


Fig. 3 DF system schematic diagram.

A detailed view of the overall simulation is given in Fig. 4 where the generators are in block 1 and the receivers are in blocks 2 and 3.
The movement of the DF system is obtained by driving the phase modulators (MODPM blocks) shown in Fig. 5 with ramp generators. The noise generators are added to study the noise effect on the overall system performances.
The blocks ANT simulate the coupling between transmitter and receiver antennas taking into account the antenna propagation path and direction of arrival of the emitting source.

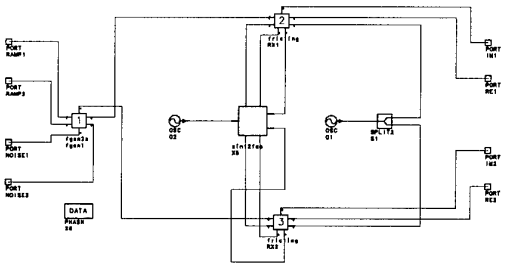


Fig. 4 Detailed view of the overall simulation.

In Fig. 6 the receiver block diagram is shown. It consists of three IF stages with three local oscillators. The IF bandwiths are 100 MHz, 20 MHz and 400 KHz.
The final IF stage utilises the I/Q demodulation and provides the input signal to the processing system where the estimated directions of arrival are calculated. In our simulation these data are put in output files RE1, RE2, IM1, IM2 which are formatted by a utility program developed in C for further processing in Mathematica.

4. Simulation results

In this section we present the statistical performance of the simulated system.
The general configuration of the doublet in motion is an array of two sensors (with d=λ/2) at the frequency of 2000 MHz with two sources at 0.2 and 0.35 rads (the angles are measured with respect to the normal to the array). The sampling is made in 2 ms taking N=8000 samples; the doublet is assumed to be installed on an aircraft which has the speed of 300 m/s.
In Fig. 7 the standard deviations and average values versus S/N are shown assuming M=15 and N'=1000. This means that n=500 and the distance between the virtual sensor is 3.75 cm. If S/N>10 dB the standard deviation is 0.04 rads and the average value starts to approach the actual correct direction of arrival.
In Fig. 8 the standard deviations and average values versus N' (the number of snapshots) are shown assuming M=15, S/N=50 dB, and n=500. It is interesting to notice that the standard deviation decreases rapidly until there is no overlap between adjacent virtual sensors; it becomes slowly decreasing when virtual sensors start to overlap. This situation happens when N'=500.
In Fig. 9 the standard deviation and average values versus n are shown assuming M=15, S/N=50 dB and N' given by equation 1. From the diagram it appears that there is an optimum value of approximately n=150 which minimizes the standard deviation.

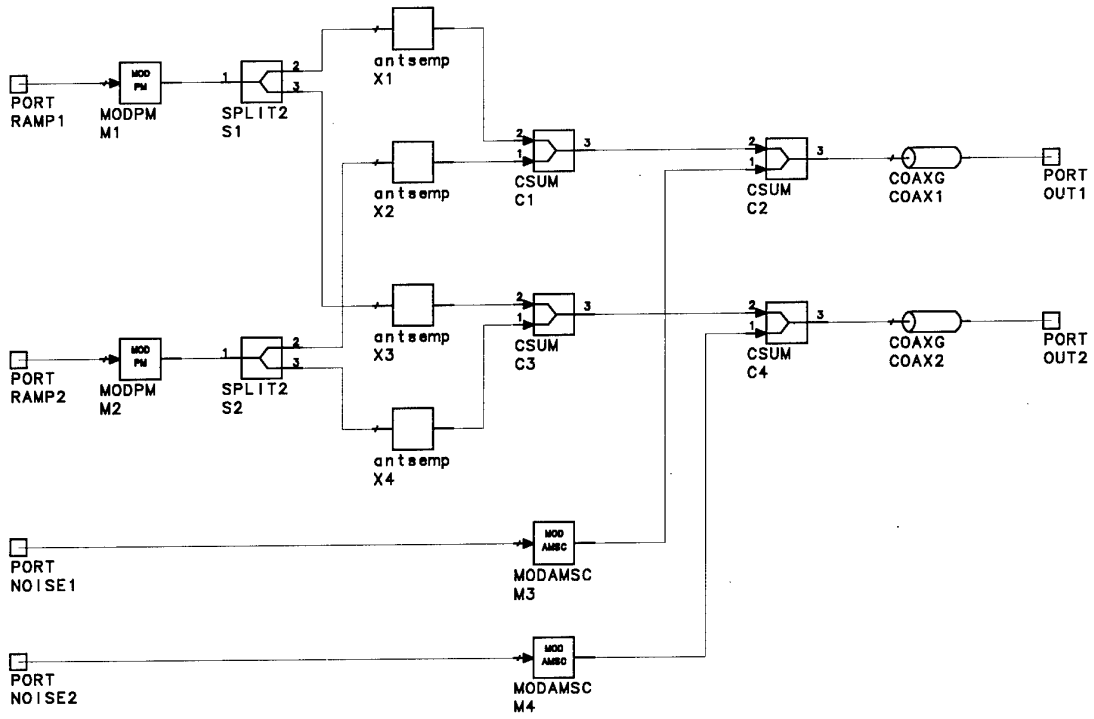


Fig. 5 Generation of receiver input signals.

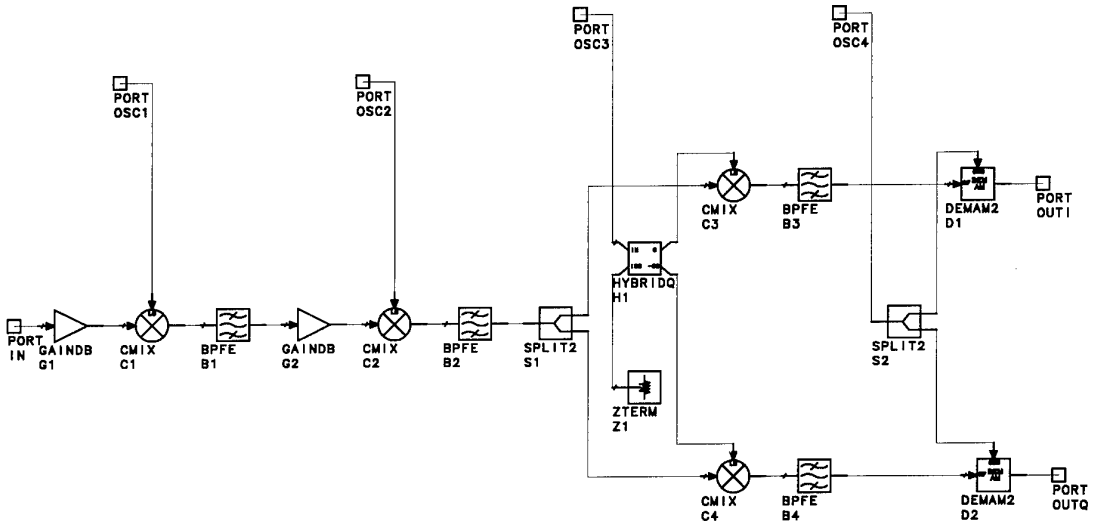


Fig. 6 Receiver Block Diagram.

On the basis of the optimum value of n the standard deviations and average values versus M are shown in Fig 10 for different values of S/N (10, 20, 50). The angle of arrival of the single source is 0.35 rads.

5. Conclusions

This study has evaluated the performances of a sensor doublet installed on an aircraft. More work has still to be done to investigate the receiver in terms of its performances with regard to the use of actual antennas and processing hardware. In the present simulation the array antennas are assumed ideal (without mutual coupling); the ESPRIT algorithm implemented in MATHEMATICA is not evaluated in terms of those round off errors which the actual processing hardware can introduce. A further improvement could be the use of a digital receiver instead of a superhet one.

References

1. R.Roy, T.Kailath "ESPRIT-Estimation of Signal Parameters via Rotational Invariance" Optical Engineering April 1990
2. Jian Li,R.T.Compton "Angle of Arrival Estimation of Coherent Signals Using An Array Doublet in Motion" IEEE Trans. on Aerospace and Electronic Systems January 1994
3. T Shan.,M.Wax,T Kailath "On Spatial Smoothing for Direction of Arrival Estimation of Coherent Signals" IEEE Trans. on Acoustic,Speech,and Signal Processing August 1985

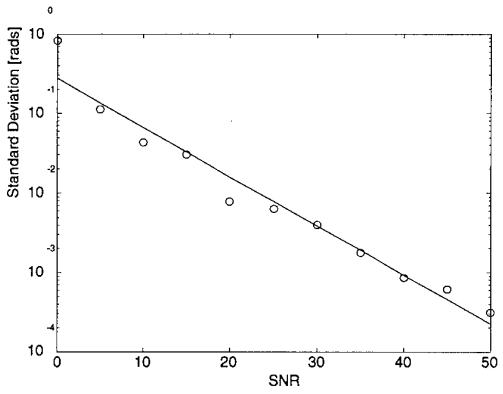


Fig. 7a

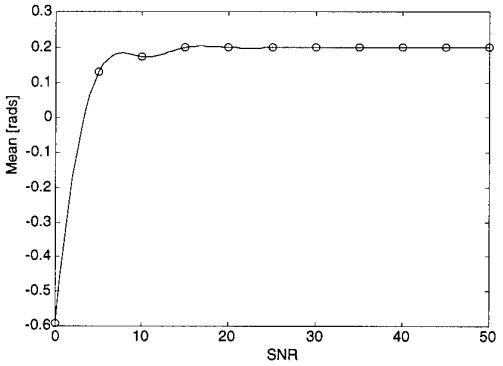


Fig. 7b

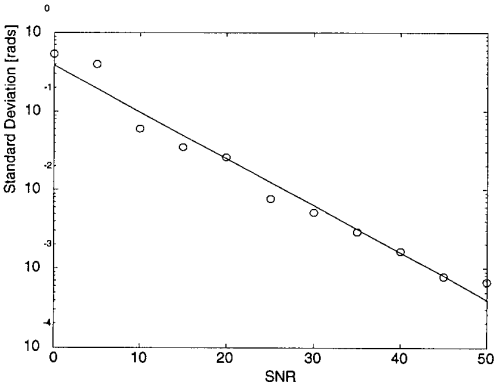


Fig. 7c

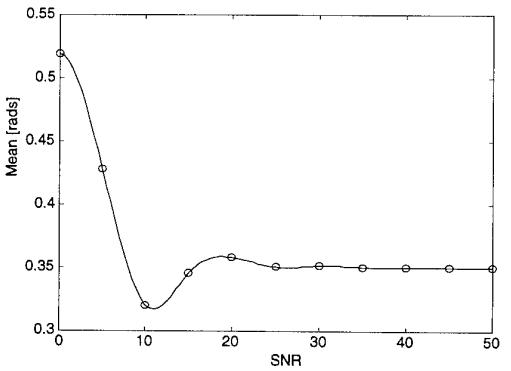


Fig. 7d

Fig. 7 Standard deviation and average value versus S/N

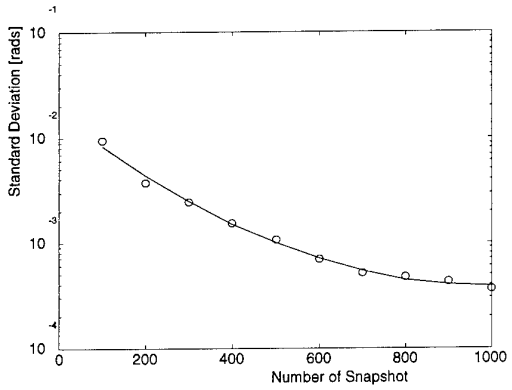


Fig. 8a

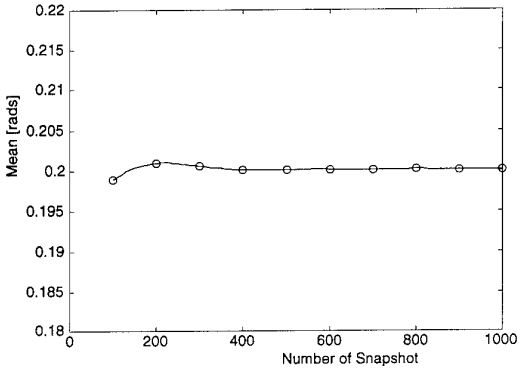


Fig. 8b

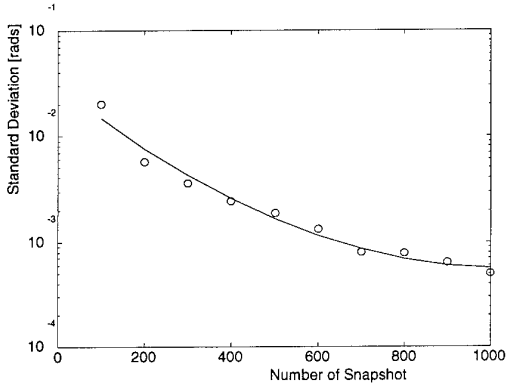


Fig. 8c

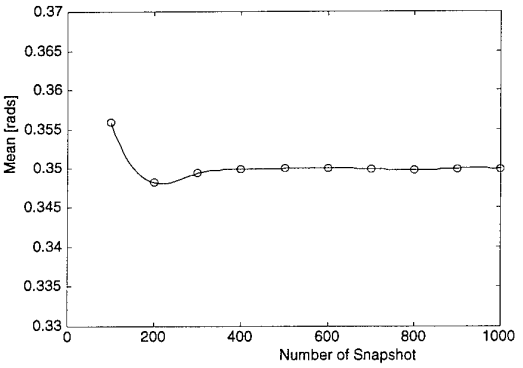


Fig. 8d

Fig. 8 Standard Deviation and Average Value versus N'

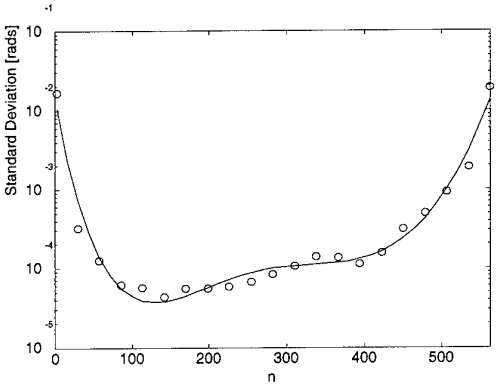


Fig. 9a

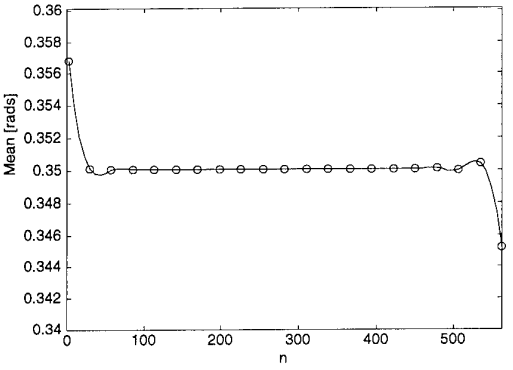


Fig. 9b

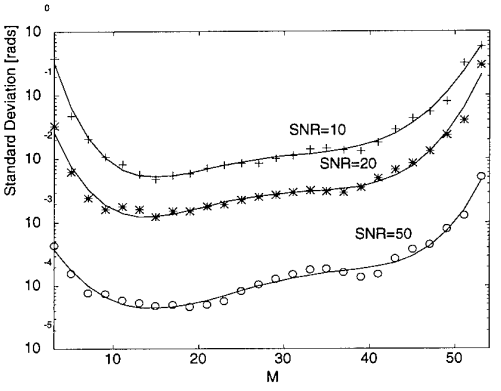


Fig. 10a

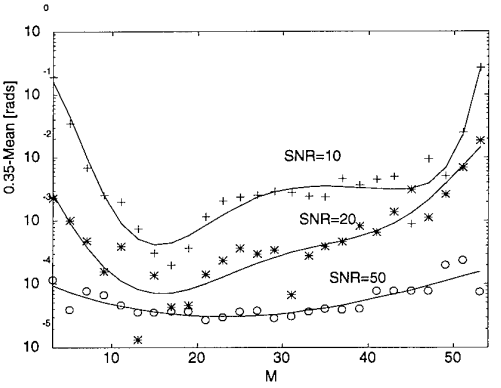


Fig. 10b

Fig. 9 Standard deviation and average value versus n
Fig. 10 Standard deviation and average value versus M.

Paper 43

N. Farsaris (GR)

Question: Have any efforts been made in the case of accelerating or decelerating aerial platform?

Comment: Inversing this technique, I think that we can detect the direction of a moving signal source. (A ship or a plane transmitting with its radio, or radar).

Author's reply:

The idea is quite good but for the moment this has not yet been investigated.

VLSI Implementation of a chip for a Systolic Array Architecture for Direction Finding Systems

B. Audone – F. Bresciani
GF – Sistemi Avionici
10072 Caselle Torinese (Italy)

D. Zerlottin
Modulo Uno
Via Cuorgne' 21
Torino (Italy)

1. Summary

Recently the new class of so-called subspace methods for high-resolution direction finding has received a great deal of attention in the literature.

In this paper a new processor designed for parallel filter structure (systolic array) is presented. The processor is a single chip VLSI implementation of an electronic board for systolic array which computes the QR, the SVD and the GSD(Generalized Schur Decomposition) of a complex matrix. The array utilizes CORDIC (Coordinate Rotation Digital Computer) arithmetic to perform the vector rotations and inverse tangent calculation in hardware. The resulting architecture is a two-layer Jacobi array that can handle all the subproblems for solving Direction-Finding problems.

This component will be used for a direction finding system of advanced characteristics to be installed on airborne platforms.

The main advantages of the high resolution technique consist in improved resolution (because it can overcome the limitations of the Rayleigh criterion) and in the possibility of detecting sources located in different positions and emitting signals at the same frequency.

2. Introduction

This paper describes a VLSI implementation of a chip for systolic processor array in order to determine in real-time the direction of arrival (DOA) of electromagnetic sources.

The direction of arrival estimation problem consists in determining the angles of arrival of a number of signals impinging on a sensor array.
Recently, the ESPRIT method of Roy and Paulraj [1] has received a great deal of attention in literature. This method belongs to the class of so-called subspace methods.
The application of ESPRIT method needs two identical sensor arrays, displaced over a distance Δ, and with m sensors each.

Let d narrow band plane waves at a known frequency ω_0 impinge on the two arrays from directions $\{\vartheta_1, \vartheta_2,...\vartheta_d\}$. Using complex representation, the signal received on the arrays from a direction k is: $s_k(t) = x_k * \exp(j\omega_0 t)$ where $x_k(t)$ is the unknown slowly time-varying complex amplitude.
Noise is present at each sensor, and is assumed to be additive, stationary, and zero mean. If N>>m is the number of the snapshots incident on the first array of m sensors, then it is possible to collect all these observations into one matrix X, with N rows and m colums:

$$X=S*A+N_x \qquad (1)$$

S is the signal matrix, with $S(i,j) = s_j(t_i)$, A is the array gain matrix (Vandermonde Matrix), with A (i, j) the gain of sensor j in the direction of signal s_i and N_x is the noise matrix with N_x (i, j) the noise present at time t_i for sensor j. The signals measured by the relevant sensors on the second array are equal up to a phase shift, which for signal s_k is given by

$$\phi_k=\exp(-j\omega_0\Delta\sin(\vartheta_k)/c) \qquad (2)$$

c is the signal propagation velocity
ϑ_k is the unknown angle of arrival
Δ is the displacement between the two arrays

With the data collected on the second array, it is possible to build a matrix Y.
This matrix satisfies the equation

$$Y=S*\Phi*A+N_y \qquad (3)$$

where $\Phi=\mathrm{diag}(\varphi_1,\varphi_2,...\varphi_d)$
The direction finding problem now reduces to estimating Φ, from which ϑ_k 's can be computed directly.
In the special case of no noise the DOA's are the solutions of the matrix pencil

$$Y-\chi X=S*(\Phi-\chi I)*A \qquad (4)$$

The non-trivial generalized eigenvalues χ's correspond to the ϕ'$_k$s.

Paper presented at the AGARD SPP Symposium on "Remote Sensing: A Valuable Source of Information", held in Toulouse, France from 22-25 April 1996, and published in CP-582.

In general noise changes the rank reducing the accuracy of the solutions.
It is necessary to make the pencil Y-χX smaller in order to reduce the effect of noise. This pencil compression is obtained by multiplying the expression (4) on both sides with the matrices P_{row} [dxN] and P_{col} [mxd] respectively:

$$P_{row} *Y*P_{col} -\chi*P_{row}*X*P_{col}=P_{row}*S*(\Phi-\chi I)*A*P_{col} +Noise=>0 \quad (5)$$

The rank of the pencil is reduced to a [dxd] matrix.
The approach of choosing the compression matrices P_{row} and $P_{col,}$ is an adaptive algorithm implemented on a systolic array.
This adaptive algorithm is based on three principal trasformations.
The first one is the QR which puts a zero in a specific position of a matrix.
QR is used to triagularize the data matrix. This is done by applying the trasformation on both sides of the data matrix.
The second trasformation is the SVD which calculates the singular values of the data matrix [X|Y]. This is done through matrix rotations.
The last trasformation is the GSD; it obtains the solution of the compressed pencil dxd. The GSD is an iterative algorithm; in only two steps it is possible to achieve good accuracy of the results.

3. Algorithm description

The algorithm implemented [2]is:

$\boldsymbol{V} \Leftarrow \mathrm{I}_{mxm}$
$\boldsymbol{R} \Leftarrow \mathrm{O}_{mxm}$
$\boldsymbol{X}_{ss} \Leftarrow \mathrm{O}_{mxm}$
$\boldsymbol{Y}_{ss} \Leftarrow \mathrm{O}_{mxm}$
$\boldsymbol{W} \Leftarrow \mathrm{I}_{mxm}$
$\Pi \Leftarrow \mathrm{I}_{mxm}$
for k = 1, ..., ∞

1. input of the new snapshots

$$\tilde{z}\left(t_k\right)^T \Leftarrow z\left(t_k\right)^T \cdot \boldsymbol{V}$$

$$\tilde{x}\left(t_k\right)^T \Leftarrow x\left(t_k\right)^T \cdot \boldsymbol{W}$$

$$\tilde{y}\left(t_k\right)^T \Leftarrow y\left(t_k\right)^T \cdot \boldsymbol{W}$$

2. QR of the new vectors calculate

$$\left[\begin{array}{c|c|c} \boldsymbol{R} & \boldsymbol{X}_{ss} & \boldsymbol{Y}_{ss} \\ \hline 0 & \cdots & \cdots \end{array}\right] = \tilde{\boldsymbol{Q}}(k)^{\mathrm{H}} \cdot \left[\begin{array}{c|c|c} \alpha \cdot \boldsymbol{R} & \alpha \cdot \boldsymbol{X}_{ss} & \alpha \cdot \boldsymbol{Y}_{ss} \\ \hline \tilde{z}\left(t_k\right)^T & \tilde{x}\left(t_k\right)^T & \tilde{y}\left(t_k\right)^T \end{array}\right]$$

3. combined SVD e GSD
for i = 1, ..., *m*−1

if $\boldsymbol{R}_{[i,i+1]}$ has two big values in the diagonal
then

a) To determine $\tilde{\boldsymbol{U}}_{[i,i+1]}$ and $\tilde{\boldsymbol{T}}_{[i,i+1]}$ by the GSD of matrixs $\boldsymbol{Xss}_{[i,i+1]}$ e $\boldsymbol{Yss}_{[i,i+1]}$.

b) To take $\boldsymbol{U}_{[i,i+1]} = \tilde{\boldsymbol{U}}_{[i,i+1]}$ and calculate $\boldsymbol{T}_{[i,i+1]}$ to obtain $\left[\boldsymbol{U}_{[i,i+1]}(k)^{\mathrm{H}} \cdot \boldsymbol{R}\right]_{[i,i+1]}$ upper triangular.

else

a) Calculate the matrixs $\boldsymbol{U}_{[i,i+1]}$ e $\boldsymbol{T}_{[i,i+1]}$ from the SVD of $\boldsymbol{R}_{[i,i+1]}$.

b) To take $\tilde{\boldsymbol{U}}_{[i,i+1]} = \boldsymbol{U}_{[i,i+1]}$ and $\tilde{\boldsymbol{T}}_{[i,i+1]} = \mathrm{I}_m$.

endif.

$$\boldsymbol{R} \Leftarrow \Pi_{[i,i+1]} \cdot \boldsymbol{U}_{[i,i+1]}(k)^{\mathrm{H}} \cdot \boldsymbol{R} \cdot \boldsymbol{T}_{[i,i+1]}(k) \cdot \Pi_{[i,i+1]}$$

$$\boldsymbol{X}_{SS} \Leftarrow \Pi_{[i,i+1]} \cdot \tilde{\boldsymbol{U}}_{[i,i+1]}(k)^{\mathrm{H}} \cdot \boldsymbol{X}_{SS} \cdot \tilde{\boldsymbol{T}}_{[i,i+1]}(k) \cdot \Pi_{[i,i+1]}$$

$$\boldsymbol{Y}_{SS} \Leftarrow \Pi_{[i,i+1]} \cdot \tilde{\boldsymbol{U}}_{[i,i+1]}(k)^{\mathrm{H}} \cdot \boldsymbol{Y}_{SS} \cdot \tilde{\boldsymbol{T}}_{[i,i+1]}(k) \cdot \Pi_{[i,i+1]}$$

$$\boldsymbol{V} \Leftarrow \boldsymbol{V} \cdot \boldsymbol{T}_{[i,i+1]}(k) \cdot \Pi_{[i,i+1]}$$

$$\boldsymbol{W} \Leftarrow \boldsymbol{W} \cdot \tilde{\boldsymbol{T}}_{[i,i+1]}(k) \cdot \Pi_{[i,i+1]}$$

$$\Pi \Leftarrow \Pi \cdot \Pi_{[i,i+1]}$$

end

$$\text{output} = \Pi \cdot \left[\frac{\boldsymbol{Yss}(1,1)}{\boldsymbol{Xss}(1,1)} \quad \cdots \quad \frac{\boldsymbol{Yss}(m,m)}{\boldsymbol{Xss}(m,m)}\right]$$

This algorithm shows three different structures of data: *V* and *R* representing the level *Z*, X_{SS} and Y_{SS} representing the levels *X* and *Y* . *Z* , *X* and *Y* represent three different sensor arrays of *m* sensors; with these arrays we calculate the directions of arrivals of the impinging signals.
The first level, *Z* (composed by matrix *R* e *V*) calculates the SVD of the input snapshots. Each snapshot, represented by *z(t)*, arrives from the sensor array; the snapshot is a vector of *m* complex numbers arriving from the sensor array. On the level *Z* the inputs are multiplied for the matrix *V* and then the resulting vector becomes an input for the *QR* phases on matrix *R*. Finally the matrix *R* undergoes an *SVD* transformation.
The calculated operations on the first level permit to determine the number of impinging signals and so to know where the signal components and the noise components are located in the structure of the space recognized from the sensor arrays.
When the first level recognizes where the signal components are located in the structure (this action is done in the *if-then-else* test of the algorithm) *GSD* or *SVD* in the particular positions of the *R*, X_{SS} and Y_{SS} matrices are calculated.
The matrices Π are used to obtain a full mixing of all parts of the data matrices. These matrices are similar to matrices of *ODD-EVEN* algorithm [3].
All the operations in the algorithm are performed with 2*x*2 matrix transformations inserted in a matrix with the appropriate dimension. These matrices have the following form:

$$\begin{array}{c} \begin{matrix} 0 & \cdot & i & i+1 & & m \end{matrix} \\ \begin{bmatrix} 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 1 & 0 & \cdot & \cdot & \cdot \\ \cdot & 0 & a & b & 0 & \cdot \\ \cdot & 0 & c & d & 0 & \cdot \\ \cdot & \cdot & \cdot & 0 & 1 & 0 \\ 0 & \cdot & \cdot & \cdot & 0 & 1 \end{bmatrix} \begin{matrix} 0 \\ \cdot \\ i \\ i+1 \\ \cdot \\ m \end{matrix} \end{array}$$

The dimension of the matrices used are *mxm*; these matrices are similar to diagonal matrices. The transformations are located in subdiagonals of 2*x*2 dimensions.

The operations *QR, SVD* and *GSD* should be carried out with easy complex rotations; also the vector-matrix products should be performed with a suitable number of these transformations.

As it is known , the proper hardware architecture to calculate and apply rotations is the module CORDIC [4]. It's also possible to realise products with the module CORDIC. It only needs some variants in its architecture.

4. CORDIC

The CORDIC processor is the algorithm *VLSI* architecture for computing and applying the rotations as shown in Fig. 1:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \mathrm{R}(\theta) \cdot \mathbf{v}$$

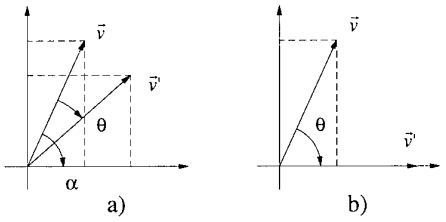


Fig. 1 CORDIC Processing

It is possible either to compute or to apply the angle θ only by initialising the processor in different ways. In Fig. 1a) the vector $\vec{v}$ rotated by the angle θ is shown; the procedure to calculate the angle θ is shown in Fig. 1b.

With these transformations it is possible to do all the operations described before. In fact it is possible to calculate the *SVD* (or *GSD*) of the generic matrix *A* only with two transformations; the former on the left hand side and the latter on the right hand side. This is true if the matrix *A* is composed of real elements.

5. Systolic Arrays

Systolic arrays are hardware architectures used in real time applications. The algorithm described above can be implemented in a systolic array.

Our application requires high speed of calculations and high response speed; these performances are suitable to parallel realizations. The systolic array is a hardware electronic architecture that has got a big level of parallelism. It is easily implementable in *VLSI*.

The structure of a systolic array is suitable to matrix trasformations (mainly for *Jacobi* rotation). In fact in our case there are a series of processors that work on submatrices of 4 elements that are placed in particular positions of the data structure. There are 4 different ways to place the processors in the structure; they are named Phase 1-2-3-4. In this way the systolic array operates the trasformations on all elements [5].

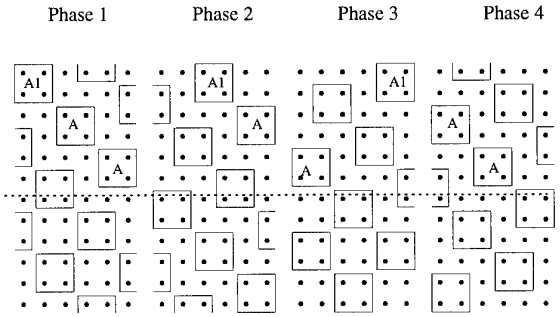


Fig. 2 SVD and DOA data flow

In our problem we work on 2 different square matrices: the former, in the high part, represents the matrices *V* and *W* and the latter represents *R*, X_{SS} and Y_{SS} . The processors or cells are mapped on the data as shown in Fig. 2.

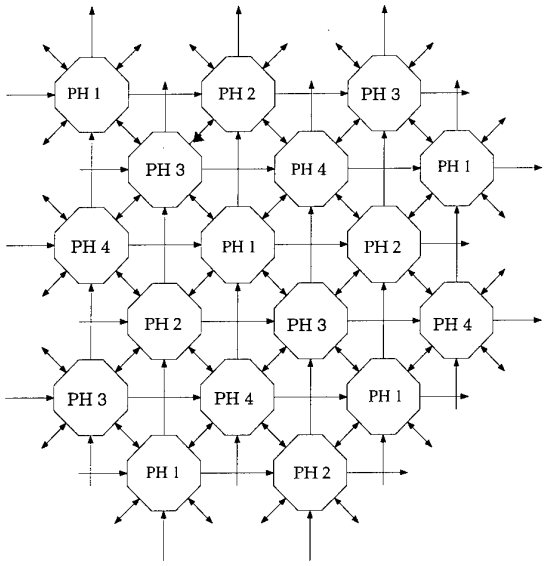
There are 4 different ways to map the cells on the data. At the time zero the processors of the Phase 1 are active, at the end of the processing they pass the data to the near cells, which will be active in the following phase.

There are three different behavioural modes for the cells: the first one for the processors that work on the data of the matrix above the dotted line (*V*), the second one for the processors that work on the diagonal of the matrix in the bottom (*R*) and the last one for the cells that work in the other data of the matrix in the bottom.

The most important types of cells are the diagonal ones; these processors calculate and apply the rotations for *QR, SVD* and *GSD* and propagate this value to the others phases.

The cells that work on the data of the matrix *V* (and *W*), called upper cells, apply only the rotations calculated by the diagonal cells on the right hand side of the data matrix *R* (X_{SS} and Y_{SS}). Then they calculate the product between the input snapshot and the data which are memorized only if they are in a particular position (A). Finally the processors put the results of the product

vector-column in a structure that propagates this data toward the diagonal cells.
The so called internal cells, that are positioned near the diagonal cell on the data of matrix *R* apply the rotations calculated by the diagonal cells to all data.[6]
There are two or more ways to realize the systolic array. The choice depends on the complexity of the processsor. The first way is the one shown in Figure 2. Every processor that knows the three kinds of the above mentioned functions map its structure on the data. This method is possible if the operations that the systolic array has to do are simple and if the exchanges of the structure data among the cells is not very frequent.
The second way provides processors that memorize the data of all levels (three 2x2 complex matrices one for every sensor array and three vectors which contain the value of the input snapshot). The data are shared with the near cells; they receive the rotation parameters from the cells situated on the left hand and bottom sides, apply them and then transfer them to the cells. This choice requires a bigger number of processors. Only a quarter of them work simultaneously, but the processor architecture is easy to implement and to test; besides, in this case, there is no problem to propagate a big quantity of data.
This choice provides the structure of connections among the processors shown in Fig. 3:



PH = PHASE

Fig. 3 Systolic Array Connection

6. Datapath

The elementary cell of the systolic array for the direction finding problem resolution contains two different parts:

- The Datapath;
- The Control Unit.

The Datapath is composed of the components required to execute the mentioned Matrix Pencil algorithm over the data coming from the sensor arrays. It will be implemented as a chip, while the control unit will be an external support of each elementary processor.
The control unit tells the datapath which operation it has to execute and which are the data to be elaborated.
One of the advantages arising from the above mentioned division is that the datapath must be unique for the three different kinds of cells of the systolic array, while it would be easier to realise a different control unit for each of them. In this prototype it is hard to know the real complexity and dimension of each control unit. Later it would probably be better to incorporate the three kinds of control units into a single one, and perhaps add it to the datapath as a chip.
Figure 4 shows the whole structure of the systolic array elementary cell and the relevant connections.
Both the Control Unit and the Operative Unit receive the clock and the reset signals.

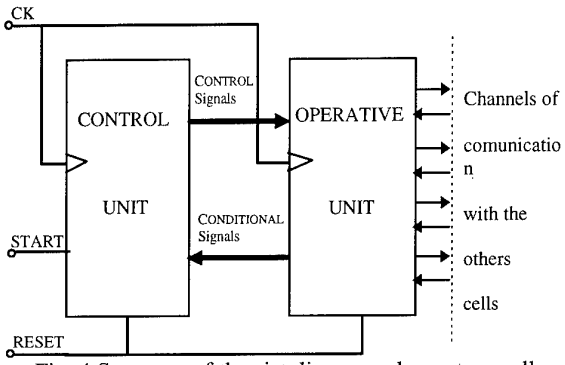


Fig. 4 Structure of the sistolic array elementary cell

The control unit receives a "start" signal coming from the general controller of the systolic array which indicates the beginning of its working time. It also receives a flag from the datapath and tests it sometimes during the execution of the algorithm.
Besides there are 44 control signals transferred from the control unit to the datapath, by which every operation is specified.
In the following their detailed functions will be explained.
The datapath also has 8 serial input and 8 serial output channels for the connection with the other cells of the array. Every input/output pair of channels is dedicated to the communication between two near cells.

In fact every elementary processor must communicate with its 8 nearest cells.

The Control Unit has been developed by a VHDL description to simulate the functionality of the datapath.
This kind of description is suitable to help the actual manufacturing.
The control unit has not been yet designed at the level of an electronic circuit; a VHDL description of its behaviour has been developed. This description could be almost immediately translated into a state machine based on a RAM or some similar memory element. Due to the great number of states needed to perform all the steps described in the [Matrix Pencil] algorithm (in the

order of some thousands), it is not conceivable to implement the control unit with registers and a combinatorial net.
If the address input of the RAM is clocked, it works as a state register: it contains the present state of the system, and on the clock edge it loads the future state. Depending on the present state (the address), the proper memory word is selected and part of its contents are sent to the data path: these are the control signals that tell the data path the operation it has to compute. The remaining part of the memory word contains the future state, and is fed back directly to the address input, or it contains signals that allow to select, by means of some external logic, the proper future state that is presented to the address input.
This external logic, that can be implemented by a programmable logic, will be more or less complex depending on which mechanisms the control unit will be able to manage. Since the behaviour of the systolic array has to be fixed and well defined, it will include a clock counter and a phase counter to allow the correct timing of the operations on the data path. Obviously, the same clock signal goes to all the elementary cells of the systolic array, synchronising the whole system. Sometimes the control unit has to check some flags from the data path, and, depending on their state, it has to make conditional jumps. In the simplest case, still sufficient for our purposes, a single flag is selected and tested to perform binary decisions: the state of the flag determines whether to load the next future state or the one provided by the output of the RAM.
Another mechanism that the control unit has to handle is a sort of procedure call. The algorithm requires the application of a certain number of rotations on different plans of data and of different angles. Writing the expanded control sequence for every rotation would be tedious, memory consuming and prone to errors. It is better to provide general procedures, one for every kind of rotation, capable to work on different data and on different angles. The data path is designed accordingly to this requirement: the input buffer that receives the command signals from the data path is duplicated in the sections regarding the address bus to its internal dual port RAM. The control unit can send either the command word to the data path, which is stored as usual in the buffer, or can send the base address of the matrix or vector on which the rotation has to be performed, and the address of the rotation angle: these will be loaded on the alternative section of the buffer. The rotation procedure will only have to specify the offset of the specific element of the matrix that has to be accessed, and to tell the data path to combine base and offset to obtain the effective address of the element. The matrices and vectors data are disposed in the dual port RAM such that the effective address can be computed simply putting base and offset in *or*. The same can be done for the rotation angles.
There are three different control units, one for the diagonal cells, the second for the internal cells, and the third for the superior cells. This is due to the different behaviour of the three types of cells.
If the size of the RAM of the control unit is sufficiently large, it would be better to put the states of all the three control units in the same chip. In this way the structure of each cell is the same, and the behaviour will be different depending on some external pins that tell the cell of which type it is (and consequently which is the starting state at the reset signal). Moreover some economy can be made sharing the rotation procedures, that are the same for every kind of cell.
The diagonal cells, when they are in its active phase, have to compute the modules of the eigenvalues on the Z plane: if they are greater than a certain threshold, a *GSD* will be performed, a *SVD* otherwise. After the longest of these operation is concluded, the data and the rotation angles computed in this phase are serialised at those fixed times and transmitted to the neighbouring cells. These ones will be in an inactive phase and, knowing that data will arrive at those fixed instants, they will be ready to receive and deserialise them. The only care that must be taken is avoiding conflicts on the communication channels. The activity of the diagonal cell is the most complex and the most time consuming. The duration of the *GSD* (that is the worst case) determines the overall performance of the systolic array.
The internal cells, when they are active, have to apply the rotations they received from the left hand side and from the bottom on the data they store. After that, they will propagate their data and the rotation angles.
The superior cell has only to apply rotations coming up from the bottom. Then the behaviour differs depending on the position of the cell in the array. Certain cells, in a so called *even* position, will compute the product between the input vectors and the matrices; then they will pass the partial results to the cells on their top. The cells on the top of the array will take these results, now completed, and prepare them to be transferred to the lower cells. The other ones, in *odd* positions, will execute only the previous rotations. Finally, all of them will transmit their data to the neighbouring cells, at fixed times.

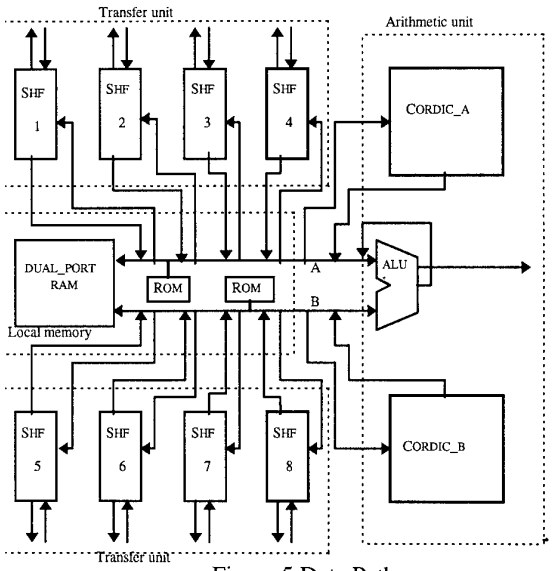


Figure 5 Data Path block diagram

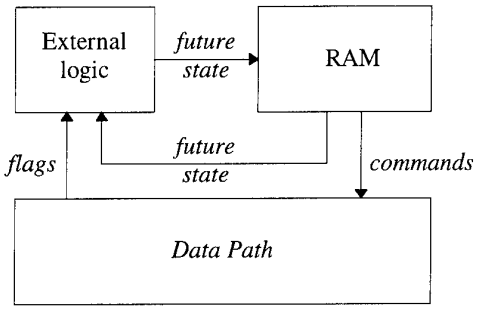


Fig. 6 Functional parts of the operative data path

The Datapath (Fig. 5) is composed of three different functional parts:

- Local Memory, which contains all the data to be elaborated and transferred to the near cells. Besides there are some other fixed values useful during the initialisation and the comparison to choose the GSD or the SVD transformation.
- Arithmetic Unit, which can execute all the operations requested by the Matrix Pencil algorithm.
- Transfer Unit, composed of 8 shift_registers which manage the data transfer from and to the connected cells.

Handled data are represented as 16bit fixed point binary values. A further single bit, used to manage data lying on the second or third quadrant of the complex plane, is associated to each value. Thus the busses size is 17 bits and also Local Memory locations size is the same.
To make the execution of the whole algorithm faster, the architecture of the datapath is structured to enable the execution of two operations at the same time.
So there are two independent Arithmetic Units, connected by two bidirectional busses, (A and B), to the Local Memory, which can read and write on both the busses.
It is important to point out that the busses conflicts are resolved by managing a tristate. In fact every component which can write on them is interfaced by a tristate buffer which is enabled by a control signal.
The whole structure of the operative datapath looks like in Figure 6, where the three functional parts and their elements are put in evidence.
Only the connections among the components are indicated but not the control signals associadet to them. In particular there is an additional element in the structure, that is a simple 44bit register which catches the control word from the Control Unit and distribute it to the other components.
Thus the main elements in the datapath, whose functions are explained with details in the following, are:

- A 89 locations Dual Port RAM;
- Two 5 locations ROM;
- Two CORDIC processors;
- A simple ALU connected to a 17bit Register;
- Eight shift Register
- A 44bit Register.

The main Local Memory component is a Dual Port RAM; it has 89 locations of 17bits and contains the complex values of the two dimensional submatrix to be elaborated and the 24 angles to apply the Givens rotations, all allocated in a particular and fixed order.
This element is connected to the two bidirectional busses A and B by two different ports so that it is possible, simultaneously, to write on the two busses or to put values into two different locations.
Thus there are two enable signals in input as well as two read_write signals and two 7bits address busses. All these signals come from the Control Unit: they are 18 of the above mentioned 44 control signals.
There are two further little memory elements, ROM_A and ROM_B, which contains 5 fixed 17bits values:

- "0";
- "1";
- "-1";
- "π";
- "$\pi/2$";

The first two locations are used to initialize the Dual Port RAM at the beginning of the algorithm.
The third one is useful when a diagonal cell must indicate that it has performed a GSD transformation.
The last ones are used to compute some of the GSD's angles.
These values are duplicated on two ROM'S; each one is connected to a single bus, to make the execution of the operations faster (overall initialization).
Every ROM receives an enable signal and a 3bit address bus.
The transformations to be applied to the incoming data stream are only Givens Rotations or multiplications between real numbers.
This is the reason why it was decided to design a processor which implements parts of the CORDIC arithmetic for the Arithmetic Unit.
CORDIC is a method which makes the hardware implementation of some special operations such as: trigonometrical, exponential, etc. functions quite easy.
There are only three operations which are ever to be executed over the values memorised in each cell to apply the Matrix Pencil:

- inverse tangent calcutation;
- vector rotation;
- product between two real numbers.

They are carried out with the CORDIC Processor.

To obtain the desired results it is necessary to initialize the state variables, (x_0, y_0, z_0) in a proper way, to perform n steps of the iterations and to execute a particular normalization procedure, which involves two further iterative equations similar to the previous ones.
A suitable architecture which can implement these algorithms is shown in Figure 7.

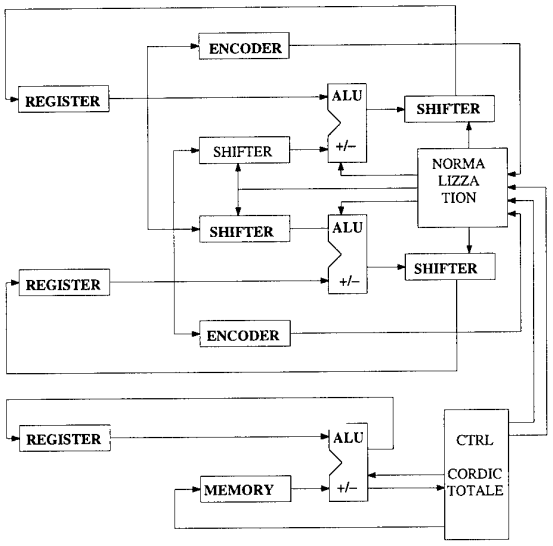


Fig. 7 Block Diagram of CORDIC

All the values are represented in fixed point binary notation.
The special arithmetic unit is composed of:

- three registers containing the state variables;
- two barrel shifters which can execute right shifts from 0 to n positions in one clock period;
- Three elementary ALU, to compute algebric additions;
- A ROM to contain the values of the θ_i angles;
- A control unit.

Some modifications have been introduced to enable the processor to handle vectors in all the four quadrants and to manage the zero-value arguments.
The CORDIC processor is an entity with:

- Six input control signals to program the operation mode and set the status of input parameters.
  - CK: the clock signal;
  - RESET: the reset signal;
  - ENABLE: the activation signal;
  - ARIN and PER: the operation selections;
  - ISIGN: true when the input rotation angle is $> \pi/2$.
- One control signal:
  - PHI1: true when the angle resulting from an inverse tangent calculation is $> \pi/2$ rad in modulus.
- Three 16bits input busses containing the initial values of the state variables;
- Three 16bits output busses containing the final values of the state variables.

Final architecture of the CORDIC processor is shown in Figure 8.

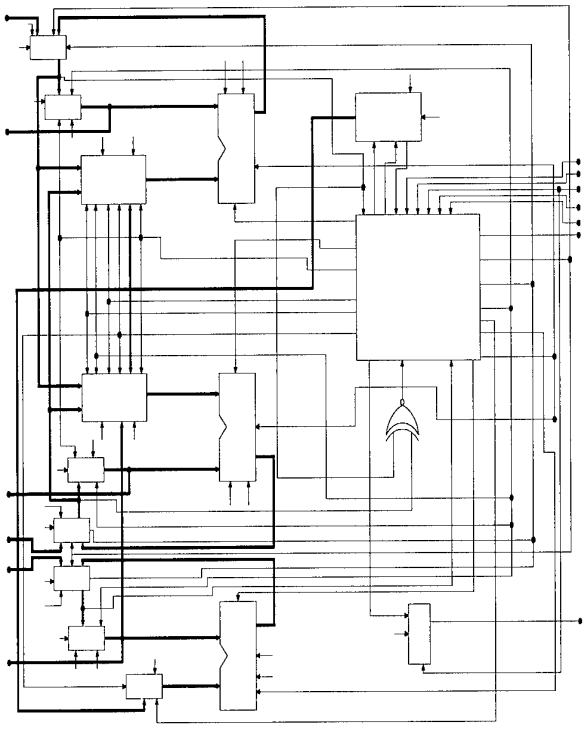


Fig. 8 Final architecture of the CORDIC Processor

It handles 20bits values to minimize round-off and truncation errors.
There is a further simple computational element in he Arithmetic Unit which is devoted to execute some algebric operations requested only during the angles computations. It receives an operand from each bus and put the result on bus A. This component can execute 8 simple different algebrical operations, thus there are three control signals to program the operation mode.
The Transfer Unit is composed of eight shift registers, each of them is connected to an internal bus and to one of the other cells of the systolic array.
They are used to catch a serial value from the autside and make it parallel to be put on the RAM. They are selectively enabled and usually work by couples.
The above described component, except the Dual Port RAM, are enabled in their different functions by a particular address_system. The 17 elements are divided in two parts: the former group is connected to bus A, while the latter group is connected to bus B. In each group each element is associated wich a 4bits address while there are a single enable_signal and a single read_write signal. The combination of these control signals enables the Control Unit to manage the behaviour of the whole datapath. There are some other control signals to specify the operation mode of CORDIC processors, ALU and shift registers so that the control word has 44 bits, as indicated before.
After the creation of a VHDL description and the execution of an accurate simulation, a synthesis tool has been used to perform the compilation of the layout of the over described architecture.
The results concerning area and working frequency are indicated in the following Table.

	AREA (mm^2)	N° TRANSISTOR	FREQ(MHz)
DATAPATH	27	115000	47

References

1. R. Roy and T Kailath " ESPRIT - Estimation of Signal Parameters Via Rotational Invariance Techniques" Optical Engineering, April 1990.

2. F. Vanpoucke, M. Moonen, E. Deprettere "Parallel and Adptive High Resolution Direction Finding" Proc. of SPIE Advanced Signal Processing Algorithm, Architectures and Implementation III, F. T. Luk, July 1992 San Diego - California.

3. Allen-Jan van der Veen, E. Deprettere "Parallel VLSI Matrix Pencil Algorithm for High Resolution Direction Finding". IEEE Transactions on Signal Processing, February 1991.

4. J. Cavallaro, A. C. Elster " A CORDIC Processor Array for the SVD of a Complex Matrix" SVD and Signal Processing II R.J. Vaccaro (Editor) Elsevier Science Pubblication 1991.

5. M. Moonen, P. Van Dooren, Joos VandeWalle "A systolic array for SVD updating" SIAM J. Matrix Anal. Appl., April 1993.

6. F. Vanpoucke , M. Moonen, E. Deprettere "A numerically stable Jacobi array for parallel SVD updating" SPIE Volume 2296/403.

Multispectral Remote Sensing of the Coastal Atmospheric Boundary Layer

C. H. Wash, K. L. Davidson and M. S. Jordan
Department of Meteorology
Naval Postgraduate School
589 Dyer Rd., Room 254
Monterey, CA 93943-5114, USA
T: 408.656.2295 F: 408.656.3061
EMail: wash@nps.navy.mil

SUMMARY

Coastal marine atmospheric boundary layer (MABL) properties derived from satellite data are compared with coastal in situ aircraft and rawinsonde measured properties. A multispectral approach using visible and IR data is tested to estimate indirectly important variables such as depth of the MABL. In situ data are obtained from aircraft, ship mounted systems and from shoreline stations. In the examined cases, remotely sensed information yields reasonable assessments of the height of the top of the boundary layer as well as of conditions immediately above the surface. Remote data describe high resolution horizontal/temporal variations, important in the coastal regions, but not described by point measured data. Limitations of the approach due to sun glint, continental aerosols and more complex MABL structures are discussed also.

1. INTRODUCTION

Knowledge of the characteristics of the marine atmospheric boundary layer (MABL) is critical to large scale investigations of air-sea interactions as well to coastal analysis and prediction. Coastal region MABL structures vary dramatically over short distances, 10-20 km. The height of the MABL is a feature of interest in several applications. For example, it is the location of elevated trapping layers affecting radio/radar waves. It is also the upper limit for mixing volume of pollution arising from shoreline activities. The depth of the MABL may be quite variable in coastal regions because it is influenced by both synoptic scales meteorological and local coastal circulations. In either case, horizontal variations occur in both the height of the MABL and strength of the capping inversion.

Conventional means of in situ measurement are necessarily limited due to the lack of vertical profile measurements over the coastal ocean. Methods using satellite atmospheric soundings are limited by the broad weighting functions resulting in poor resolution of the sharp temperature and moisture gradients at the top of the MABL. Therefore, the further development of satellite-based MABL characterization methods to estimate variations of the MABL, such as its depth and relative moisture content, are valuable both for regional and global scale studies.

This paper describes an indirect method of MABL characterization for clear regions by balancing the estimates of aerosol optical depth and the total water vapor with the relationship between relative humidity and radiative extinction. This method is referred to as the multispectral approach. The method estimates MABL height and moisture variables in the well-mixed boundary layer. First, the technique is presented followed by an evaluation during the Variability of Coastal Atmospheric Refraction (VOCAR) experiment during August/September 1993. Then results from other cases from the Persian Gulf and coastal Central California are discussed.

2. MULTISPECTRAL METHOD

The method described here relies on previously developed techniques for estimating aerosol optical depth and total column water vapor. Aerosol optical depth techniques have been developed by Refs 1, 2 and others. This study uses the red-visible radiance measurements of the NOAA AVHRR (channel 1 & 2) and their direct relation to aerosol optical depth (Ref 2). Also, several techniques have been developed to detect atmospheric water vapor variations from satellite measurements (Refs 3 and 4). The method used here follows Ref 4 that relates total column water vapor to the difference between the split window brightness temperatures from the NOAA AVHRR sensor (channel 4 and 5). Since both estimates are derived from the same sensor, the boundary layer estimates described below can be derived from a single data source.

Three assumptions about the marine boundary layer and vertical profile of extinction and water vapor are needed to relate optical depth and column water vapor to MABL depth and relative humidity. First is that the MABL is well-mixed. Second, the only significant contribution to aerosol optical depth results from extinction by aerosols in the MABL. Finally, the

Paper presented at the AGARD SPP Symposium on "Remote Sensing: A Valuable Source of Information", held in Toulouse, France from 22-25 April 1996, and published in CP-582.

percentage of water vapor in the MABL can be estimated from in situ data or other satellite measurements.

Within the well-mixed MABL, potential temperature and specific humidity/mixing ratio tend to be constant with height. Adiabatic mixing due to buoyancy and wind shear effects, described by Ref 5 and others, is responsible for maintaining these well-mixed profiles. Fig. 1 illustrates the evolution of a well-mixed boundary layer using a time series of ship-measured rawinsonde soundings off the Southern California coast. Note the nearly constant potential temperature in the boundary layer. Mixing ratio (not shown) is also nearly constant with height in the MABL. The multispectral technique takes advantage of these simple distributions and is also based on the well-known relationships between temperature, saturation vapor density and relative humidity.

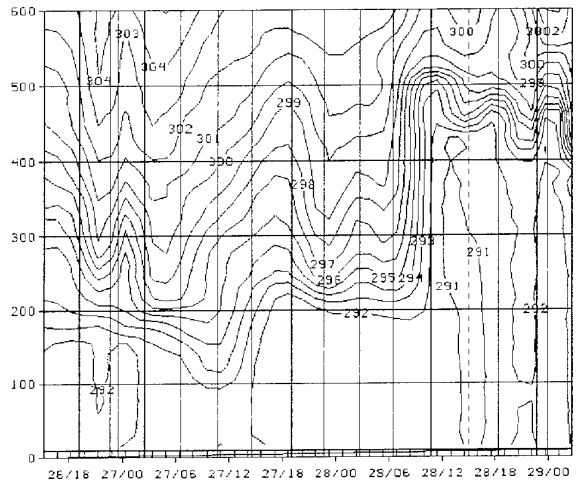


Fig. 1. Time series of MABL evolution using 17 rawinsondes off the Southern California coast, 26-29 August 1993. Contours of virtual potential temperature (K) illustrate the height (m) of the MABL.

The technique further assumes that aerosol optical depth at the red-visible and near-infrared wavelengths' results from particles that are confined primarily within the MABL. If this assumption is true, then the optical depth in the red-visible would be due to the integrated effects of extinction due to these aerosols in the MABL. Ref 6 describes atmosphere aerosol distributions and concludes that for marine particles, absorption is small and the extinction is due to scattering. This extinction coefficient is a function of the cross sectional area for a given particle radius, the extinction efficiency (dependent on the complex refraction index and particle radius), and the distribution of particles by radius. Variations in each of these three factors produce corresponding changes in extinction. Ref 7 showed that the dominant term affecting extinction is particle size. Ref (2) developed a relationship between extinction and relative humidity consistent with Ref 7. This relationship is based on aircraft measurements of extinction within the MABL off the southern California coast in 1982.

Satellite estimates of optical depth and column water vapor are both related to the MABL height and moisture. An iteration method was devised to solve for the MABL height and surface relative humidity for clear regions in a satellite pass using the AVHRR data, Refs 8 and 9. Sensitivity estimates using a model atmosphere indicated the method is reliable when the MABL satisfies the assumptions of the technique, Ref 9.

3. VOCAR RESULTS

The multispectral approach has been successfully applied to MABL analyses over coastal regions of Northern California and Southern California (Ref 10 & 11) and the Persian Gulf (Ref 12). In this paper we will present a validation of the technique using a case study during the Variability of Coastal Atmospheric Refractivity (VOCAR) experiment, off the southern California coast, in August-September 1993 using newly available aircraft from Naval Weapons Test Center, Point Mugu, CA. The first VOCAR results are contained in the thesis, Ref 10, and conference presentations at last AGARD meeting, Ref 11.

To use the VOCAR period to test the technique, an estimate of the percentage of the water vapor in the MABL was used. This is an important improvement to the technique for climatological regions where significant water vapor is found above the marine inversion.

The VOCAR period case presented here was chosen because it was a clear day in the VOCAR region and rawinsondes were launched within 30 minutes of the satellite pass time at seven locations (one ship, two island, 4 coastal locations). In addition, aircraft data are now available providing low-level temperature and moisture profiles between the coast and San Nicolas (NSI) and San Clemente (NUC) islands within two hours of the satellite overpass.

In this case, there is significant mid-tropospheric water vapor. We computed the percentage of water vapor density in the MABL, compared with the total from the

surface to 400 mb, for all rawinsonde locations. The percentage, which is representative of the MABL throughout the region, is 14% for this case. This case indicates that the multispectral technique will work when significant amount of mid-tropospheric water vapor is present.

Fig. 2 shows the multispectral MABL depth analysis for the 23:47 UT 26 August 1993 (4:47 pm local) NOAA AVHRR satellite pass. The dark areas indicate that the pixels were over land, are clouds, or the iterative method did not converge at that point. The rawinsonde observations are indicated by yellow dots on the image while aircraft profiles are denoted by white dots.

Fig. 2 shows the slope of the MABL, from shallow values to the West (75 m) to deeper values along the eastern portion of the coastal region (225 to 275 m). The thicker MABL region is associated with warmer sea surface temperatures (SST) along the California coast from Point Vincente (PVN) to North Island (NZY) near San Diego. The range of SST is from 23 to 24 C in the warmer area to 16 to 18 C south of Point Arguello (PDR) along the eastern edge of the California Current.

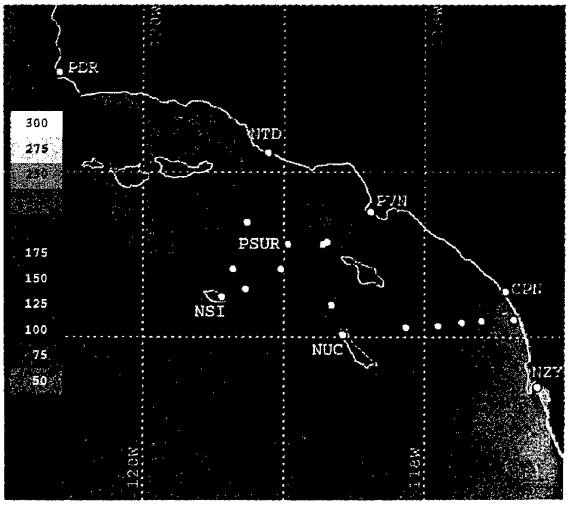


Fig. 2. Multispectral MABL depth (m) for 23:47 UT 26 August 1993 NOAA AVHRR satellite pass. VOCAR rawinsonde observations are devoted by yellow dots and aircraft profiles are given by white dots. The dark areas indicate the pixels were over land, are clouds, or the iterative method did not converge at that point.

Evidence confirming these coastal SST and MABL structures are found in the rawinsonde and aircraft data. The only overwater rawinsonde is from the R/V Point Sur (PSUR) and it's measured depth of 148 m compares favorably with the shallow 127 m satellite estimate.

A comparison of aircraft temperature and moisture profiles from the Point Mugu aircraft with the satellite assessment is presented in Table 1 and Figs. 2 and 3. The aircraft flew a vertical saw-tooth flight track between 75 m and 825 m penetrating the mixed layer. Twelve measurements from aircraft descents and ascents along flight legs between San Clemente Island and south of Camp Pendleton (CPN) and between San Nicolas Island (NSI) and Point Vincente (PVN) are used. Of particular interest is the aircraft leg of the distinct gradient in MABL height and SST west of San Diego. Two aircraft soundings in the lee of San Nicolas Island have been excluded from the analysis as the MABL appears influenced by local island effects.

The aircraft soundings confirm the slope of the MABL depth indicated by the satellite data. The satellite data underestimates the depth by 44 m but the RMS differences are less that 50 m. The aircraft temperature of the mixed layer fall within 1 C of the satellite estimated SSTs. Scatterplots of aircraft vs satellite MABL depth (Fig. 3) and temperature (Fig. 4) confirm the good correlation between the two datasets.

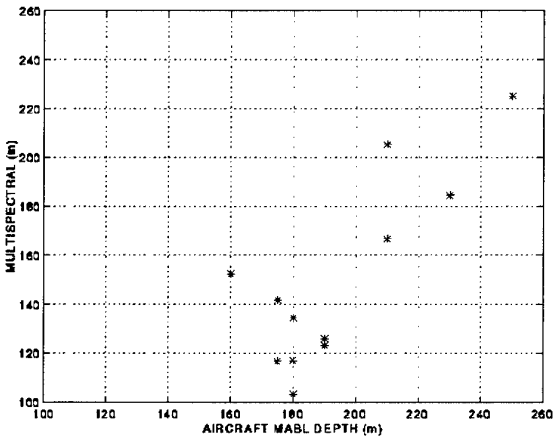


Fig. 3. Scatterplot of aircraft-measured and multispectral satellite estimates of MABL depths (m).

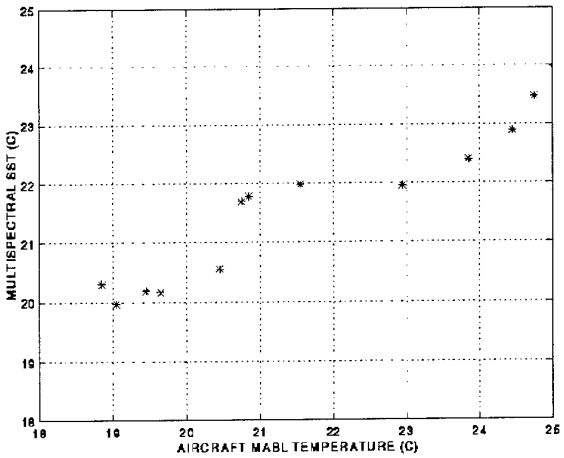


Fig. 4. Scatterplot of aircraft-measured MABL temperatures and satellite estimates of SST (C).

Table 1.
Comparison of Aircraft versus Multispectral MABL Height

BIAS	44.4 m
RMS differences	49.6 m
Correlation Coefficient	.809

Table 2.
Comparison of Aircraft MABL Temperature versus Satellite SST

BIAS	-0.07 C
RMS differences	1.04 C
Correlation Coefficient	0.944

4. OTHER VALIDATION CASES AND METHOD LIMITATIONS

The multispectral technique has been applied to several other coastal data sets (Ref 12). The results from these cases are summarized here and presented in more detail at the conference.

The SHAREM 110 exercise was held in the Persian Gulf and Gulf of Oman during the period of 6 to 18 February 1995. Rawinsondes were available from three ships during the entire period while aircraft data were available on five days. Due to the geography of the Persian Gulf, the atmosphere is never completely free from land influences. However, a Shamel occurred during the experiment with a period of moderate northwest winds from the Northwest along the axis of the Gulf. This flow with a long overwater fetch satisfied the conditions of the technique. A deep well-mixed MABL developed during this period (10 February). The MABL depth was measured to be approximately 1000 meters with most (83%) of the column water vapor in the MABL.

Table 3 presents an evaluation for this case from soundings from the USS David R. Ray and USNS Silas Bent. The difference between the sounding and satellite estimate was near 50 meters and the boundary layer slope was correctly determined by the satellite analysis. The satellite MABL analyses also was consistent with the cloud top temperature of some boundary layer cumulus over the western part of the Gulf. The multispectral technique proved to be very successful in this case even with considerably thicker MABLs than observed in VOCAR.

Table 3
MABL Height Comparisons for the Persian Gulf
10 February 1995

	D.R. Ray	Silas Bent
Multispectral	1005 m	1042 m
Rawinsonde	1055 m	1090 m
Difference	50 m	48 m

Other SHAREM and coastal California cases illustrate limitations to the approach. These limitations are sun glint, continental aerosols and more complex MABL structures. With polar-orbiting sun synchronous satellites, sun glint can be a major difficulty. In conducting these validation studies, approximately 30% of the clear (otherwise usable) passes can not be analyzed because of the solar reflection from the sea surface. This problem may be less severe when using data from the new US NOAA geostationary satellites, GOES-8 and 9.

In one of the SHAREM cases, strong off-shore continental flow with significant dust caused plumes of elevated optical depth. The multispectral technique did converge in this case and produced MABL analyses. However, the MABL depth was hundreds of the meters in error due to the continental influences present. This case illustrates that situations with significant off-shore transport of aerosols need to examined closely. The plume-type structure of the optical depth and MABL height analyses give strong indications of the impact of the continental aerosols.

Finally, cases with multiple mixed layer structure and elevated trapping layers have been found. This is particularly common over regions undergoing cold upwelling events along the west coast of continents. Here a new mixed layer is formed over the cold water,

distinctly colder than the mixed layer above it. In these situations, the multispectral approach does successful map the lower mixed layer, but does not provide any information about the inversion structure above the lowest mixed layer. In these cases, the technique can not provide any information about elevated inversions or refractive layers that are not associated with the surface but may be tactically relevant to electromagnetic-electro-optical propagation and other applications.

5. CONCLUSIONS

We have described the evaluation of remotely sensed estimates of mixed layer structure at and above the ocean surface in coastal regions using aircraft data during VOCAR and from other field experiments. Both successful and unsuccessful cases have been studied. The multispectral approach has accurately mapped the topography of the MABL, SST and optical depth conditions for cases from coastal Southern California, Persian Gulf and Central California coast. The comparisons with aircraft and rawinsondes show the satellite data to correctly describe the MABL topography and agree to within 20% with in situ observations. The method can be applied to all cases of clear skies. However, sun glint, continental aerosols and situations with complex MABL structures can not be analyzed with this approach using polar-orbiting satellite data. The sun glint problem should be less severe with the new US geostationary satellites. Future studies will evaluate this source of multispectral data for MABL studies.

6. ACKNOWLEDGMENTS

We thanks the NPS staff who contributed to the in situ VOCAR data collection and analyses (K. Jones, P. Frederickson, T. Neta). Special thanks goes to C. Skupniewicz, NPS, for his assistance in using the multispectral process code. Roger Helvey of Pacific Missile Test Center, Point Mugu provided the aircraft data and was of great assistance in its interpretation and use.

7. REFERENCES

1. Griggs, M., "Satellite Measurements of Tropospheric Aerosols," *Adv. Space. Res.*, vol. 2, pp. 109-118, 1983.

2. Durkee, P.A., D.R. Jensen, E.E. Hindman and T.H. VonderHaar, "The Relationship Between Marine Aerosol Particles and Satellite-Detected Radiance," *J. Geophys. Res.*, vol. 91, pp. 4063-4072, 1986.

3. Prabhakara, C.G., G. Dalu, R. Lo and N. Nath, "Remote Sensing of Seasonal Distributions of Precipitable Water Vapor over the Oceans and the Inference of Boundary Layer Structure," *Mon Wea Rev.*, 107, 1388-1401, 1979.

4. Dalu, G., "Satellite Remote Sensing of Atmospheric Water Vapor," *Intl. J. Remote Sensing*, vol. 7, 1089-1097, 1986.

5. Rogers, R.R., "A Short Course in Cloud Physics," Pergamon Press, New York, NY, 235 pp., 1979.

6. Shettle, E.P. and R.W. Fenn, "Models are the Aerosols of the Lower Atmosphere and the Effects of Humidity Variations on their Optical Properties," AFGL TR-79-0214, Air Force Geophysics Laboratories, Hanscom AFB, MA, 94 pp., 1979.

7. Fitzgerald, J.W., W.A. Hoppel and M.A. Vietti, "The Size and Scattering Coefficient of Urban Aerosol Particles at Washington, DC as a function of Relative Humidity," *J. Atmos Sci*, 39, 1838-1852, 1982

8. Kren, R.J., "Estimation of Marine Boundary Layer Depth and Relative Humidity with Multispectral Satellite Measurements," Masters Thesis, Naval Postgraduate School, Monterey, CA, 73 pp., June 1987.

9. Smolinski, S.P., "Marine boundary layer depth and relative humidity estimates using multispectral satellite measurements," Masters Thesis, Naval Postgraduate School, Monterey, CA, 70 pp., March 1988.

10. Walsh, D.J., "Multispectral NOAA Marine Atmospheric Boundary Layer (MABL) Estimates During VOCAR," Masters Thesis, Naval Postgraduate School, Monterey, CA, 106 pp., June 1994.

11. Davidson, K.D. and C. H. Wash, "Remote Measurement of Atmospheric Refraction Conditions in the Coastal Region," AGARD-94, Propagation Assessment in Coastal Environments, Bremerhaven, Germany, Sept 1994.

12. Teadt, T.L., "Studies in Satellite Multispectral Analyses of MABL Depth," Masters Thesis, Naval Postgraduate School, Monterey, CA, 55 pp., March 1996.

Paper 45

T.Wahl (NO)

Does your pixel-by-pixel method converge well to a "smooth" surface, or is it necessary with much filtering?

Author's reply:

There is no filtering in the method. The MABL depth retrieval varies in the order of 10-30 meters pixel to pixel, giving a relatively "smooth" depth estimate.

R. Philbrick (US)

1. You indicated that you have observed a bias in the boundary layer thickness in the VOCAR data with the retrieved boundary layer being too thin by about 50 m. Do you know why that bias exists?

2. You indicated some large differences in the optical depth when the source for the aerosols was from the land. Did you collect samples or measure the index of refraction of the particles?

Author's reply:

1. We believe our estimate of the fraction of water vapour in MABL is an underestimation of the true value and leads to an estimated MABL that is too shallow. There was a case where there was significant water vapour above the MABL.

2. No samples were available, unfortunately.

Use of Remote Sensing for Monitoring The Lagoon Köyceğiz And Its Environment in Turkey

D. Maktav[1]
F. Sunar[1]
S. Kapdaşlı[2]
N. Musaoğlu[1]

Istanbul Technical University, Faculty of Civil Engineering
Department of Remote Sensing[1]
Department of Hydraulics[2]
80626 Ayazağa, Istanbul, Turkey

SUMMARY

The Köyceğiz Lagoon in Turkey is a part of an area declared as Specially Protected Area. The goal of this study was to monitor this lagoon by integrating conventional water quality data and satellite remote sensing data.

1. INTRODUCTION

The need for better monitoring of the lake and lagoon environment has increased in recent years, in particular along coastal boundaries (Johannessen *et al.*, 1993; Johnson *et al.*, 1980). The objective of this study was to use Landsat Thematic Mapper digital data combined with water quality data for monitoring the Lagoon Köyceğiz at the Mediterranean Region in Turkey. The water quality parameters included salinity, turbidity, temperature and oxygen (Verdin, 1985). Regression models were developed between the water quality parameter measurements from boat and the Landsat Thematic Mapper data. The digitally processed satellite images were interpreted from coastal engineering point of view.

2. DATA AND METHODS USED

2.1 Water Quality Measurements

Landsat 5 Thematic Mapper data in seven wavebands at a 30-m spatial resolution (thermal : 120 m) on 8 August 1984, on 19 August 1988, on 27 July 1991 and on 11 November 1995 were used. A single channel Raystar 390 GPS Receiver for navigators with the memory capacity to store 300 waypoints, for measuring the coordinates of the selected stations in the lagoon was used. Microprocessor Oxygen-meter OXI 196 was used for oxygen measurements. Salinity and temperature measurements were made by Microprocessor Conductivity Meter LF 196. Depth measurements were accomplished with the FMV-603 Video Sounder which determines the distance between its transducer and underwater objects and displays the results on a screen. For turbidity measurement the Secchi Disk was used. Also, a topographic map at a scale of 1: 25,000 was utilized.

2.2 Remote Sensing Methods

For classification of the remotely sensed data, a supervised classification method is used where classes are defined using sample areas of an image before classification and then similar pixels are allocated to each class. A maximum likelihood discriminant function, which basically allocates pixels to the class to which they have the highest probability of belonging, was used. This approach assumes that the real classes to discriminate are represented in the image data as normal distributions around class mean values. The form of each normal distribution is determined by mean and variance/covariance of the class. Allocation of pixels to classes involves using class distribution statistics as a probability density function for the class which indicates the likelihood of a particular pixel value belonging to the class (Swain *et al*, 1978; Harrison *et al*, 1990). Also, the image is rectified for removing geometric distortions from remotely sensed data and to match the geometry of the satellite image to the geometry of the map at a scale of 1:25,000.

The general approach of the study involved was to analyze the measurements of several water quality parameters taken near simultaneously with the Landsat overpass (on 11 November 1995). Correlations between the water quality parameters and either raw Landsat band (also different band combinations) DNs or calibrated radiances were determined. Regression models were developed to establish the relationship between Landsat spectral data (raw and calibrated radiances) and water quality measurements at the selected stations. Digital numbers (DNs) for the six reflective and one thermal band were extracted for single pixels. The land/water interface was determined by masking land values obtained from clustering to apply the statistical models.

Paper presented at the AGARD SPP Symposium on "Remote Sensing: A Valuable Source of Information", held in Toulouse, France from 22-25 April 1996, and published in CP-582.

3. STUDY AREA

The Köyceğiz Lagoon lies in the southwestern part of Turkey very near to the county of Köyceğiz in the province of Muğla. Its water surface area encompasses more than 51 sq. kms. It is approximately 12 kms long and 5 kms wide and contains more than 30 kms of shoreline. The connection of this lagoon to the Mediterranean Sea is via a 10 km long channel and the Wetland Dalyan which provides critical food and habitat for a wide variety of fish and wildlife (Maktav *et al.*, 1994a,b). Therefore, the lake is in a type of lagoon. Its eastern coasts mostly are covered with approximately 2-3 m high (from water surface) reeds, and most of the freshwater entering the lagoon comes from the mountains. However, it is also fed by underground water.

The lagoon is a part of an area declared as Specially Protected Area (Köyceğiz-Dalyan) on July 1988 and the National Park on 1996 (Fig. 1). The aim of the declaration was nature conservation and protection of environmental values (TÇV, 1993). Today, this protection area is one of the Turkey's most valuable natural treasures.

4. METHODOLOGY

To monitor the Lagoon Köyceğiz, water quality data, satellite remote sensing data and digital image processing methods were used. Water quality data in the lagoon were obtained near simultaneously with the satellite data at 37 stations in the whole protection area, but only 20 stations were taken into consideration for the Köyceğiz Lagoon (Fig. 1). These data were salinity, oxygen, temperature, turbidity and depth (although not a water quality parameter). They were measured using the instruments given in the previous paragraphs. The locations of the lagoon station points were determined by GPS measurements. Later, the GPS coordinates were transformed to UTM coordinate system based on WGS-84 ellipsoide and to UTM system based on Hayford ellipsoid, successively.

As a preliminary step, plots of each water quality parameter versus each band (or selected band combination) were produced to analyze if any trends among them were evident (Fig. 2). Correlations between the parameters and the seven TM band values for each station were determined. The general approach involved development of linear model (univariate or multivariate) based upon the spectral characteristics of the various parameters. The TM band values or band combination values were used as *independent* variables in the models. Water quality parameters were the *dependent* variables in the model. Selection of the band values to use initially for model development was based on plots, correlations, and the physical relationship between the respective parameters and the spectral values (Khorram *et al.*, 1991). The overall appropriateness of each model was evaluated by the R-square value (a measure of the amount of total variation), by the significance of the F- and t-tests (single-tailed tests, Alpha value of 0.05). Secchi Disk depth was the most easily modeled parameter. This would seem reasonable since it is very closely related with the spectral characteristics of the water bodies (Lathrop *et al.*, 1992; Estep, 1993). No satisfactory results were obtained between spectral data and salinity or oxygen or depth parameters.

The following model was found to be the best predictor of the natural log of Secchi Disk depth:

$$\mathbf{LNSECCHI} = 39.612 + 0.78503 \ln(\mathbf{B2}) + 1.07421 \ln(\mathbf{B3})$$

$$\mathbf{n} = 19 \quad \mathbf{R^2} = 0.43$$

The F- and t- tests results showed that the two variables had an absolute value greater than the test-critic values. Therefore, two variables used in the regression model were accepted useful in predicting the Secchi Disk values in the lagoon.

The natural-log transformation was not used for calibration of the temperature model. The model is:

$$\mathbf{TEMP} = -15.042 + 0.2861\ \mathbf{B6}$$

$$\mathbf{n} = 20 \quad \mathbf{R^2} = 0.65$$

For the temperature, radiance was used (gave better results for regression) whereas for the other parameters, raw data was used.

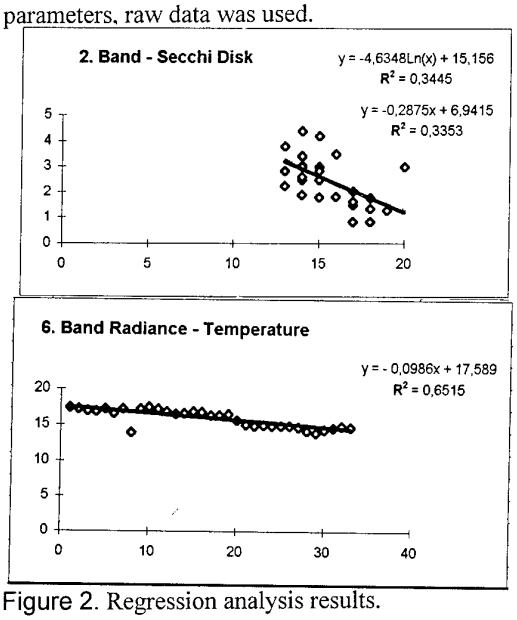


Figure 2. Regression analysis results.

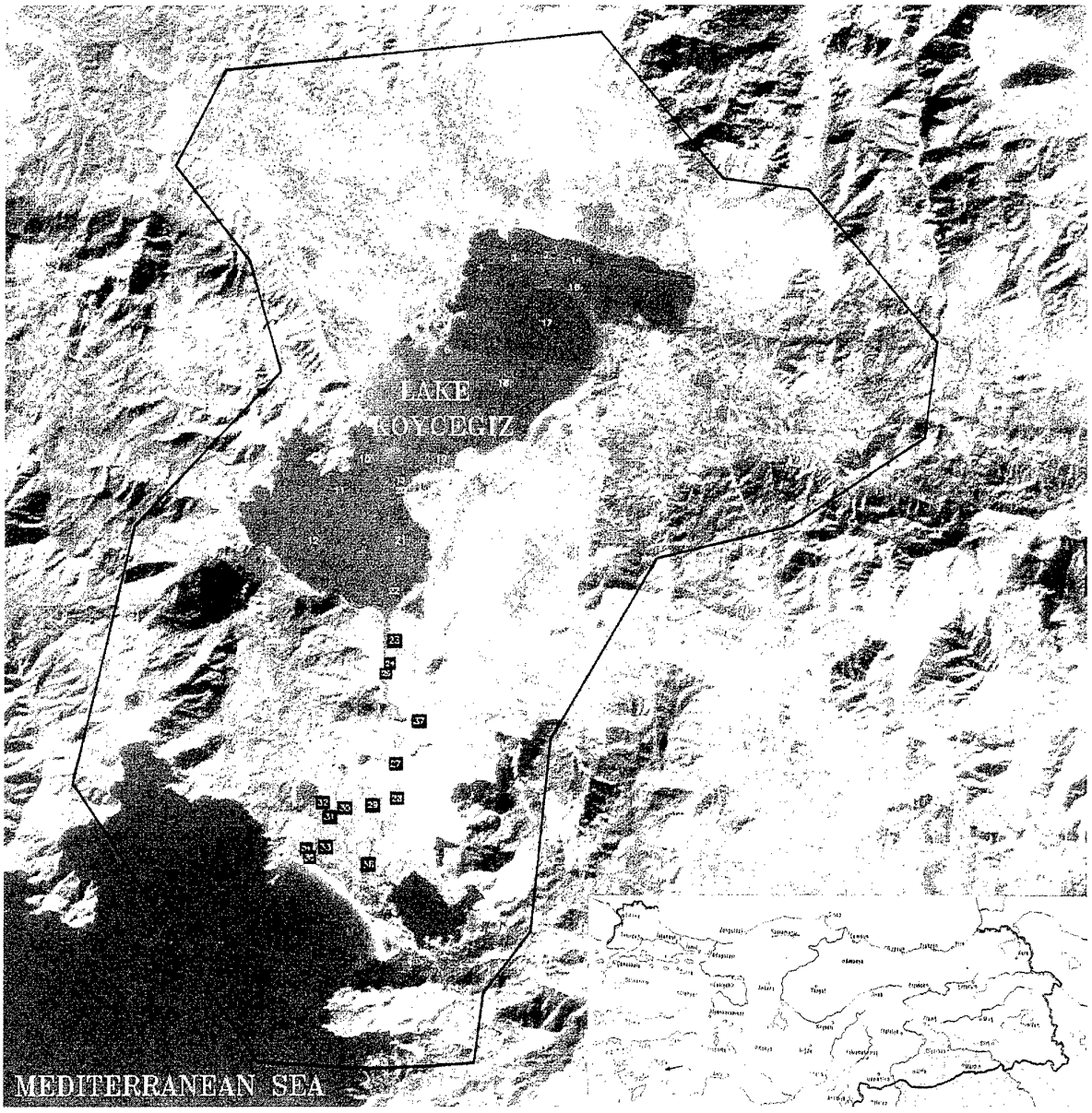


Figure 1. The Köyceğiz Lagoon as a part of the Specially Protected Area (36° 50' N and 28° 40' E).

5. DISCUSSION

The distribution and stability of a lake or lagoon ecosystem, such as of the Köyceğiz Lagoon, depend on the physical characteristics of the water: salinity, temperature, oxygen, turbidity, circulation and depth. Each affects and affected by the others.

Salinity given in number of grams of dissolved salts in 1 liter of water is a key factor affecting the physical make-up of a lagoon and increases with depth. Therefore, freshwater tends to remain at the surface. If the lagoon has some groundwater sources, the salinity gradients may be very small and the salinity distribution may remain uniform. Salinity measurements within the Lagoon Köyceğiz showed that the salinity over the depth and the area does not vary widely (2.9-3.1 mg/lt), because in addition to the feeding river waters the lagoon is also fed by groundwater and the mixing process is quite strong due to the wind effects (Table 1).

Temperature dramatically changes the rate of chemical and biological reactions within the water. Because the lagoon is generally measured shallow at the coasts (2-5 m, except some points in the middle 10 or 30 m), its capacity to store heat over time is relatively small. As a result, lagoon water temperature fluctuates throughout the year. These changes in water temperature influence when plants, such as reeds, cattails, sedges and animals, such as carp, bream, different species of birds (gray heron, sanderling, coot, stork etc.) feed, move locally or migrate.

The water temperature measurements within the lagoon showed that the temperature profile is fairly constant (16-17 C°). Dissolved oxygen contents were also approximately constant (8.5-10 mg/lt). These results showed that the mixing process between the higher and lower water bodies is highly strong causing the heat and oxygen transfer from water surface to the bottom.

Circulation of lagoon water transports plankton, fish eggs, sediments, dissolved oxygen, minerals and nutrients mostly throughout the northern part of the Köyceğiz Lagoon. Circulation is driven, primarily, by the movements of lagoon water from the south and the west, and causes nutrients and sediments to be mixed and resuspended. This mixing creates a zone of maximum turbidity that, due to the amount of available nutrients which all living organisms need for growth and reproduction, is often used as a nursery area for fish and other organisms, such as the reeds (Fig. 4a). Also, the supervised classification results of the images taken at four different years show the distribution of these turbid areas (Fig. 3). The main causes of the circulation in the Köyceğiz Lagoon are the rivers Değirmen, Çakmak etc. which arrive at the lagoon at the western coasts, and the northerly winds blown through the valley of the river which is the significant wind direction in this area. Therefore, the river water with % 0.1 salinity is forced to move towards the eastern side of the lagoon by the wind effect.

Thus, wind plays an important role in the mixing of the lagoon's surface water. Regarding the water quality measurements done in the lagoon, the turbid areas seen in the masked satellite image are the sediments transported one day ago by a very strong wind (Fig. 4b).

Probably, the most interesting result is that the main circulation route which is in the north half of the lagoon, occurs in opposite clockwise direction, namely from south to north (Fig. 4b). This is almost opposite the wind direction and is an unusual circulation in general. This result also shows that any point- or non point pollution well away the town Köyceğiz can affect the town. Additionally, the pollution coming from the town Köyceğiz will not reach the southern part of the lagoon and will return.

6. CONCLUSIONS

The distribution of water quality parameters depicted in the color coded maps follow the expected distributions of the variables. Landsat Thematic Mapper data proved useful in modelling selected water quality parameters.

The use of a ratio of bands does not improve the strength of the correlation over the use of single bands.

In general, the integration of the remote sensing technologies with the conventional hydrographic measurements for environmental research purposes gives very valuable information which can be only obtained by making long-term in-situ measurements that are time consuming and very expensive. It has been concluded that satellite remote sensing data can be used successfully in monitoring the lakes or lagoons.

ACKNOWLEDGMENTS

This research is a part of a still ongoing project "Investigation of the Hydrology of the Lagoon Köyceğiz Using Remote Sensing Methods". The authors would like to thank to the Istanbul Technical University Research Fund who fully supported the project.

Table 1. Water quality measurements.

Station #	Latitude	Longitude	Salinity (mg/lt)			Temperature (C°)			Oxygen (mg/lt) %			Depth (m)	Secchi disk (m)	Time	Explanations
			from surface	from 4m.	from 5m.	from surface	from 4m.	from 5m.	from surface	from 4m.	from 5m.				
1	36 57.40	28 41.21	2.9			17.4			9.1			4.2	3.00	13.35	
2	57.35	40.6	2.9			17.2			8.7			5.0	2.50	13.5	
3	57.23	39.94	2.9			17.0			8.6			5.4	3.00	14.05	Reeds
4	56.48	39.6	2.9			16.9			8.5			4.9	3.50	14.23	
5	55.95	39.25	3.0			17.2			8.8			1.9	1.90	14.31	Reeds
6	55.62	38.81	2.9			16.6			8.6			5.3	1.80	14.4	Turbid Bay
7	55.27	37.77	3.0			17.2			8.5			3.9	0.85	14.56	In the river mouth
8	55.32	37.69	1.1			13.9			9.5			2.0	0.85	15.00	In the river mouth
9	54.26	37.61	3.0			17.2			9.5			3.3	3.05	15.15	
10	53.80	37.11	3.1			17.5			9.5			-	-	15.25	No currents
11	53.04	36.56	3.1			17.2			9.6			31.6	3.40	15.45	
12	53.90	38.50	3.0			16.9			9.9			2.0	3.00	16.12	
13	57.30	41.76	2.9			16.5			10.0			11.1	2.85	16.21	
14	56.89	41.69	2.9			16.6			9.8			20.0	2.85	9.58	
15	56.32	41.18	3.0	3.0		16.9	16.9		9.1	8.9		25.0	3.80	10.07	
16	55.41	40.31	3.0		3.0	16.8		16.6	9.0		8.7	22.9	3.42	10.26	
17	54.25	39.07	3.0	3.0		16.4	15.8		10.4	10.3		5.2	4.40	10.46	
18	53.78	38.54	3.0			16.4			10.4			2.5	2.25	10.46	Reeds
19	53.00	38.25	3.0		3.0	16.5		16.3	9.3		9.3	5.3	4.20	11.14	
20	52.23	38.12	2.9			15.7			9.4			4.3	2.85	11.25	Backside of the channel

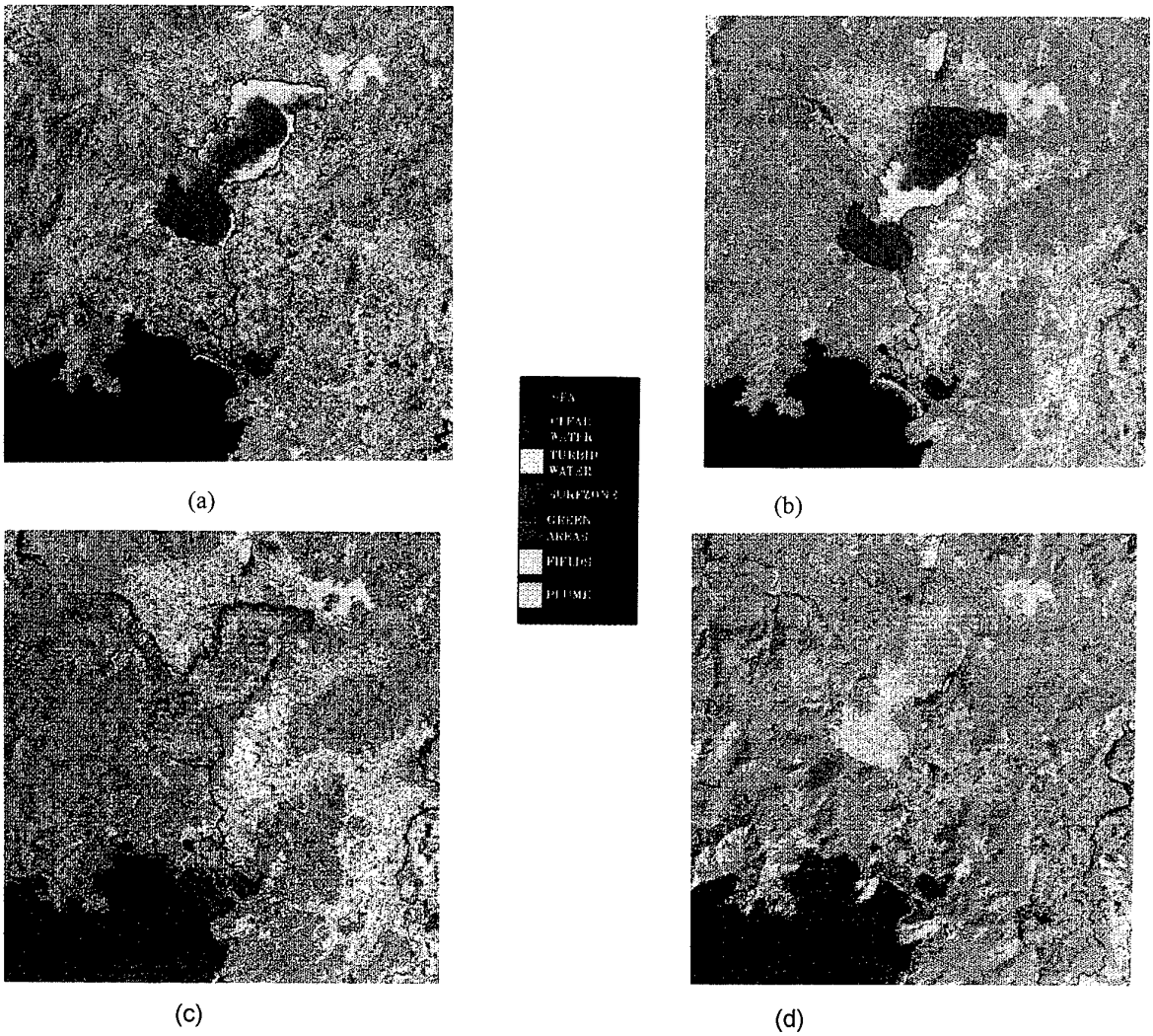
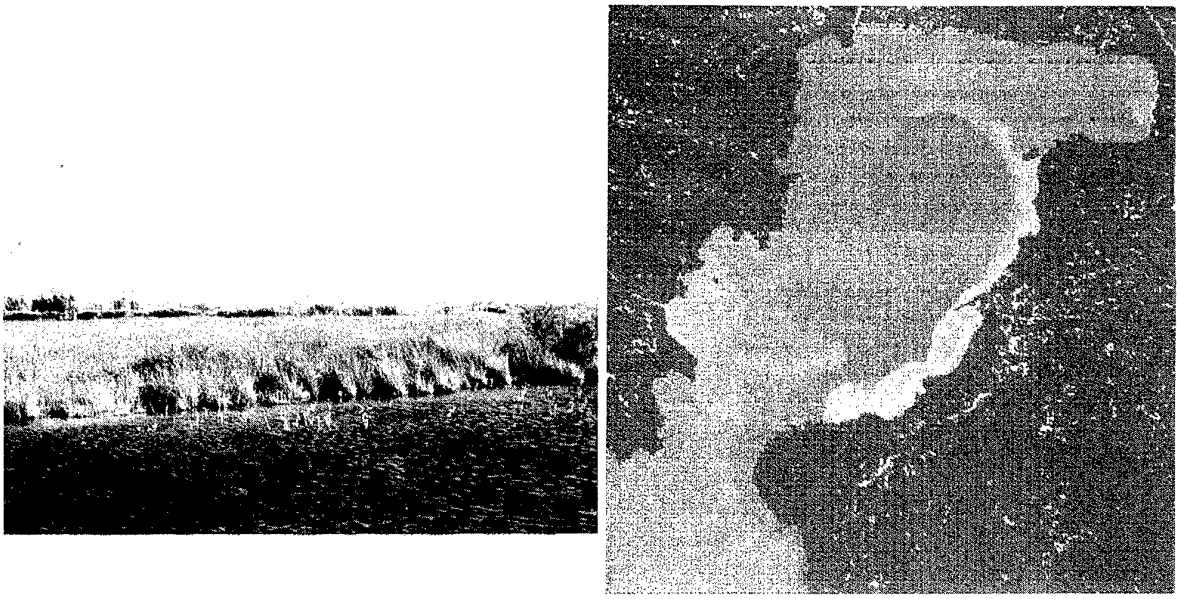


Figure 3. Multitemporal classification results showing the circulation of the turbidity (Landsat TM images at the dates, (a) August 1984, (b) August 1988, (c) July 1991, (d) November 1995).



(a) (b)

Figure 4. (a) Photograph of reeds in the lagoon taken with an amateur camera.
(b) The masked Landsat TM image showing the circulation in the Lagoon Köyceğiz (11 November 1995).

REFERENCES

Johannessen, J.A., Roed, L.P., Johannessen, O.M., Evensen, G., Hackett, B., Pettersson, L.H., Haugan, P.M., Sandven, S., Shuchman, R., "Monitoring and Modelling of the Marine Coastal Environment", *Photogrammetric Engineering and Remote Sensing*, Vol. 59, No 3, 1993, pp.351-361.

Johnson, W. R., Harriss, C., R., "Remote Sensing for Water Quality and Biological Measurements in Coastal Waters", *Photogrammetric Engineering and Remote Sensing*, Vol. 46, No 1, 1980, pp.77-85.

Verdin, J. P., "Monitoring Water Quality Conditions in a Large Western Reservoir with Landsat Imagery", *Photogrammetric Engineering and Remote Sensing*, Vol. 51, No 3, 1985, pp.343-353.

Swain, P. H., Davis, M. S., "*Remote Sensing : The Quantitative Approach*", McGraw Hill, USA, 1978.

Harrison, B. A., Jupp, D. L. B., "*Introduction to Image Processing*", CSIRO, 1990, Australia.

Maktav, D., Kapdaşlı, S., "Monitoring of Coastal Processes at İztuzu in Turkey with the Help of Satellite Remote Sensing Methods", *Marine Technology Society Journal*, Vol. 28, No 2, USA, 1994a, pp.53-56.

Maktav, D., Kapdaşlı, S., "An Engineering of Landsat Data to Wetland Investigations in Turkey, *10th. Thematic Conference on Geologic Remote Sensing*, San Antonio, Texas, USA, 1994b, pp. 515-520.

TÇV, "*Wetlands of Turkey*", TÇV (Turkish Environmental Foundation), Önder Publication, Turkey, 1993.

Khorram, S., Cheshire H., Geraci, A.L., Rosa, G., "Water Quality Mapping of Augusto Bay, Italy from Landsat TM Data", Int. Journal of Remote Sensing, 1991, Vol.12, No. 4, pp. 803-808.

Lathrop, G.R.,"Landsat Thematic Mapper Monitoring of Turbid Water Quality", *Photogrammetric Engineering and Remote Sensing*, Vol. 58, No 4, 1992, pp.465-470.

Estep, L., Arnone, R., "Correlation of CZCS Surface Ks with Ks Derived from Secchi Disk", *Photogrammetric Engineering and Remote Sensing*, Vol. 59, No 3, 1993, pp.345-350.

Synoptic Variability Revealed by Satellite and Equivalent Altitude

Paul Greiman, Jay Rosenthal and Roger Helvey
Geophysics Branch, Code 521400E
Naval Air Warfare Center Weapons Division
Point Mugu, California 93042-5001
USA

1. SUMMARY

Both real time assessments of EM propagation conditions, (in the absence of direct measurements), and inferential techniques of forecasting duct occurrence, height, and intensity require the a priori knowledge of the relation between these duct parameters and the standard weather variables which describe large and mesoscale air mass features. Through pattern recognition of stratus/stratocumulus cloud features on meteorological satellite imagery, and by identification of air mass sectors and wind circulations, it is possible to estimate the present and future state of the radar propagation environment by first determining the spatial and temporal variability of inversion conditions, weather map features, and cloud features on weather satellite imagery.

Quantitative and largely automated approaches of using satellite imagery over the sub-tropical ocean areas are particularly useful in measuring refractive variability, but simple techniques of using grid field data at two altitudes, which are independent of cloud information, are also valuable.

2. INTRODUCTION

An important operational need of the Navy is the ability to anticipate and exploit radar propagation conditions over ocean and coastal regions where ducting and radio 'hole' phenomena are most prominent. These phenomena, caused by elevated refractive layers resulting from 'inversion conditions' in which warm, dry air overlays cooler, moist air, are strongly influenced by both large scale air mass conditions as well as mesoscale features induced by coastlines, mountains and islands. Any attempt to assess or forecast radio propagation conditions must therefore necessarily take into account the synoptic and mesoscale variability of the lower atmosphere. Over the open ocean, it is believed that the assumption of horizontal homogeneity is valid about 80% of the time for typical radio path lengths of importance to the Navy (Glevy,1976). However, near fronts (air and sea), and in the vicinity of islands and coastal terrain, horizontal and temporal variations can be much more prominent.

Under the Navy's electromagnetic EM/EO Propagation Assessment Program, several approaches are being explored to obtain an assessment or prediction of atmospheric conditions in sufficient detail to estimate occurrence, height and intensity of elevated ducts (Richter, 1994). One method, underway at the Naval Research Laboratory (NRL) in Monterey involves development of a high resolution numerical forecast model and data assimilation system in which all available observations are incorporated into a model forecast for use in range-dependence assessments of refractive structure.

Another approach is represented by the Naval Air Warfare Center Point Mugu, Geophysics Branch's attempts to relate refractive structure to synoptic, mesoscale and satellite data parameters by application of statistical relationships established during work under the EM/EO Program. This effort is focussed on the (generally) elevated ducts due to refractive (trapping) layers aloft; the behavior of the evaporative duct attached to the ocean surface is not considered. Implicit is a conceptual model of an elevated trapping layer associated with an inversion-dominated weather regime with low stratus or stratocumulus clouds that are lowest and flattest in the south-eastern parts of the sub-tropical oceans, closely associated with anticyclones (high pressure regions) where the overlying inversion is lowest and strongest. This general trend in both cloud and inversion characteristics correlates well with the occurrence, height and intensity of elevated ducts as observed in extensive radiosonde data sets from the eastern and central north Pacific Ocean.

Since the parameters which describe the distribution, height and intensity of elevated ducts are directly related to the sub-tropical inversion layer and the underlying marine stratus/stratocumulus cloudiness, it is possible to infer information about the duct characteristics from data derived at the cloud tops. Initial efforts to do this focused on recognition of cloud type and appearance on satellite imagery. It became obvious that satellite cloud patterns, particularly those evident in higher-resolution visible imagery, served as good signatures of marine layer depth, strength of the overlying inversion, and by implication the height and intensity of accompanying elevated ducts (Rosenthal et al, 1985). Figure 1 is a characteristic GOES visual image covering a portion of the northeastern Pacific ocean. Changes in appearance of stratus from smooth, to small granular elements and then to larger 'closed' cells have been statistically correlated to elevated duct transitions from low and strong, to high and weak. Areas of open cells and frontal bands correspond to deep, well-mixed conditions and the absence of elevated ducts. The intuitive and experiential

Paper presented at the AGARD SPP Symposium on "Remote Sensing: A Valuable Source of Information", held in Toulouse, France from 22-25 April 1996, and published in CP-582.

conclusions were reinforced by radiosonde refractive profiles at San Nicolas Island, about 60 miles offshore Point Mugu.

3. SATELLITE-IR DUCT TECHNIQUE

Early interest at the Geophysics Branch in the possibility of quantitatively and more objectively determining elevated duct heights from satellite data resulted in the so-called Satellite IR-Duct Technique (Lyons, 1985). It is based on the observation that 1) the tops of the ubiquitous marine stratocumulus clouds off the west coast of the USA and other regions are often co-located with the duct optimum coupling height (base of the trapping layer aloft); 2) cloud-top temperatures are related to their height, and 3) satellite IR data can provide an approximate value of that temperature. Subsequent efforts have been directed towards refining the several components of the technique so that it could be applied as an aid in mapping elevated duct heights over extended areas, using an automated or computer-assisted approach. Figure 2 illustrates an empirical fit of inversion base altitude versus temperature offset from sea-surface temperature (SST); the inversion base lies at the top of the marine layer and is the upper limit for any imbedded stratocumulus cloud tops., This relationship permits conversion of satellite-derived cloud top temperatures into duct height estimates. Complications include variations in sea surface temperature (here subtracted from inversion temperatures to minimize their effect) and other factors which affect temperature lapse rates in the marine layer, and inaccurate IR temperatures due to contamination from moisture above the marine layer, or because of partial cloud coverage. Procedurally, the Meteorological Interactive Data Display System (MIDDS) was used to compute duct heights from the IR duct technique over the Pacific Ocean between the west coast of North America and the Hawaiian Islands.

For analysis of duct heights, the MIDDS cursor is moved around the domain of the satellite images, allowing the optimum coupling height (OCH) to be calculated in one of two situations:

(1) First, the region of the exploring cursor (15 by 15 pixels) may contain a collection of clear, partially-filled, and cloudy pixels. In this case, a scatter plot of IR temperature versus corresponding visual brightness for each pixel can be inspected or processed to extract sea surface temperature (SST) for clear pixels and cloud top temperature for cloudy pixels, using thresholds.

(2) In the event the sampled region contains partially cloudy pixels, but no completely clear pixels are present in the immediate vicinity of the cloud of interest, it is still possible to estimate the underlying SST. The cursor is shrunk to a single pixel and then placed over a number of clear points nearby the sample region. Using upwelling IR radiance data, we extract the SST for each of these locations, and determine how much they are offset from climatology for that latitude and longitude. These offsets are interpolated into the cloud-covered point of interest using a Barnes-Cressman weighting scheme. The resulting interpolated offset is then applied to the climatological SST for the location hidden by the clouds to obtain a value for comparison with the satellite IR cloud-top temperature.

One of the inherent limitations of the IR-duct method occurs with the presence of contaminating high level cloudiness. In our analysis, we are examining low, marine stratocumulus in fair weather conditions over the eastern Pacific Ocean, but cirrus cloudiness can occasionally move into the field of view, due to the proximity of jet streams, tropical outflow or encroaching frontal boundaries from the mid-latitudes. IR-based temperatures can go well below 273K in this type of situation, invalidating use for estimating lower marine layer stratus top temperatures. However, peculiarly cold temperatures in the upper atmosphere have unique signatures on the bispectral scatter plot which allow them to be detected and discarded. Similarly, inadvertant inclusion of land areas can be avoided.

Another problem arises when we are analyzing large overcast stratus areas, since the clear pixels needed for SST can be remote from the location of interest and the interpolated offsets may not be accurate.

4. EQUIVALENT-ALTITUDE TECHNIQUE

Using the hydrostatic assumption and a simple two-layer representation of the lower atmosphere,it is possible to calculate the altitude of the interface between the layers, if temperature and pressure/height at reference levels above and below the region of interest are known (Helvey,1979). Given 700mb and 1000mb data, this 'equivalent' altitude so obtained can be used to estimate the height of the inversion base and optimum coupling height of an accompanying elevated duct. Statistical analysis of historical radiosonde data sets provides the basis for determining optimum coupling height from Equivalent Altitude, as illustrated in Figure 3. The likelihood of a duct aloft can be estimated from climatology, reinforced by consideration of the overall stability of the1000mb-700mb layer; ducting is favored by large stabilities (temperature increase with height). This technique is also being explored as the basis for a first guess estimate for extrapolating climatological elevated duct statistics over data-sparse ocean regions. The advantage of the Equivalent Altitude approach is that grid fields from numerical analyses can be overlain on corresponding satellite imagery over the entire domain, yielding duct heights regardless of cloud conditions, providing a completely independent and complementary set of information which can be cross-calibrated and blended with results from the satellite technique.

5. USING IR DUCT AND EQUIVALENT ALTITUDE TOOLS TO ASSESS SYNOPTIC VARIABILITY

We may relate duct heights computed from these techniques to the accompanying synoptic situation using a case study from the VOCAR (Variability of Coastal

TABLE I.

DATE	Satellite IR-Duct OCH [meters]	Radiosonde OCH [meters]	Duct Strength
26 August	350	305	very strong
28 August	150	206	extreme
29 August	250	358	moderate
30 August	250	371	extreme
31 August	300	603	weak
01 September	350	475	very weak
02 September	50	14	moderate

Atmospheric Refractivity) field tests. Some aspects of this interesting sequence of events were described by Helvey et al(1994). In Figures 4a-c/5a-c we see the results of progression of a vorticity maximum off the coast of California. Figure 4a shows the accompanying low pressure system at 1000 millibars (mb) transversing the northern California coast as it moves westward. High pressure prevails at 700mb, and the air mass is stable, as depicted in figure 5a by the dashed lines(see later description of equivalent altitude analyses). The low moves west-southwestward underneath the 700 mb during the next two days, gradually expanding as it moves. Duct heights reach a local maximum within the swirl of clouds that mark the low, as a result ofthe rising motion "deepening" the marine layer upward. Duct heights to the north and southeastof the vorticity maximum are lower, because the mixed layer has been compressed by sinking airassociated with the strong overlaying inversion. Further west and southwest, duct heights are considerably higher and reach 700-1000 meters in this sequence of photographs. This is consistent with observations, as marine layer depths normally increase as one transits from the west coast toward Hawaii as described previously by Rosenthal and Helvey (1991).

A day later at 0030Z on 2 September the remains of the low are shown well out to sea with stratus broken up into large, fairly high, dense pieces but with clear conditions prevailing along southern and central California and well out to sea. Off Baja California, far removed from the effects of the low pressure influence, a much lower and more uniform sheet of stratus began to spread northward as anticyclonic conditions aloft again took over.

To assess the accuracy of the IR-duct technique, estimated values were compared to computed values from actual radiosonde measurements at San Nicolas Island. In a preliminary sense, results compare quite well to previous duct verification analyses performed at Point Mugu as documented by Lyons (1985) and Szymber and Fox (1989). Comparisons between duct heights computed in this study and radiosondes are presented in Table I. Again, these apply to the VOCAR experiment, which was conducted during August-September, 1993. Daily radiosondes were available within a few minutes before or after 0000 GMT each day, and are compared to the 0000 GMT run of the MRF model. The relative strength of the elevated duct or OCH is indicated in the last column.

Note that differences between calculated OCH values and radiosonde observations are largest for weak ducts, but are relatively small for ducts that are moderate or greater in strength. Calculated OCH values average 27% lower than radiosonde values. However, when the weak and very weak ducts are eliminated, the difference is only 16%. Clearly, some skill is demonstrated by analyzing elevated ducts with the IR-Duct technique, despite the small sample. The average difference is 90 meters (296 feet) counting all ducts, and 41 meters (134 feet) with the weak ducts eliminated. Fox and Szymber found 80% of duct estimates from satellite data were within 400 feet of radiosonde measurements, supporting use of the present technique for evaluation of elevated ducts.

The 'equivalent altitude' analysis of marine layer depth for the VOCAR case is shown in Figures 5a-c, as calculated from the global MRF (Medium Range Forecast) model gridded fields for 700 and 1000 mb heights and virtual temperatures. Locally, a maximum in equivalent altitude is located near the vorticity center, offset by a small distance. In Figures 5a and 5b, the lowest values appear to the southeast of the vorticity max, consistent with the pattern of duct heights observed earlier. Values increase as one goes west and southwest of the vorticity max, in agreement with computed duct heights and actual observations. Note that low duct heights spread northward along the coast behind the dissipating vorticity max, as we see in Figure 4c. This also appears in the corresponding equivalent altitude map, Figure 5c. A northward "alongshore" surge of marine air is observed in the stratus deck accompanying these features. These West Coast surges have been discussed by Mass and Albright (1987).

Although values for our gridded analysis differ somewhat from calculated duct heights, the overall pattern for the marine layer is in general agreement. Note that equivalent altitude is not applicable over large landmasses, because the unique two-layer subtropical marine structure is not present.

The IR-duct and equivalent altitude techniques were used to detect and map refractive variability (topography of the elevated duct optimum coupling height surface) for several other cases. From 17 to 19 September 1994, a broad area of inversion-dominated stratus and stratocumulus clouds prevailed along and offshore the U. S. West Coast in advance of a frontal band and vigorous low pressure system in the central Pacific. Figures 6a-c

show the visual GOES-7 views with heavy numbers indicating IR-duct values, and thin lines corresponding to 700 mb and 1000 mb height contours. In Figures 7a-c are displayed contours of the computed eauivalent altitude values, as well as contours of stability values (700 mb minus 1000 mb potential temperatures).

As high pressure builds in advance of the frontal system, duct heights remain low along the California coast, increasing in height towards the storm system. The gradient of duct height is greatest in the region about 500 miles west of the central California coast.

In another case, from 25 to 27 July 1995 (Figure 8a-c), a weak upper level disturbance about 600 miles west of southern California has destabilized the marine layer, resulting in large clusters of thick stratocululus arranged within an otherwise partially clear region near the disturbance center. Meanwhile, along the coast, the marine layer is shallow under higher pressure, and the stratus is consequently more uniform and continuous. IR-duct heights are low along the coast, but are approximately 500 meters higher in the vicinity of the offshore low pressurre disturbance. A similar gradient in duct height is exhibited in the equivalent altitude contours, shown in Figure 9a-c.

6. CONCLUSION

In the absence of direct measurement of duct parameters, the occurrence, height and intensity of these refractive features can be estimated either from IR cloud-top temperatures processed from satellite imagery using the IR duct technique, or from gridded data at 2 levels using the equivalent altitude technique. These can be used to assess the horizontal variability of the EM propagation environment as well as to infer future horizontal variability by changes in the synoptic patterns. Efforts to relate refractivity features to specific synoptic events are continuing, and results so far show good agreement with synoptic-refractive models developed in earlier studies. Since the IR-duct technique requires the presence of marine cloudiness, collaborative efforts are underway to combine it with a technique working only in the absence of clouds, developed by the U. S. Naval Postgraduate School and described elsewhere in these proceedings (Wash, Davidson and Jordan, 1996).

7. REFERENCES

Glevy, D. F. "An Assessment of Radio Propagation Affected by Horizontal Changes in Refractivity", Naval Electronics Lab. Cen. Technical Note 3153, May 1976.

Helvey, R. A., "Diagnosis of Elevated Ducts by means of Equivalent Altitude", Proceedings of Conference on Atmospheric Effects Assessment, 23-25 January 1979, NOSC Technical Document 260, 15 June 1979.

Helvey, R., J. Rosenthal, L. Eddington, P. Greiman, and C. Fisk, "Use of Satellite Imagery and Other Indicators to Assess Variability and Climatology of Oceanic Elevated Ducts". AGARD CP-567, pp. 33.1-33.13, 1994.

Richter, J. "Structure, Variability, and Sensing of the Coastal Environment", AGARD CP-567, pp.1.1-1.13, 1994.

Rosenthal, J. and R. Helvey, "Refractive Assessments from Satellite Observations", AGARD CP-502, pp. 8.1-8.9, 1991.

Rosenthal, J., S. Westerman, and R. A. Helvey, "Inferring Refractivity Conditions from Satellite Imagery," Pacific Missile Test Center, Geophysical Sciences TN 96, 1985.

Coakley, J. Jr. and F. Bretherton, "Cloud Cover from High-Resolution Scanner Data: Detecting and Allowing for Partially Filled Fields of View", Journal of Geophysical Research, Vol. 87, No.C7, pp. 4917-4932, 1982.

Lyons, S. "SPADS Automated Duct Height Statistics", Geophysical Sciences Technical Note No.100, NAWCWPNS, Point Mugu, CA, 1985.

Szymber, R. and A. Fox, "Evaluation of Accuracy of the PMTC Satellite-Infrared (IR) Automated Duct Height Algorith",. Geophysical Sciences Technical Note No. 146, NAWCWPNS, Point Mugu, CA, 1989.

Mass, C. and M. Albright, "Coastal Southerlies and Alongshore Surges of the West Coast of North America", Monthly Weather Review, Vol. 115, pp. 1707-1738, 1987.

Wash, C. H., K. Davidson, and M. S. Jordan. "Multispectral Remote Sensing of the Coastal Atmospheric Boundary Layer", AGARD SPP Symposium on Remote Sensing, April 1996.

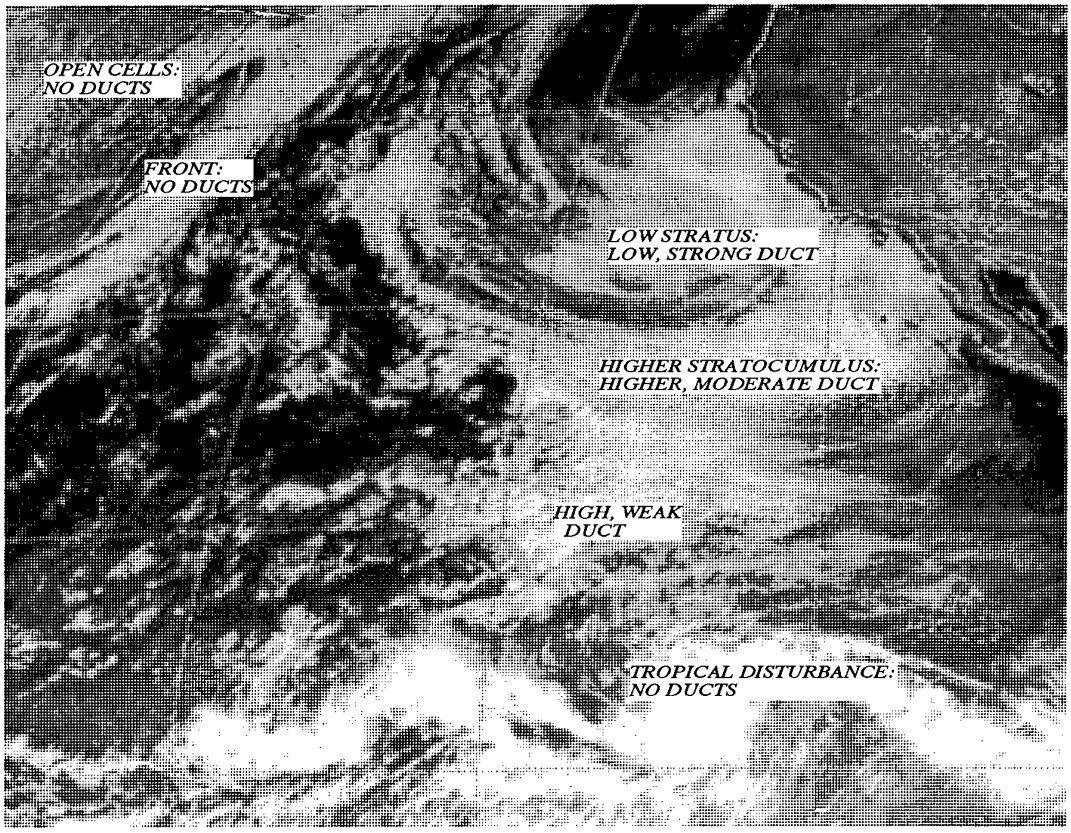


Figure 1. Typical cloud cover in GOES visual image covering portion of northeastern Pacific Ocean

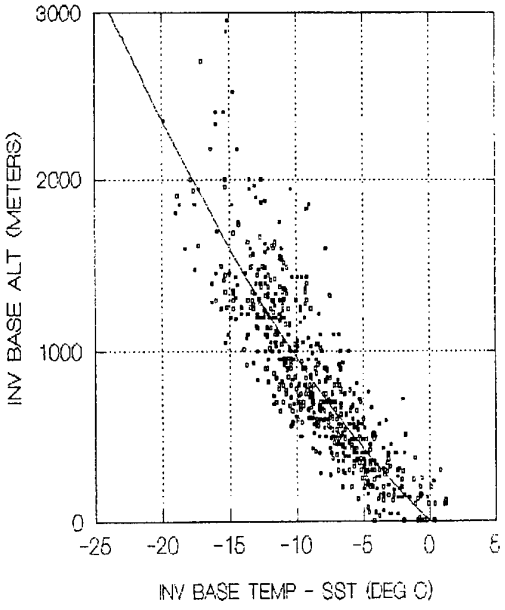


Figure 2. Inversion base altitude versus temperature relative to surface (SST)

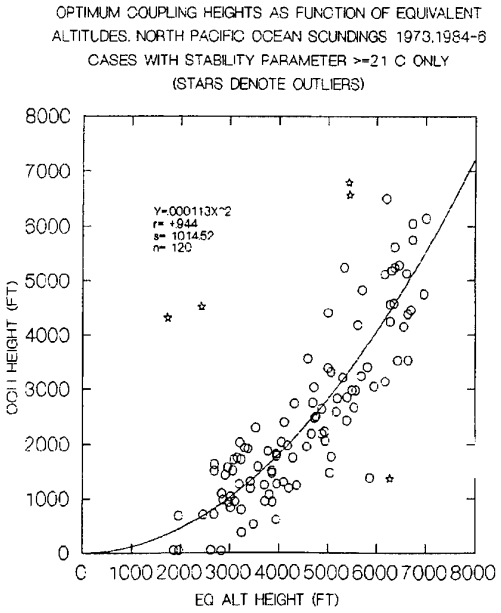


Figure 3. Optimum Coupling Height versus Equivalent Altitude

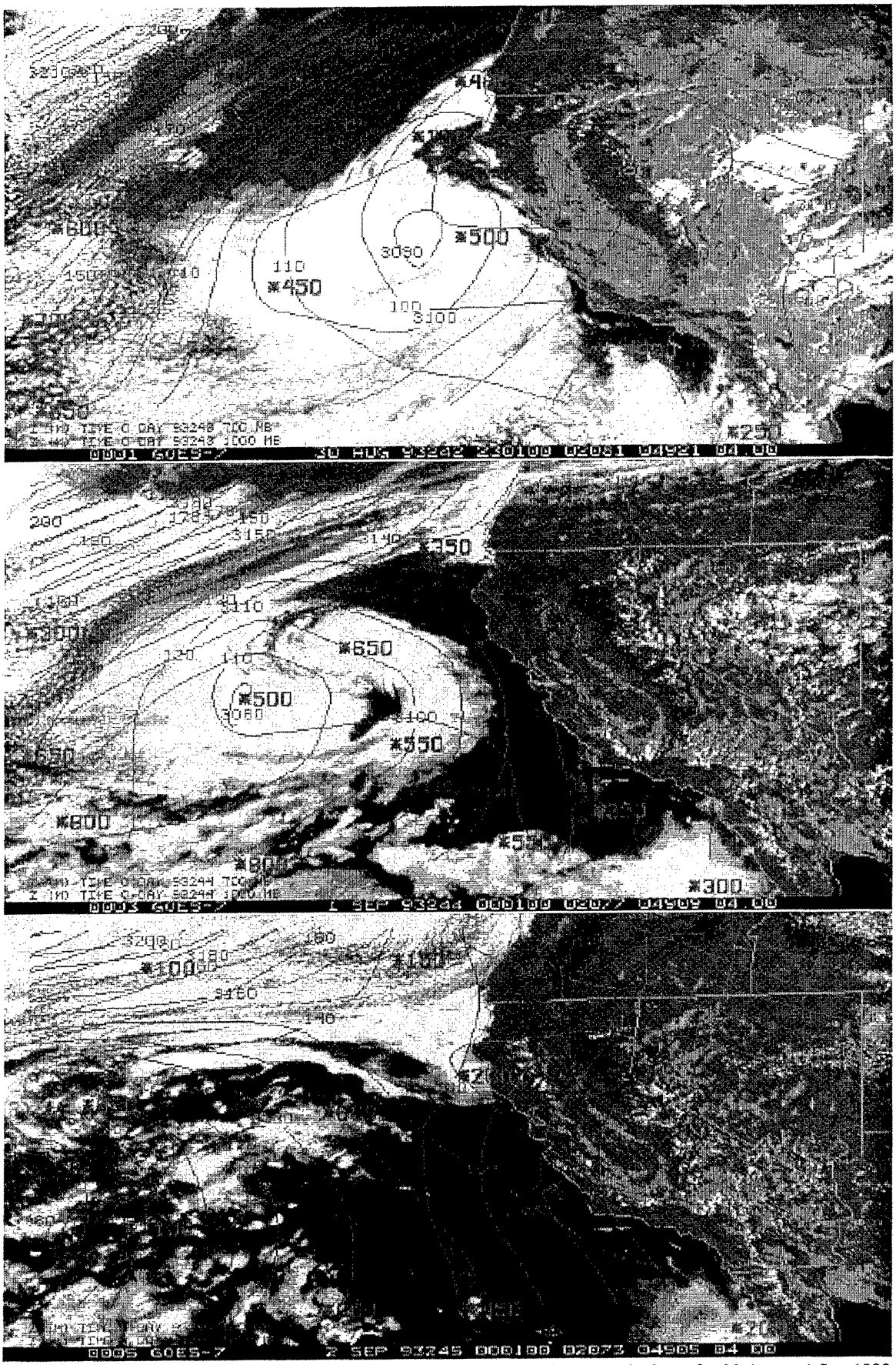


Figure 4a,b,c. GOES visual images, IR-duct altitudes [meters], 1000mb & 700mb charts for 30 Aug to 1 Sep 1993 (top to bottom).

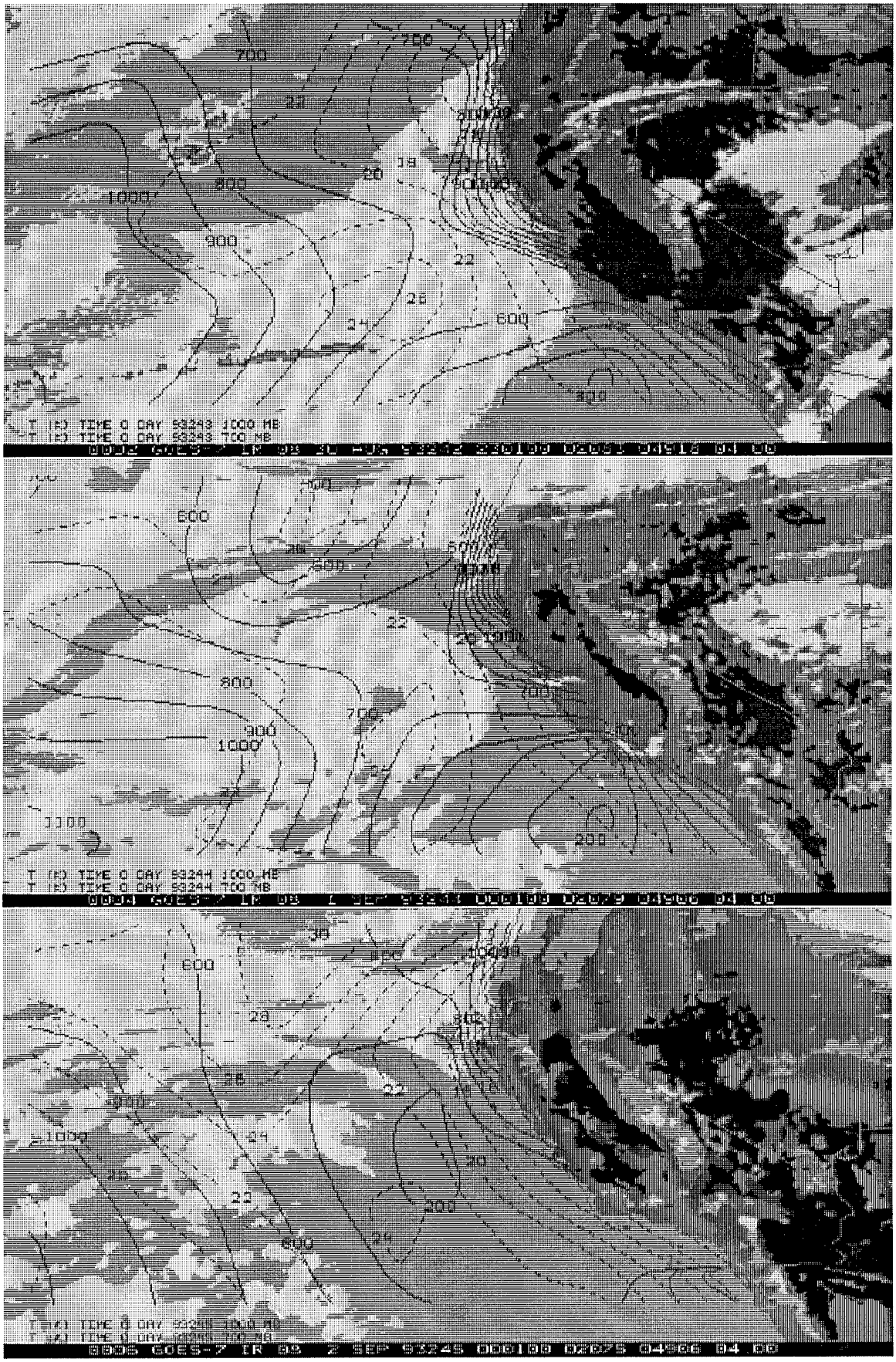


Figure 5a,b,c. IR images, equivalent altitude & stability index (dashed), 30 Aug to 1 Sep 1993 (top to bottom). Equivalent altitude is in meters, stability index is 700mb minus 1000mb potential temperature difference in degrees C.

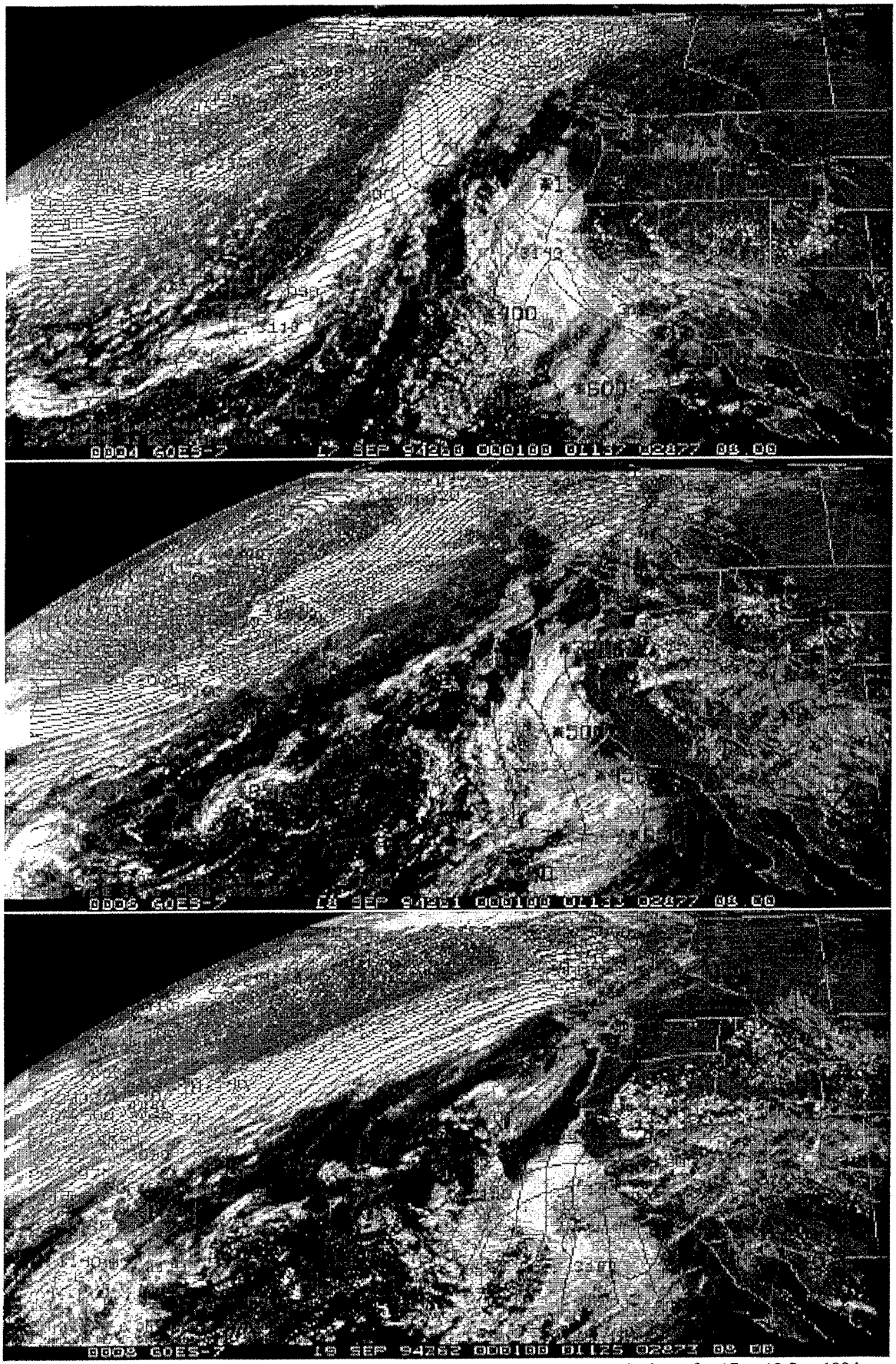


Figure 6a,b,c. GOES visual images, IR-duct altitudes [meters], 1000mb & 700mb charts for 17 to 19 Sep 1994 (top to bottom).

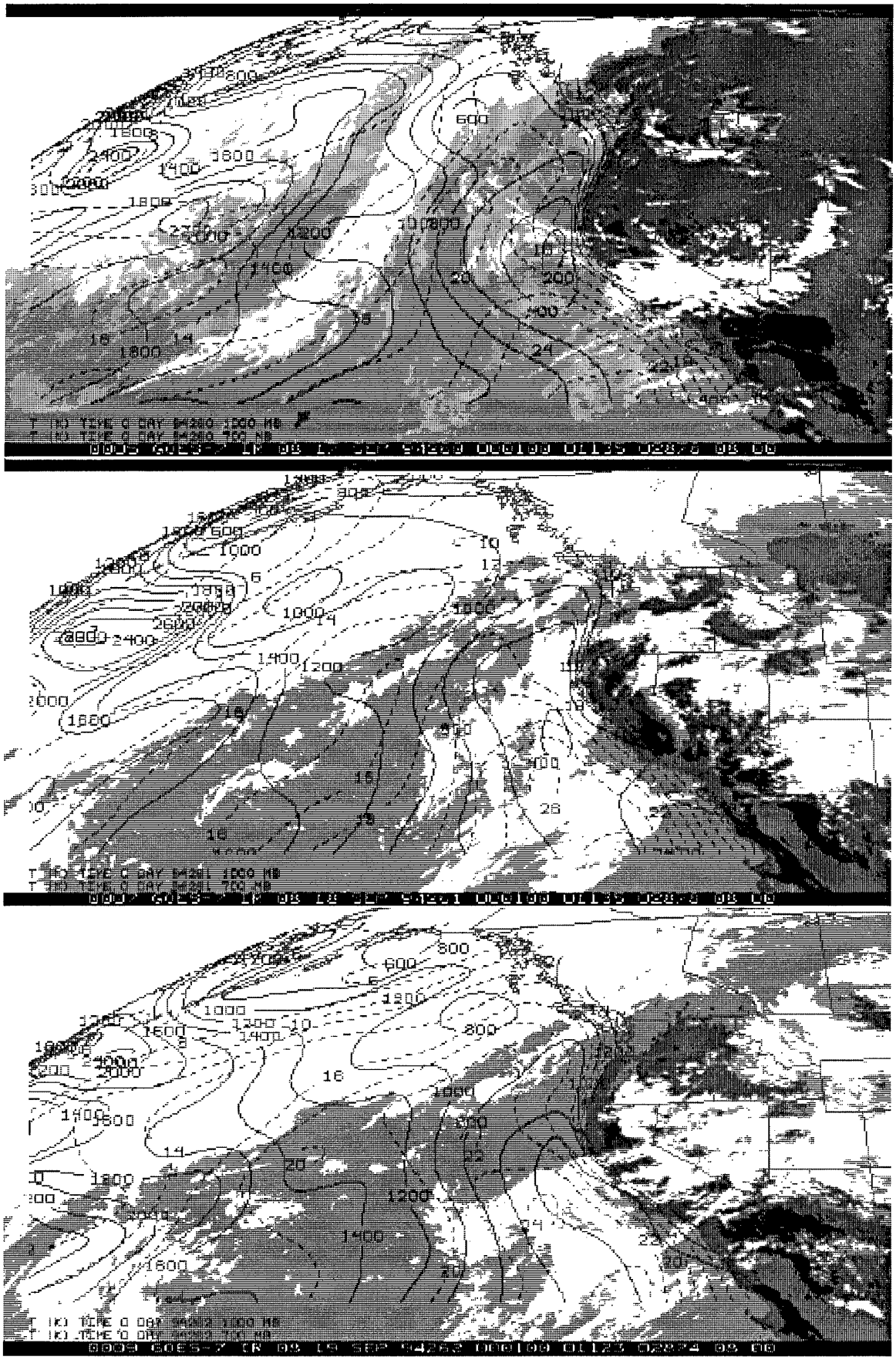


Figure 7a,b,c. IR images, equivalent altitude & stability index (dashed), 17 to 19 Sep 1994 (top to bottom).
Equivalent altitude is in meters, stability index is 700mb minus 1000mb potential temperature difference in degrees C.

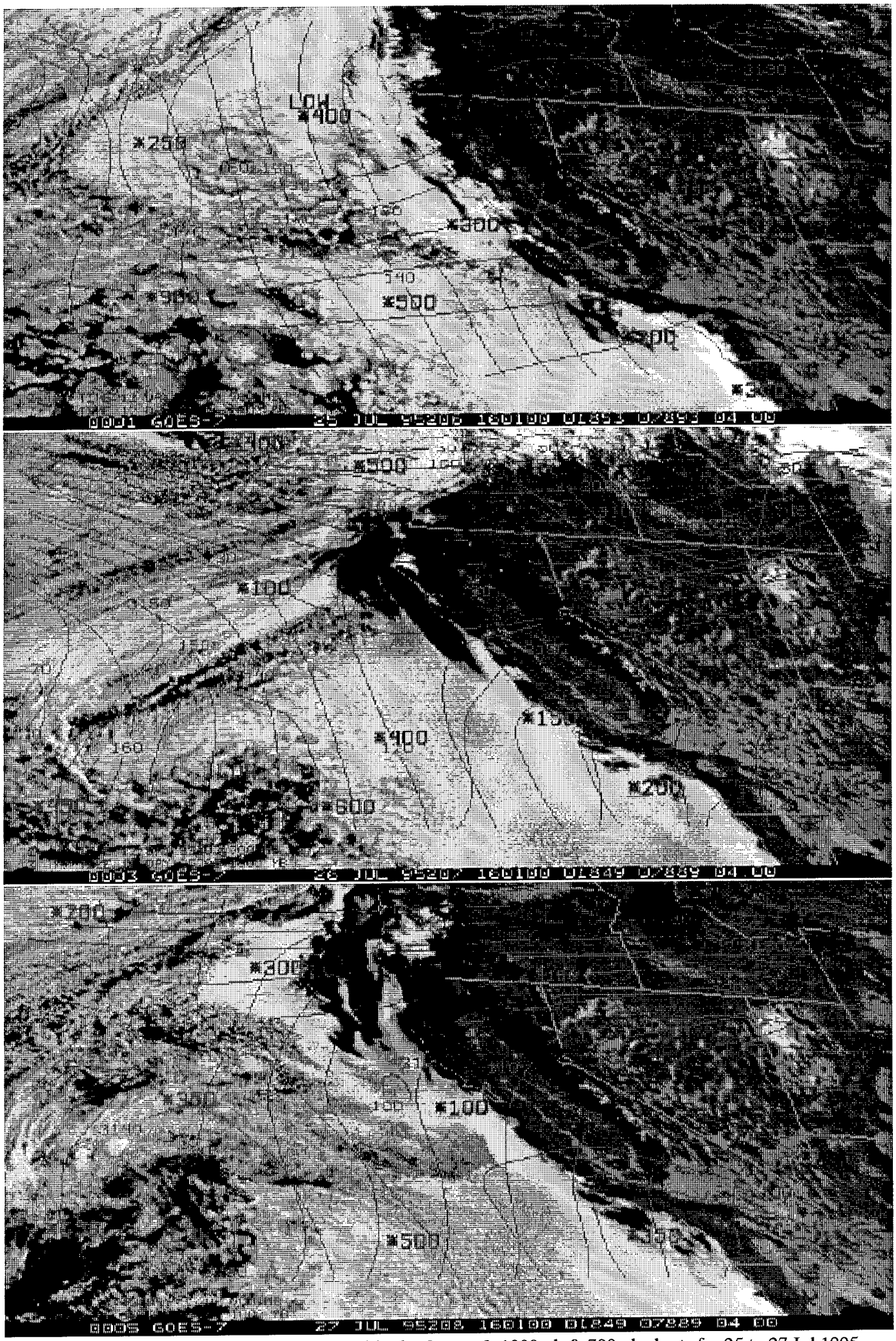


Figure 8a,b,c. GOES visual images, IR-duct altitudes [meters], 1000mb & 700mb charts for 25 to 27 Jul 1995 (top to bottom).

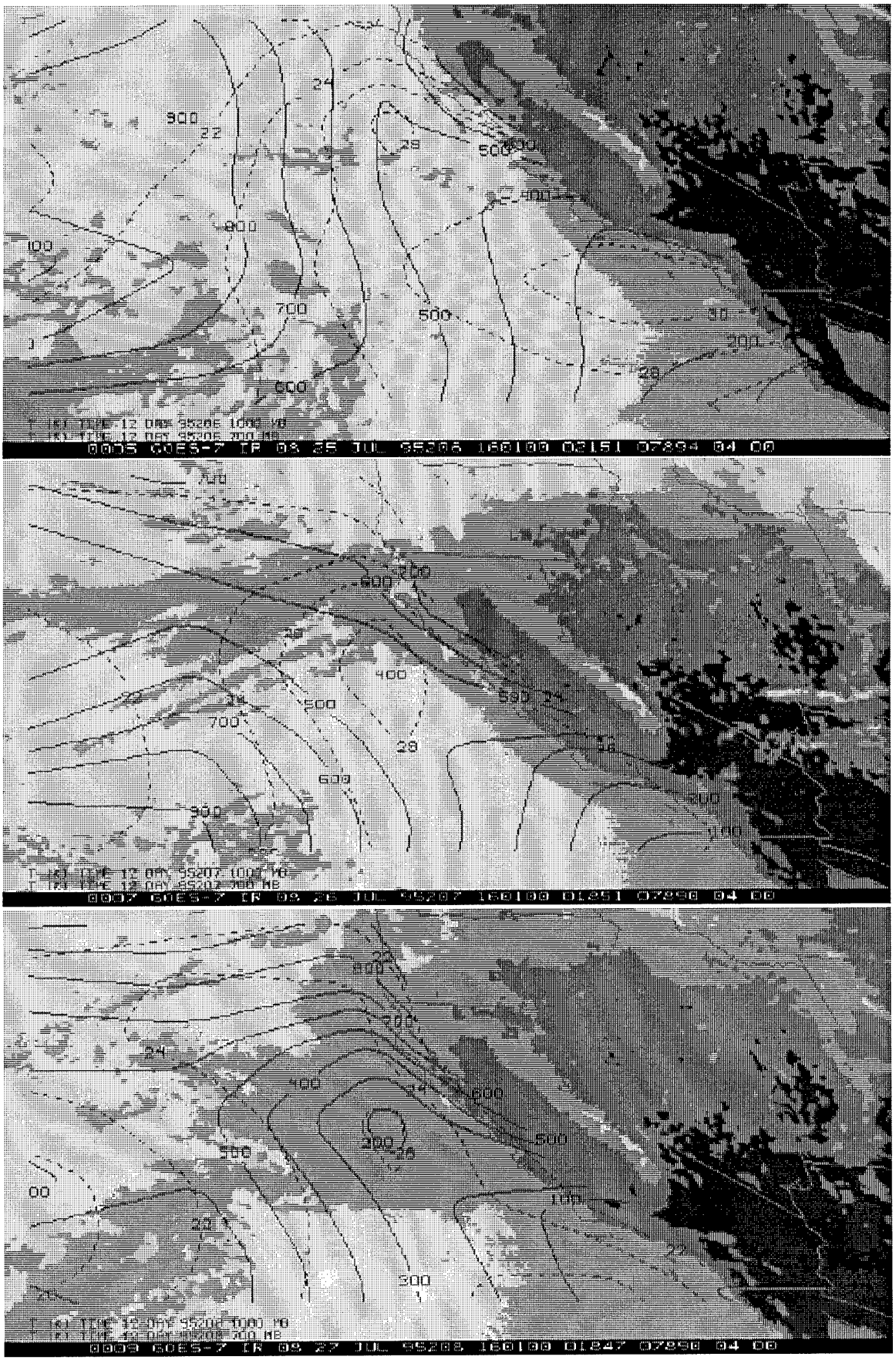


Figure 9a,b,c. IR images, equivalent altitude & stability index (dashed), 25 to 27 Jul 1995 (top to bottom).
Equivalent altitude is in meters, stability index is 700mb minus 1000mb potential temperature difference in degrees C.

Paper 47

M. McKean (CA)

Have you done any storm tracking or prediction including assessment of sea state?

Author's reply:

We are of course very interested in tracking and prediction of storms (and associated sea states) by means of satellite imagery. However, in this talk, the focus of the techniques is to determine characteristics of the propagation environment under subsidence inversion conditions where a well mixed marine layer is capped by a stable layer of warm, dry air. These conditions occur away from storms. In the talk that Mr Helvey gave yesterday, he described techniques he is developing to extrapolate (predict) low cloud features. This is an excellent tool for estimating low level winds (and sea conditions) as well as forecasting the movement of convective clouds and storms. For major storm features like fronts, we are collaborating with contractors to NRL to develop a short-term prediction technique of cloud systems responsible for widespread precipitation.

N. J. Farsaris (GR)

What frequency band was used in the studio soundings mentioned in the first method?

Author's reply:

The "IR-duct" technique uses statistics compiled from a large number of radiosondes (balloon soundings) that in effect allowed the conversion of inversion base altitude to inversion base temperature. These (approximatetly 4000) soundings consisted of standard V1Ƶ, Vaisala and older UCLA sondes operating at the standard radiosonde telemetry frequency of 403 MHz, or in some cases 1680 Mhz.

T.Wahl (NO)

In SAR images we frequently see ocean surface footprints of atmospheric gravity waves. The variations in surface wind speed can be computed from a backscatter model. Could such observations be used to get more information about the height of the marine boundary layer?

Author's reply:

1) We have observed that the size of the marine stratocumulus clouds provides a rough indication of the depth of the marine layer. Presumably, the scale (and shape) of wind variations shown by SAR imagery could be related in the same way to the size of internal mixing structures in the marine atmosphere, and hence its depth.

2) If the speed of gravity waves propagating along the top of the marine layer can be detected by SAR imagery, and if this speed can be related to marine layer depth, then the SAR could be a useful tool for this application.

Remote Sensing of Evaporation Ducts Using SHF Propagation Measurements

L. Ted Rogers
Propagation Division
NCCOSC RDTE DIV 883
49170 PROPAGATION PATH
SAN DIEGO, CA 92152-7385

1 SUMMARY

A maximum likelihood method is implemented to determine the height of the evaporation duct from multi-frequency, slightly beyond-line-of-sight SHF propagation measurements using data from the NATO sponsored propagation experiment that took place in the vicinity of Lorient, France during the fall of 1989 [1] [2]. The model for the propagation environment is horizontally homogeneous with a vertical structure having two parameters, the evaporation duct height and the trapping layer base height. The environmental parameter space is discretized and propagation factors are computed using a parabolic equation model for each point in the parameter space. The relative likelihood of combinations of the environment parameters are computed from the mismatch of the vectors of modeled and measured propagation factors. Good agreement is found with evaporation duct heights calculated using bulk meteorological models.

2 INTRODUCTION

Non-standard refractivity structures affecting low altitude electromagnetic (EM) propagation in the over-water environment can (usually) be divided into evaporation ducts and elevated trapping layers. The evaporation duct is formed by the rapid decrease in humidity above the sea surface and is the most likely mechanism for non-standard propagation [3]. Elevated refractive layers are further classified as surface based ducts and elevated ducts which are formed by the capping inversion of the stable marine atmospheric boundary layer. When the elevated refractive layer forms a surface based duct, signal levels at beyond-line-of-sight ranges can increase to 10 dB or more above free space levels at radio and microwave frequencies.

Two parameters, the evaporation duct height δ and the trapping layer base height h, are *inverted* from propagation magnitude measurements at SHF frequencies on a near-LOS path using a maximum likelihood method. As described in [4] the advantages of the determination of the refractive environment from propagation measurements is that the effects that are sensed are path-integrated and are measured instantaneously. The author believes that the measurement and methods utilized:

1. Reliably indicate evaporation duct heights when evaporation ducting is the dominant mode of propagation and the evaporation duct height is less than approximately 22 meters.

2. Indicate a feasible combination of trapping layer parameters when the dominant mode of propagation is due to a surface based duct.

2.1 Current evaporation duct sensing

The refractivity profile of the evaporation duct is usually calculated from bulk measurements of air temperature T_a, wind speed U, relative humidity RH and sea surface temperature T_s. Those measurements are input to a model based upon Monin-Obukhov similarity theory [5] such as the *Liu, Katseros and Bussinger* (LKB) model [6], the *Jeske* model [7], or a modification of the Jeske model such as the neutral model and the Paulus model (both in [8]). In the neutral model it is assumed that $T_s = T_a$ and in the Paulus model it is assumed that $T_s = T_a - 1$ if $T_a - T_s \geq 0$. These models are referred to collectively as *bulk* models. The height at which the change in modified refractivity [9] with respect to height is zero is called the evaporation duct height δ. When the dominant mode of propagation is the evaporation duct, propagation calculations computed using the evaporation duct profiles determined by models based upon the LKB, Jeske, neutral and Paulus methods have been found to be of comparable overall accuracy [10].

2.2 Experiment setup

Figure 1 depicts the layout of the *Lorient '89* experiment. Transmitters were located at the Quiberon peninsula and receivers were located at Gavres. The length of the transmission path was 27.7 km. Transmission frequencies, and transmitting and receiving antenna heights referenced to mean-low-low-water level are given in table 1. The mid-path buoy shown in figure 1 was equipped with sensors for wind direction, wind velocity U, relative humidity RH and air temperature T_a. The three sensors were mounted 4.5 meters above the sea surface. The buoy was also equipped with a sea temperature (T_s) sensor which

Paper presented at the AGARD SPP Symposium on "Remote Sensing: A Valuable Source of Information", held in Toulouse, France from 22-25 April 1996, and published in CP-582.

was mounted 1.5 m below the sea surface. The tide was monitored at the receiver site in Gavres.

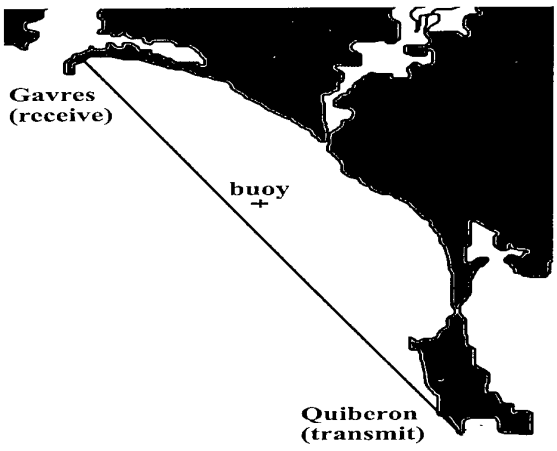


Figure 1: Transmission path for Lorient 1989 propagation experiment.

Table 1: Frequencies and antenna heights referenced to mean-low-low-water.

Freq. (GHz)	Trans. Ht.	Rec. Ht.	Pol.
3.0	11.37	13.41	HH
5.6	11.37	13.41	HH
10.5	11.35	13.50	HH
16.0	11.36	13.40	HH

Both meteorological and propagation measurements became available on September 23, 1989; 0000Z on that date is considered to be day 0.000 of the experiment. Full data were available until day 52.0 of the experiment. In this paper only days 0.0 through 8.0 are considered as the instances of surface based ducting interspersed among the periods of evaporative ducting during that period of time provides for a well-balanced test of the inversion procedure.

2.3 Environmental model

Equations 1, 2 and 3 constitute the environmental model for calculations of the modified refractivity [9] profile $M(z)$ from two parameters, the evaporation duct height δ and the trapping layer height h as illustrated in figure 2.

$$M(z) = .13z - .13 \cdot \delta \cdot \ln\frac{z + z_0}{z_0} \quad z < h \qquad (1)$$

$$M(z) = M(h) - 2(z - h) \quad h < z < h + 10 \qquad (2)$$

$$M(z) = M(h + 10) + .118(z - h) \quad h + 10 < z \qquad (3)$$

Equation 1 is the negative exponential profile for a neutral stability evaporation duct which asymptotically approaches a gradient $(dM(z)/dz)$ of .13 M/m (M-units / meter) in the mixed layer as given in [8]. Equation 2 is for the trapping layer (in this case synonymous with the capping inversion) and is chosen based upon the modeling experience of the author. The modified refractivity gradient within the trapping layer is -20 M/m, for a total negative change of 20 M-units from the bottom to the top of the trapping layer. The negative change in modified refractivity is referred to as the M-deficit. The trapping layer is included to account for propagation effects that are not due to by evaporative ducting. At heights greater than 100 m the trapping layer has little or no effect on propagation for the transmission links that are considered in this paper. Above the trapping layer a standard gradient of .118 M/m is used as given in equation 3. The model used for the shape of the evaporation

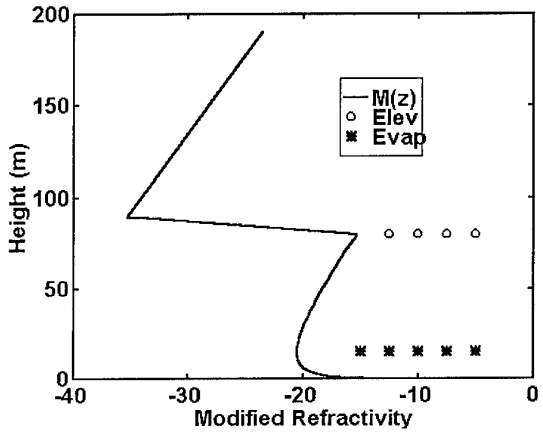


Figure 2: Environmental model having evaporation duct and trapping layer.

duct is well accepted in the propagation and meteorological communities. To the author's knowledge there is yet to be a consensus regarding the shape of the trapping layer so its shape may differ significantly from the true shape. Nonetheless, when the correspondence between measured propagation factors and those associated with the modeled trapping layer is stronger than the correspondence between the measured vectors and evaporation ducts, it is a strong indication that the evaporation duct is not the dominant mode of propagation.

2.4 Environmental effects

The effects that evaporation ducts and elevated trapping layers have on propagation are different. Those effects are examined separately to provide insight into the inversion process. It should be noted that the graphs presented in this section are examples where the tide height is 2.4 meters with reference to mean-low-low-water level. The tide is explicitly taken into account in the refractivity inversion procedures described in this paper.

Evaporation duct effects: Plots of propagation factors

F versus evaporation duct heights δ with no trapping layer present (i.e. having the profile entirely described by equation 1) appear in figure 3. The frequencies and path geometry used in the calculations are as described in section 2.2 with the tide set to 2.4 m MLW. In figure 3 it is observed that only at 3.0 GHz is the propagation factor always an increasing function over the interval of 0 to 40 meters. At the higher frequencies δ is ambiguous for values of F that are greater than approximately -2 dB. At values

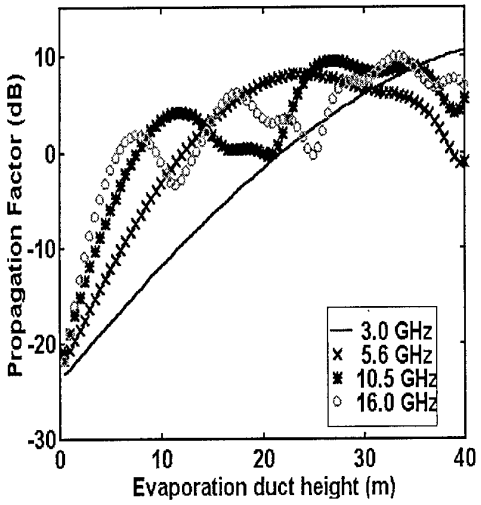


Figure 3: Propagation factor versus evaporation duct height.

of F greater than -2 dB, the corresponding vector of propagation factors becomes "less unique" implying less certainty with a given inverse problem solution in the presence of measurement noise and horizontal inhomogeneity. It is quite clear from figure 3 however, that frequency diversity will improve the accuracy of remotely sensed evaporation ducts heights.

Elevated trapping layer effects: The effects of elevated trapping layers are shown in figure 4 where propagation factor F as a function of trapping layer height h is plotted for four different combination of h and M-deficit dM (the negative change in modified refractivity in the vertical direction at the trapping layer). When the height of the layer is less than 60 meters, the signals are at, or are above, free space levels (i.e. $F \geq 0$dB). When the height is greater than 90 meters, the propagation factors are close to standard atmosphere levels (i.e. $F \cong -23$dB). At layer heights of greater than 60 meters but less than 90 meters the difference between the low and high frequency propagation seems to be largely independent of the M-deficit so the benefit of frequency diversity is minimal. Thus it is unlikely that both layer height and M-deficit can be determined from propagation measurements for the frequencies and path geometry under consideration. Rather than trying to determine both a layer height and M-deficit from the propagation data, a 20 M-unit

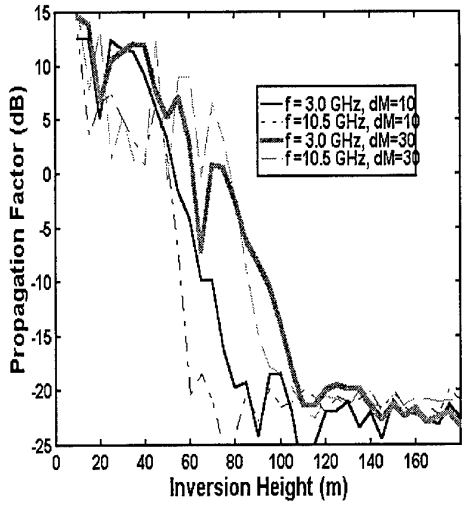


Figure 4: Propagation factors for various trapping layer geometries.

M-deficit will be assumed and only a layer height corresponding to that M-deficit will be inverted. The M-deficit that is assumed and the layer height that is inverted are but a single *feasible* combination of the two parameters.

2.5 Relative likelihood formulation

In the preceding section it was seen that solutions to this non-linear inverse problem are expected to be ambiguous. So rather than having the inversion procedure calculate a single "best" estimate of the refractivity parameters of interest, it is desired to know the likelihood of different parameter estimates over the whole parameter space. Unfortunately, not enough is known about the errors in the inversion process, at this time to determine the true likelihood of estimates. It is the authors decision to use the term *relative likelihood* to describe the certainty of parameter estimates. In other respects, the procedures used are similar to those used for Bayesian estimation as described in [11]. The use of *relative likelihood* in this instance is analogous to the implementation of matched field processing in underwater acoustics where ambiguity surfaces are used without reference to a true probability (see [13] and [14]).

Relationship to Bayes theorem: Let Y be observation, and X be parameters related by the function G such that $Y = G(X)$. Let $\hat{X}_i \subset \hat{X}$ be discrete vectors for modeled values of X such that $\hat{Y}_i = G'(\hat{X}_i)$ where G' is the model of G. The probability that a estimated parameter vector $\hat{X}_i$ is equal to the true parameter vector X, given the observation vector Y is found using Bayes theorem is as shown in equation 4:

$$P(\hat{X}_i = X|Y) = \frac{P(\hat{X}_i)P(Y|\hat{X}_i)}{\sum_j P(\hat{X}_j)P(Y|\hat{X}_j)} \qquad (4)$$

Consider the case where:

1. There is no *prior* knowledge of the distribution of X (i.e. X is be uniformly distributed, otherwise known as the *reference state of information*).
2. The mismatch $Y(t)-\hat{Y}_j$ when $X = \hat{X}_j$ is entirely due to measurement noise that has a zero-mean Gaussian distribution with covariance matrix Q.
3. There are no modelization errors, i.e. $G' = G$.

Then 4 reduces to

$$P(\hat{X}_{i,j} = X|Y) = \frac{e^{(Y-\hat{Y}_i)^T\hat{Q}^{-1}(Y-\hat{Y}_i)}}{\sum_j e^{(Y-\hat{Y}_j)^T\hat{Q}^{-1}(Y-\hat{Y}_j)}} \quad (5)$$

where "T" indicates the transpose. In the inverse problem implementation described in this paper, equation 5 is used.

Relative likelihood method: The two environmental parameters to be estimated are the evaporation duct height δ and trapping layer height h. The frequency f and the tide height d are parameters that are used in the forward problem calculations but are not parameter to be inverted. The discretization of the four parameters is given in table 2 where Δ is the discretization interval and the column labeled "values" lists the actual discrete values of each variable.

Table 2: Discrete parameter space.

Parameter	Δ	Values
$\hat{\delta}_i; i = 1, 2, \ldots 19$	2 m	$\{1, 3, \ldots 39\}$
$\hat{h}_j; j = 1, 2, \ldots 19$	10 m	$\{20, 30, \ldots 200\}$
$f_k; k = 1, 2, \ldots 4$	N/A	$\{3.0, 5.6, 10.5, 16.0\}$
$d_l; l = 1, 2, \ldots 7$	0.6 m	$\{0.6, 1.2, \ldots 4.2\}$

Using the Radio Physical Optics (RPO) propagation model [12] a four-dimensional propagation factor matrix $\hat{\mathbf{Y}}$ is pre-computed for the entire $19 \times 19 \times 4 \times 7$ parameter space described in table 2.

In this paper, time series of the inversion results are examined for data taken at discrete intervals so time is discretized $t_n; n = 1, 2, \ldots N$. For any given t_n, the tide is known, so by use of linear interpolation, the the $19 \times 19 \times 4 \times 7$ matrix of propagation factors is reduced to $\hat{\mathbf{Y}}(t_n)$ a three-dimensional, $19 \times 19 \times 4$ matrix corresponding to the tide at t_n. From $\hat{\mathbf{Y}}(t_n)$, 19×19 different column vectors, each consisting of *model* of propagation factors at the four frequencies are determined as given in equation 6.

$$\hat{Y}_{i,j}(t_n) = G'(\hat{X}_{i,j}, t_n) \begin{pmatrix} \hat{Y}_{\hat{\delta}_i,\hat{h}_j,f_1}(t_n) \\ \hat{Y}_{\hat{\delta}_i,\hat{h}_j,f_2}(t_n) \\ \hat{Y}_{\hat{\delta}_i,\hat{h}_j,f_3}(t_n) \\ \hat{Y}_{\hat{\delta}_i,\hat{h}_j,f_4}(t_n) \end{pmatrix} \quad (6)$$

The parameter vector is $\hat{X}_{i,j} = (\delta_i, h_j)^T$. Note that in general, $\hat{Y}_{i,j}(t_n) \neq \hat{Y}_{i,j}(t_k)$ for $n \neq k$.

The *measurement vector* at time t_n is $Y(t_n)$, where (like the model vector) the elements of $Y(t_n)$ are the measured propagation factors at each of the four frequencies. The squared error $S_{i,j}(t_n)$ of estimating $Y(t_n)$ using $\hat{Y}_{i,j}(t_n)$ is given by equation 7 where I is the identity matrix:

$$S_{i,j}(t_n) = (Y(t_n) - \hat{Y}_{i,j}(t_n))^T I (Y(t_n) - \hat{Y}_{i,j}(t_n)) \quad (7)$$

The parameter vector that minimizes S is $\hat{X}_{i^*,j^*}(t_n)$, or (using simpler notation) $\hat{X}^*(t_n)$.

Estimate of the covariance matrix Q: The mean value of $S(\hat{X}^*, t_n)$ over the time interval that is examined in this paper was found to be 9.414 and is denoted μ_{S^*}. The estimate for covariance matrix Q of the mismatch vector $(Y - \hat{Y}_{i,j})$ was then approximated using equation 8:

$$\hat{Q} = \frac{1}{\mu_{R^*}} I \quad (8)$$

The implicit assumption of using this estimate of covariance is that $X = \hat{X}^*$. This is unlikely to be true; if anything it is expected that the magnitude of the true error covariance Q is larger, perhaps quite a bit larger than $\hat{Q}$.

Continuing with relative likelihood: The normalized squared error is then:

$$R_{i,j}(t_n) = (Y(t_n) - \hat{Y}_{i,j}(t_n))^T \hat{Q}^{-1} (Y(t_n) - \hat{Y}_{i,j}(t_n)) \quad (9)$$

The effect of normalization is that the average value of $R_{i,j}(t_n)$ over the interval 0.0 - 8.0 days is 1. The *relative likelihood* $RL(\hat{X}_{i,j}(t_n)|Y(t))$ of the estimated parameter vector $\hat{X}_{i,j}(t_n)$ being equal to the true parameter vector $X(t_n)$, given the observation vector $Y(t_n)$ is appears in equation 10:

$$RL(\hat{X}_{i,j}(t_n)|Y(t_n)) = \frac{e^{-R_{i,j}(t_n)}}{\sum_{n,m} e^{R_{n,m}(t_n)}} \quad (10)$$

At each time increment t_n, computations are performed for the for the relative likelihood for all combinations of δ_i and y_j.

3 CASE STUDY

Two illustrative examples are now examined. The first is from a period where evaporation ducting is the dominant mode of propagation (day 7.01) and the second is from an earlier observation (day 2.90) where the dominant mode of propagation is thought to be due to a surface based duct. To better illustrate the ambiguity, the Q matrix has been multiplied by 2.5.

The measurement vector for day 7.01 is given in equation 11. Since the propagation factor at both 3.0 GHz and 5.6 GHz are well below free space and below the

propagation factors at 10.5 and 16.0 GHz it is reasonable to assume that dominant propagation mechanism at that time was evaporative ducting.

$$\begin{pmatrix} F_{3.0} \\ F_{5.6} \\ F_{10.5} \\ F_{16.0} \end{pmatrix} = \begin{pmatrix} -16.7 \\ -9.4 \\ 1.1 \\ 0.6 \end{pmatrix} \tag{11}$$

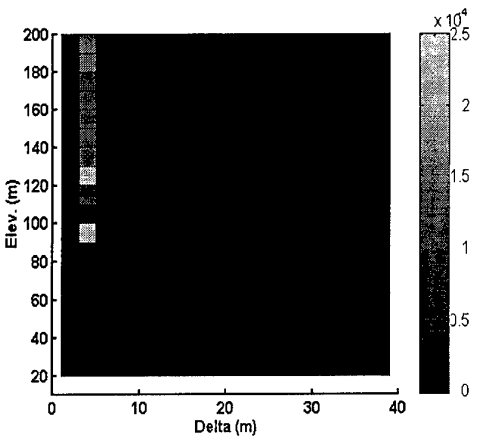


Figure 5: Ambiguity surface for day 7.0.

Figure 5 contains the relative likelihood of combinations of δ and h corresponding to the measurements in equation 11. The region of high likelihood is $7 < \delta < 9, h > 90$. The unambiguous results shown here are much different than the results when sensing at a single frequency; sensing using the 3.0 GHz signal alone is too insensitive at low evaporation duct heights to accurately indicate the duct height. Sensing using just one of the the four higher frequencies give results are too ambiguous.

Even with sensing at four frequencies the unambiguous results as observed in figure 5 are by no means the rule; a case in point is provided by looking at the ambiguity surface that results from using the factors measured on day 2.9 of the experiment as given in equation 12.

$$\begin{pmatrix} F_{3.0} \\ F_{5.6} \\ F_{10.5} \\ F_{16.0} \end{pmatrix} = \begin{pmatrix} .7 \\ 7.2 \\ 5.3 \\ 4.6 \end{pmatrix} \tag{12}$$

In figure 6 it is seen that at that time there are several likely combinations of δ and h corresponding to the propagation measurements and those combinations are not close together. Referring back to figure 3 it is clear that the measured propagation factors at all four frequencies are "saturated". A point this illustrates is the need for a likelihood approach to the inverse problem. Each of the lightly colored areas in figure 6 represents a local maxima of likelihood of equation 10, or equivalently a local minimum of equation equation 9. Had a gradient based optimization been used to search for a single maximum likelihood point, where the solution would be found in parameter space would depend upon the point in parameter space where the search was started. Knowing the certainty of solutions to non-linear problems like the problem here is as important as the solution itself.

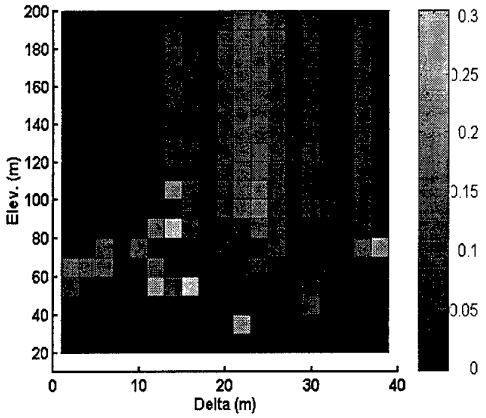


Figure 6: Ambiguity surface for day 2.9.

4 TIME SERIES

The following time series are used to examine the performance of the inversion methods that are utilized in this research:

1. Figure 7: Measured propagation factor at 3.0 and 10.5 GHz.
2. Figure 8: Likelihood contours of remotely sensed trapping layer height.
3. Figure 9: Likelihood contours of remotely sensed evaporation duct height.
4. Figure 10: Evaporation duct heights calculated using the neutral and LKB from measurements of T_a, T_s, U and RH at the buoy location shown in figure 1.

In each case the abscissa is the time in days since midnight on the first full day that both meteorological and propagation data became available. The data is now divided into periods where it appears that surface based ducting is the dominant mode of propagation and periods where evaporation ducting is the dominant mode. The evaporation duct heights determined by remote sensing under the two different propagation conditions are compared to duct heights calculated using bulk models.

Surface based ducting as the dominant mode of propagation: It is hypothesized that surface based ducting is the dominant mode of propagation for the time intervals of days 1.6 - 1.9, 2.6 - 2.9, and 3.5 - 3.8. The

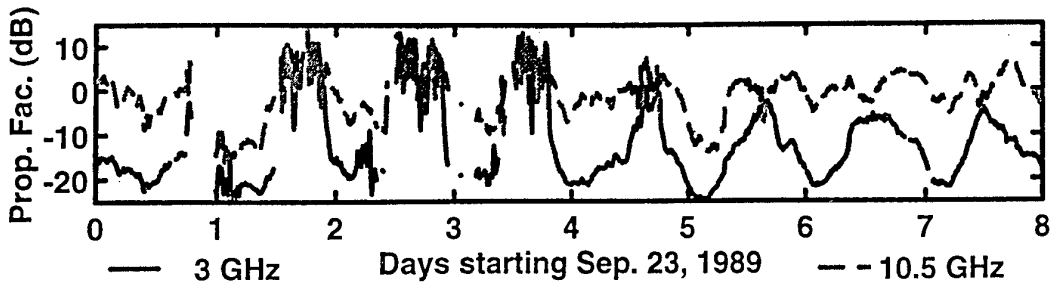


Figure 7: Measured propagation factors at 3.0 and 10.5 GHz.

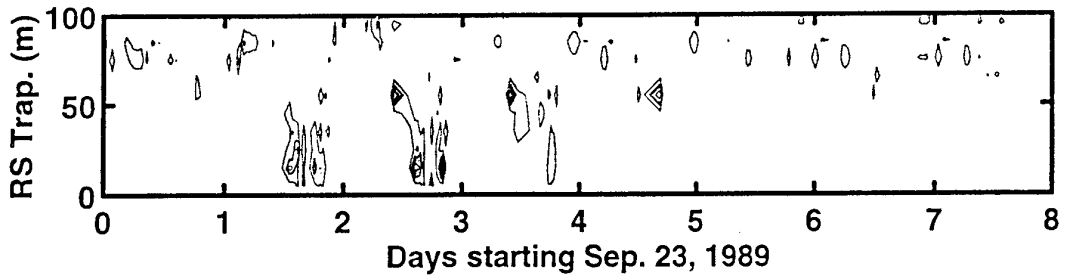


Figure 8: Relative likelihood contours of trapping layer height.

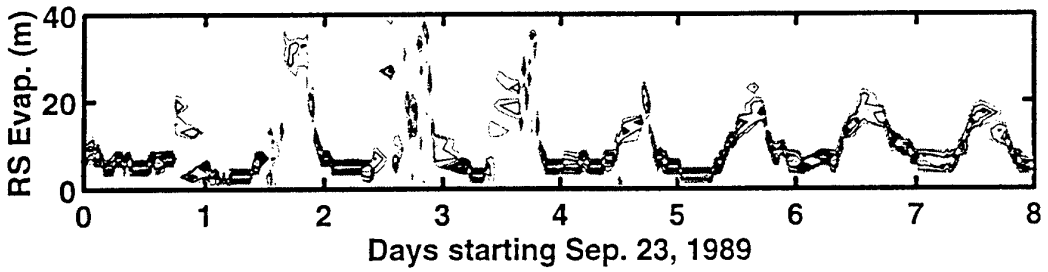


Figure 9: Relative likelihood contours of evaporation duct height.

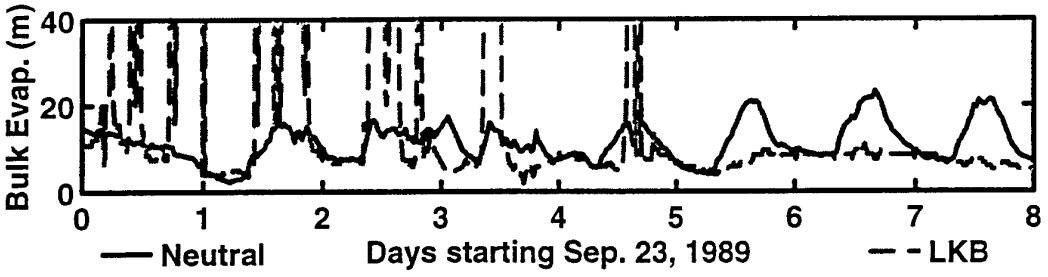


Figure 10: Evaporation duct heights calculated from meteorology.

following observations (from those periods) are consistent with that hypothesis:

1. From figure 7: Measured propagation factors at both 3.0 and 10.5 GHz are at or above free space levels and significant fading occurs. While the former is associated with both very high evaporation ducts as well as surface based ducts, the latter is more typically observed only with surface based ducts.

2. From figure 8: Trapping layer contours are concentrated at heights of less than 80 meters. Note that in figure 4, that only surface based ducts at height of less than 80 meters significantly affect propagation for the path and frequencies of interest.

3. From figure 9: Contours of relative likelihood for the height of the evaporation duct, as opposed to their appearance during other period, are ambiguous and, for the most part, at height of greater than 20 meters.

Duct heights calculated with the neutral and LKB models appear in figure 10. In each of the three intervals of suspected surface based ducting the neutral evaporation duct height is less than 15 meters, while the duct height calculated using the LKB method is either undefined or is less than 15 meters. Investigation of the episodes where the LKB duct height is undefined is beyond the scope of this paper, nonetheless, it is quite clear that there are large discrepancies between the evaporation duct heights estimated by remote sensing and estimates calculated using the neutral and LKB models.

It is the assessment of the author that during these periods of time, surface based ducting is the dominant mode of propagation and the remotely sensed evaporation duct heights are unreliable. The indication of surface based ducting is felt to be reliable in the sense that the inverted trapping layer height and assumed 20 M-unit M-deficit constitute a feasible combination of surface based duct parameters.

Evaporation ducting as the dominant mode of propagation: Evaporative ducting is hypothesized to be the dominant mode of propagation 0.0 - 1.6, 1.9 - 2.6, 2.9 - 3.5, and 3.8 - 8.0. The following observations from those periods of time are consistent with that hypothesis:

1. From figure 7: The propagation factor at 3.0 GHz is less than that at 10.5 GHz.

2. From figure 8: There is little indication of a trapping layer below 80 meters.

3. From figure 9: The likelihoods for the remotely sensed evaporation duct heights are concentrated within a few meters.

By visual inspection the root mean squared (RMS) difference between remotely sensed height in figure 9 and the neutral duct height in figure 10 is around 2 meters. The correspondence between the remotely sensed evaporation duct height and that calculated using the LKB model shown in figure 10 is not as good. Based upon the analysis of propagation estimation accuracy reported in [10], it is the duct heights calculated using the LKB model that are in error in the periods of strong disagreement between the two bulk models such as around the day 5.6, 6.6, and 7.6

It is the assessment of the author that during the periods of time where evaporation ducting is the dominant mode of propagation the remotely sensed evaporation duct heights are reliable. It is also the assessment of the author that an indication of no surface based ducting from remote sensing is reliable.

5 Summary

During periods where evaporative ducting has been the dominant mode of propagation and the evaporation duct heights have been less than 20 meters, the results of the inversion process have been relatively unambiguous. At other times, varying degrees of ambiguity have been observed in the results. Comparisons of the remotely sensed evaporation duct height to heights calculated with bulk meteorological models using environmental measurements taken at mid-path (i.e. very high quality) are quite favorable for periods of time where the remotely sensed duct heights are unambiguous.

The greatest problem with implementation of the inverse problem methods used here is the difficulty in quantifying the certainty of parameter estimates. Due to uncertainty of the error covariance matrix used in the inversion procedures it can only be said that some parameter values are more likely than others but not the probability of the values. In operational use, remote sensing of refractivity as described in this paper is expected to be used in conjunction with predictions from a mesoscale model, data from the various *in situ* and remote sensors, and *priors* from climatic data. Determination of the joint probability of refractivity parameters using inputs from the different sensors requires knowing the conditional probability of the parameters for a given set of predictions or measurements.

Acknowledgments

This work is sponsored by the Office of Naval Research. The data used were provided by NATO AC243 Panel 3, Research Study Group 8.

References

[1] Claverie, J. and Y. Hurtaud, "Propagation tranhorizon en atmosphere marine modelisation et noveaux resultats experimentaux," AGARD CP 502, pp. 4.1 - 4.14, 1992

[2] Christophe, F., N. Douchin, Y. Hurtard, D. Dion, R. Makarushka, H. Heemskerk and K. Anderson, "Overview of NATO/AC 243/Panel 3 activities concerning radiowave propagation in coastal environments," AGARD CP 567, pp. 27.1 - 27.9, 1992

[3] Hitney, H.V., "Refractive effects from VHF to EHF part A: propagation mechanisims," NATO AGARD Lecture Series no. 196, 7 Rue Ancelle, 92200 Neuilly-Sur-Seine, France, pp. 4A1-4A13, Sep., 1994

[4] Richter, J.H. , "Sensing of radio refractivity and aerosol extinction," AGARD Lecture Series 196, pp. 3.1-3.17 Sep., 1995

[5] Monin, A.S. and A.M. Obukhov, "Basic laws of turbulent mixing in the ground layer of the atmosphere," Tr. Geofiz. nst. Akad. Nauk., SSSR, 151, pp. 163-187, 1954

[6] Liu, W.T., K.B. Katsaros, and J.A. Bussinger, "Bulk parameterization of air-sea exchanges of heat and water vapor including the molecular constraints at the interface," J. of Atmospheric Sci., 36, 1722-1735, 1979

[7] Jeske, H., "The state of radar-range prediction over sea," NATO AGARD conf. proc. 70(2), 50.1-50.10, 1971

[8] Paulus, R.A., "Specifications for environmental measurements to assess radar sensors," Naval Ocean Systems Center, San Diego, CA 92152-5000, Tech. Doc. 1685, Nov. 1985

[9] Livingston, D.C. "The Physics of Microwave Propagation, Prentice-Hall," pp. 60 and 105-106, 1970

[10] Paulus, R.A., K.D. Anderson and L.T. Rogers, "Simplified modeling of the surface layer for evaporation ducting," National Radio Science Meeting, Commission F, January 14, 1996

[11] Tarantola, A., "Inverse Problem Theory," Elsevier, Amsterdam, 1987

[12] Hitney, H.V., "Hybrid ray optics and parabolic equation methods for radar propagation modeling," IEE Radar 92 Conference Proceedings No. 365, pp. 58 - 61, 1992

[13] Tolstoy, A., "Matched field processing for underwater acoustics," World Scientific Publishing, Singapore, 1993

[14] Baggeroer, A.B., W.A. Kuperman, and P.N. Mikhalevsky, "An overview of matched field processing in underwater acoustics," IEEE J. of Ocean Engineering, vol. 18, no. 4, October, 1993

Paper 48

M.F. Levy (UK)

What is the purpose of inverse techniques in the context of refractivity sensing?

Author's reply:

The long term solution for estimating refractivity is to calculate optimal refrativity estimates using inputs from in-situ sensors, remote sensors (the "inverse propagation method" I've described is one of these), and mesoscale models. To the best of my knowledge, the inverse propagation method is the only remote sensing method except the LIDAR that is highly sensitive to humidity gradients in the troposphere. Ultimately a ship may be able to measure propagation loss from emitters of opportunity to estimate refractivity. Even the simplest measurements may indicate whether there is or is not a surface based duct.

In the short term, I expect the inverse propagation method to become an important tool for studying the marine atmospheric boundary layer. The presentation here is a case in point. In order to evaluate evaporation duct models using this data set, it was necessary to determine when evaporative ducting was the dominant mode of propagation. I have been able to do this via the inverse propagation method, which was far less expensive than launching radiosondes during the experiments would have been.

C.R. Philbrick (US)

You raised the point about the meteorology measurements only being available at mid-path. Based on the data sets that you have examined, which location is more important for the met profiles: the transmitters, receivers or mid-point?

Author's reply:

I found the mid-path meteorology for the Lorient transmission path to be highly representative of the meteorology at the transmission path. During the Variability of Coastal Atmospheric Refractivity (VOCAR) experiment, it was found that atmospheric soundings taken from shore sites negatively biased propagation estimates. Soundings taken from over-water mid-path location and from the seaward path end-point (at a small island) were found to be unbiased.

REVIEW OF MILLIMETER WAVE RADIOMETRY POTENTIALITIES. ACTIVITIES IN CERT-ONERA AND DLR-HF

POTENTIALITES DE LA RADIOMETRIE EN ONDES MILLIMETRIQUES. ACTIVITES AU CERT-ONERA ET AU DLR-HF

J. LEMORTON[1], H. SUESS[2]

(1) CERT-ONERA/DERMO, BP 4025, 2 Avenue E. Belin, 31055 Toulouse Cedex, France (Email : lemorton@onecert.fr)

(2) DLR-HF, PO Box 1116, Oberpfaffenhofen, 82230 Wessling, Germany (Email : suess@ohf015.hf.op.dlr.de)

ABSTRACT

Millimeter wave radiometry has been widely studied for more than twenty years and can provide specific advantages as compared to infrared/visible imaging and radar detection as well. This paper presents a brief review of millimeter wave radiometry techniques and applications. Recent advances in MMIC (Microwave Monolithic Integrated Circuit) technology and in focal plane array imagers are particularly highlighted. Research work at CERT-ONERA and DLR-HF which have been active in the field for more than ten years are also described.

RESUME

La radiométrie en ondes millimétriques a donné lieu à de nombreuses études depuis plus d'une vingtaine d'années et peut procurer des avantages spécifiques par rapport aux techniques d'imagerie infrarouge ou visible, et de détection radar. Ce papier décrit brièvement les principes de la radiométrie millimétrique et passe en revue les applications principales. Les récents développements dans le domaine de la technologie des circuits intégrés monolithiques microondes et des imageurs à réseau de détecteurs dans le plan focal sont particulièrement mis en évidence. Les recherches menées au CERT-ONERA et au DLR-HF sont également décrites.

INTRODUCTION

Millimeter wave radiometry has been widely studied during the last twenty years. Much theoretical work has been achieved to understand and represent radiative properties of natural surfaces and radiative transfer in the atmosphere. Extensive advances in hardware millimeter wave components and improved system understanding have resulted in better performances and higher reliability. Today's sensors are mainly devoted to Earth and Atmosphere remote sensing. Military applications have been evaluated, as well as landing and navigation aids for aircrafts. For all these applications, the advantages of millimeter wave radiometry are provided by its passive nature, moderate complexity and easy interpretation. Limitations come from poor geometrical resolution and reduced range. However recent developments on MMIC millimeter wave technology and on focal plane arrays imagers associated to advanced processing techniques are likely to overcome these limitations.

The objectives of this paper are to review the state of the art of millimeter wave radiometry and to present the research work conducted at CERT-ONERA and DLR-HF which have been active in the field for more than ten years and have established bilateral collaboration since 1989.

BASIC PHYSICAL PRINCIPLES [1]

The capability of passive millimeter wave radiometry is based on the permanent thermal radiation of objects and materials which are at temperatures above absolute zero. In such a way millimeter wave radiometry is an extension of optical or infrared radiation measurements techniques but in the microwave part of the electromagnetic wave spectrum.

The power emitted by an object (and which can be received by an adequate sensor) is given by :

$$P_e = e\,k\,T_p\,B$$

where e is the emissivity of the object
k is the Boltzmann's constant
T_p is the physical temperature of the object
B is the bandwidth in which the radiation is considered, which is in fact the sensor receiver bandwidth

$k\,T_p\,B$ corresponds to the emitted power of a blackbody at temperature T_p. In that way, the emissivity of a material can be considered as the ratio of the power emitted by this material to the power emitted by a blackbody at the same temperature. Alternatively, as the Planck law can be simplified with its Rayleigh Jeans approximation in the microwave domain, the emissivity of a material can also be considered as the ratio of the brightness temperature of the material to its physical temperature.

However in the microwave domain, the power received from an object is not only composed of the power emitted by this object but also of the power coming from elsewhere (the environment) and reflected by the object (figure 1). The reflectivity r of most materials in the microwave domain is not negligible. For an object that allows no incident RF to pass through (i.e. its transmissivity is zero), the thermodynamic equilibrium condition indicates that $e+r = 1$. As the reflectivity is known to depend on the incidence angle, on the dielectric permittivity and on the surface roughness (and of course somehow on the frequency), the microwave emissivity is also a function of all these parameters.

Paper presented at the AGARD SPP Symposium on "Remote Sensing: A Valuable Source of Information", held in Toulouse, France from 22-25 April 1996, and published in CP-582.

Finally the apparent radiometric temperature of a surface or an object consists both of a thermally emitted power and a reflected power. Atmospheric extinction and emission properties have also to be taken into account. The contrasts between various surfaces or objects will then be related to differences in emissivity and reflectivity and to a minor extent, for usual situations, to physical temperatures. Table 1 gives examples of emissivities of some common surfaces at 35 GHz and 94 GHz. Of course these values are somehow variables depending of the roughness surface, humidity of the soil, ...

Consequently, objects with low emissivity will mainly reflect the apparent temperature of the surroundings which will generally be a combination of the downwelling sky temperature and of neighboured objects temperature. This will often lead to low apparent temperatures. On the contrary, most portions of the Earth components (soil, vegetation, ...) are mostly emissive and will give high apparent temperatures.

The sky brightness temperature highly depends on meteorological conditions. Bad weather conditions lead to a more absorbing atmosphere, therefore more emissive and to higher brightness temperatures. This is the reason why radiometric contrasts will be degraded in bad weather conditions as compared to clear sky but still leading to detectable features, depending on the characteristics of the receiver.

	35 GHz	94 GHz
Forest	0.99	0.99
Grass	0.91	0.91
Asphalt	0.83	0.99
Concrete	0.76	0.91
Snow	0.60	0.78
Water	0.47	0.65
Dielectric coated metal	0.35	0.46
Painted metal	0.21	0.28
Metal	0.075	0.10

Table 1 : Examples of emissivities at vertical incidence [2]

MILLIMETER WAVE RADIOMETRY ADVANTAGES

Visible and infrared detection is more usual for imaging or remote sensing purposes in general. However millimeter wave radiometry has the advantage of being able to run well even at night, as opposed to visible systems. Passive IR sensors can perform well at night also, but millimeter wave radiometry can still operate properly through poor visibility conditions (i.e. through fog, clouds, smoke, ...), whereas IR systems performances are degraded. This possibility has strong implications either on civilian or military applications as it will be seen later on.

As compared to millimeter wave radar systems, radiometric sensors are passive and therefore discrete. No emitter source is needed and so the instrument can be made less complex and costly than radar. Furthermore another main advantage is that the interpretation of passive millimeter wave images is similar to the interpretation of IR and visible light images. This is due to the physical principles of the measurements. Visible pictures are formed by visible emission and reflection of diffuse sunlight. Millimeter wave images are formed by millimeter wave emissivity and reflection of cold atmospheric temperature. High correlation exists between the power emitted or reflected by adjacent pixels in the image. This is not the case of course for active radar images where the speckle or glint effects are very confusing for the interpretability of images.

MILLIMETER WAVE RADIOMETRY LIMITATIONS

The main disadvantage of passive millimeter wave radiometry, specially for imaging systems, is that the spatial resolution (or the pixel size) is not as good as for IR or visible systems. Spatial resolution is directly a function of the beamwidth of the antenna or collecting aperture. For a circular aperture with diameter D, this beamwidth is given approximately by $\theta = \lambda/D$ where λ is the wavelength. At a distance R on the ground, the spatial resolution will then be approximately $R\lambda/D$ for a vertical incidence. Therefore large antennas should be used at millimeter wave to achieve required resolution.

A possible way of dealing with this limitation and improving the spatial resolution is to use specific processing techniques such as superresolution, antenna beam deconvolution or inverse methods. The principle of these techniques is that, as the antenna can be considered as a low pass filter for spatial frequencies, therefore the high frequency content of the image is sharpened by the antenna. It is specially true if apodized illumination of the antenna is used in order to have low sidelobes as in radar systems. Uniform distribution is preferable in that case. Then using a priori information on the antenna behaviour (for example the antenna pattern or the point spread function), the high frequency content can then be enhanced by a processing highpass filter technique. This can improve the spatial resolution by a factor of 2 [3]. But it is usually not yet sufficient for some applications.

As up to some years ago, most imaging systems were single beam scanning instruments, the way of producing an image was by a mechanical scanning compatible with a TV raster scan. Furthermore the dwell time on a pixel is related to the sensitivity of the radiometric instrument. The more the receiver stays on the same pixel, the more radiated energy is integrated, and so it will be easier to compare it with internal thermal noise. So the overall time to produce an image was quite prohibitive for a lot of applications requiring 2D scanning. It could be suitable if one dimension could be given by the movement of the aircraft or satellite on which the radiometer is boarded. These limitations are becoming less and less problematic as it will be seen further on by the introduction of focal plane arrays.

MAIN CIVILIAN APPLICATIONS

Millimeter wave radiometry can lead to a large set of civilian applications particularly in the field of remote sensing of the atmosphere for meteorology and climatology studies, remote sensing of the Earth surface for agriculture or oceanography, and also in the field of landing aids for civil aircrafts.

Remote sensing of the atmosphere consists in the evaluation of the quantity of specific gases in the atmosphere. This is based on a measurement of the brightness temperature of the

atmosphere around frequencies which correspond to gases absorption lines. For example, measurements around 22 GHz or 183 GHz are highly sensitive to the water vapour content. The quantity of O_2 in an atmospheric layer (from which the physical temperature of the layer can be calculated) can affect the radiometric measurements around 60 GHz and 118 GHz. Specific gases in the high atmosphere can also be studied, for example ozone, ClO, N_2O at frequencies between 200 and 300 GHz. Atmospheric sensing from ground based radiometers at 20.6 and 31.6 GHz has been used for more than a decade for measuring atmospheric water vapour and cloud liquid contents. These measurements have been proved very useful for millimeter wave propagation studies which have strong implications on future Earth space telecommunications (OLYMPUS and ITALSAT propagation experiment [4]). Similar instruments have been used on satellite, still for the moment for feasibility validation of such applications (for example SSM/I [5]). Future meteorological satellite such as the next generation of METEOSAT will probably board millimeter radiometric channels. The MLS radiometer onboard the American satellite UARS is devoted to the study of the chemistry at high altitudes in the atmosphere (O_3, ClO; ...) [6].

As far as remote sensing of the Earth surface is concerned, millimeter wave radiometry has been widely used for the study of oceans, polar ice-sheets and vegetation, either from airborne or satelliteborne platforms. Not only the radiometric temperature of the ocean can give information on wind speed above the surface, but also radiometric measurements can also be used to characterise the atmosphere above the ocean and then allow other instruments measurements (for example, ATSR radiometer on ERS1 for altimeter measurements,...). Ice-sheets characterisation is possible with multifrequency radiometer in order to distinguish between snow, first-year ice and multi-year ice. A better knowledge on polar ice-sheets will find implications in climatology studies [7]. Finally radiometric measurements are also used to assess some specific aspects of vegetation grow-up processes and can lead to applications in agriculture studies [8].

Regarding navigation and landing aid applications for helicopter and civilian airliners, a lot of potential applications exist in the approach phase to the runway, as well as in the final landing phase and moving phase from the runway to the gates. Very impressive results have been obtained by DRA (UK Defence Research Agency) at Heathrow airport [9]. These applications participating to the general concept of Enhanced Vision Systems are tested in conjunction with other sensors and will require the large technological improvements appeared in the last five years (see below).

MILITARY APPLICATIONS

Military applications of millimeter wave radiometry are mainly focused on target detection even if some terrain mapping and landing aids for military aircrafts or helicopters can also be dealt with.

Target detection is clearly based on the brightness temperature contrast between targets and the background. Military targets are mostly made of metal, which even painted remains poorly emissive at millimeter wave. On the contrary, the surroundings mainly consist of vegetation, asphalt or concrete and are quite emissive. However it has to be noticed that non reflective material can be coated on targets to enhance their emissivity. Target detection is mainly concerned with battlefield surveillance concepts or missile and bomb guidance systems. In the first case, radiometer instruments on the ground or boarded on helicopter can provide information on targets at moderate range even in adverse weather or obscurant presence. In the second case, brightness temperature contrasts between targets and their surroundings can be used for the detection and guidance at low range towards the target. Battlefield damage assessment through smoke can also be valuable, whereas burning areas are also highly detectable due to high physical temperature contrast. Ship detection at sea is another possible application leading to routes surveillance concepts.

Mapping applications could probably mainly be used for navigation (cruise missile, airplanes or UAVs) [10]. Real-time millimeter wave images of an area can be correlated with stored optical reference maps for determining relative location of the platform.

All these potential applications explain the important research and development efforts made in the eighties in the USA [11], but also in Europe. However due to the lack of spatial resolution resulting in too short detection range or insufficient mapping capabilities, these studies have not led to real operational use of the millimeter wave radiometry technique for military systems.

However new technology advances as well as use of innovative concepts achieved since the beginning of the nineties are probably going to change this situation. A renewal of military interest for millimeter wave radiometry has been observed for two or three years [12].

NEW DEVELOPMENTS

Recent advances in radiometry techniques are driven by the progress in MMIC technology, specially at millimeter waves. Two aspects are concerned by these developments, on one hand the technology and principle of the radiometric receiver itself, on a second hand the concept of the imaging system.

Up to the end of the 80's, millimeter wave radiometer receivers were based on a superheterodyne approach. In that way they used a mixer on a first stage, then IF amplifier to achieve the large gain necessary for so weak signals detection. This approach was almost mandatory in the absence of low noise amplifier in the millimeter wave band. However it suffers from high conversion loss and high noise figure at the mixer stage and also requires very good local oscillator. In terms of volume and cost, this approach made millimeter wave passive imaging arrays almost impossible.

An alternative approach is to use a direct detection receiver, simply composed of a high performance LNA (Low Noise Amplifier) and a detector circuit. In that case, no local oscillator is needed and DC power consumption as well as cost and volume can be highly reduced. This direct detection approach has been made possible for two or three years by the development and advances of millimeter wave MMIC LNAs. This is specially true in the USA due to the MMIC

ARPA program. Similar actions have been undertaken on a smaller scale on a trilateral basis in France, UK and Germany.

A W-band monolithic LNA has been reported using pseudomorphic HEMT in 1991 [13]. New improvements have been achieved each year since that time. Modules requiring several MMIC LNAs in series due to insufficient gain for a single LNA have been replaced by a multistage LNA single chip. Circuit unstability and assembly difficulty have therefore been improved and performances as well in terms of higher gain, larger bandwidth and moderate noise figure [14].

Advances in MMIC technology have enabled a new generation of millimeter wave radiometric imagers by allowing the design and realisation of focal plane array systems. It has been seen above that usual single channel scanning systems were unable to fill all the technical requirements for fast high resolution imaging systems. An imager with a multibeam capability, either a plane aperture phased array or a focal plane array, can simultaneously view all the pixels of its field of view. Then the dwell time on each pixel is no longer constrained by the frame time. This results in better radiometric sensitivity. MMIC radiometric receivers have been shown to be ideally suited for focal plane array systems. For example, Millitech has developed a 16x24 channels receiver at 94 GHz capable of producing a 32x48 pixel image. Each channel is however still of the superheterodyne type and gives a sensitivity between 1 K and 2 K for an integration time of 33 ms [15]. A multibeam focal plane array for airborne imaging systems has also been reported (with 24 receivers part of 192 fully populated array) [16]. TRW has concentrated considerable research effort in that domain reporting first a 1x8 receiver array [17], then developing a W-band 1040 pixels focal plane array receiver to be tested this year [18]. These latest systems from TRW fully use MMIC LNAs for direct detection architecture. The integration time for each pixel is about 2 ms only and the temperature sensitivity about 0.3 K at 94 GHz. With this equipment known as a passive millimeter camera, the acquisition of millimetric scene images in real time is to be reached with frame rate similar to video frame rates (10 to 30 Hz).

Another domain that has been concurrently improved concerns signal and image processing techniques. As it has already been mentioned above, superresolution techniques have been demonstrated to operate real time and can improve the spatial resolution usually limited by Fresnel diffraction by a factor of 2 [3]. Other more promising processing concepts come from radioastronomy interferometric techniques and lead to interferometric or synthetic aperture radiometers [19]. The main characteristics of this technique is to use partially filled array which can image a scene with high resolution but not too many receiver channels. Signal processing is based on correlation between each pair of channels and inverse Fourier transform, and can lead to large computational load (depending on the number of channels). Advances in computer capacity and speed would still be required in that domain. The validity of the concept for microwave radiometry has been demonstrated, for example by the ESTAR instrument at NASA [20], but others experimentations have also been run elsewhere.

The association of MMIC direct detection millimeter wave receiver, with focal plane arrays imaging techniques and some interferometric signal processing is probably one of the most challenging, but realistic, objective for the following years in millimeter wave radiometry. Anyway all these recent advances are susceptible to enlarge the operational applications of radiometric imaging since it overcomes the main limitations experienced at the end of the 80's.

ACTIVITIES AT CERT

CERT-ONERA has been active in the field of microwave and millimeter wave radiometry for more than ten years. The nature of the activities has been equally shared into measurements campaign for the acquisition of radiometric signatures, feasibility of applications and theoretical modelling for the development of simulation tools.

A 35 GHz single channel mechanical scanning was developed for CELAR (Centre d'ELectronique de l'ARmenent, a French MoD technical center in Rennes) at the end of the 80's. The main objective was to get a very sensitive and accurate instrument in order to fill a signature data bank of various objects and surfaces. A prototype of 35 GHz electronically scanning radiometer has also been designed with millimeter wave ferrite phase shifter (which have been studied at CERT for more than 20 years) in order to evaluate the feasibility of mapping and navigation for cruise missile. A multifrequency radiometer for airborne remote sensing applications has also been evaluated.

Various measurements campaigns have been implemented with the 35 GHz single channel radiometer. Apart from military vehicles signature acquisition directly achieved by CELAR, CERT has run the instrument for the measurement of ship signatures, and of atmosphere or Earth surfaces brightness temperatures. For example, figure 2 shows the brightness temperature of a 120 m length military ship passing through the radiometer antenna beam. In vertical polarisation, the ship appears colder than the sea, whereas in horizontal polarisation an opposite variation is obtained. These measurements taken from a 300 m height cliff are characterised by moderate distance between the sensor and the ship (<1000 m) and low grazing angle ($\approx 15°$). These results have been used after some modelling study and extrapolation to evaluate the feasibility of ship detection from space [21].

Observations of supercooled water associated with updraft in stratiform ice clouds have been achieved using 35 GHz microwave radiometer and Doppler meteorological radar in Lannemezan [22]. In this case the high sensitivity of the microwave radiometer to the presence of liquid water has been a decisive feature in the sense that the 35 GHz radar was not too much affected by this liquid water when included in ice clouds with high reflectivity ice particles. CERT has also participated to an experiment conducted in the French Alps in order to measure snow signature on glaciers. This experiment was run with CNRS-LGGE and CNES in the aim of validating snow microwave emission models [23].

Concurrently with these experiments, a great deal of efforts have been put in theoretical modelling of emission and scattering of natural surface or elements. The main objective of this type of work is the achievement of simulation software which can constitute useful tools for the design and evaluation of radiometer systems for new applications.

For example the effects of absorption and scattering phenomenon in the atmosphere have been studied in order to assess the variations of radiometric contrasts received by a radiometric imager in various meteorological conditions. Bad weather conditions may degrade the quality of radiometric images. The analysis has been based on radiative transfer theory whose vectorial equation has been solved by the invariant imbedding method [24]. Atmospheric absorption by gases and hydrometeors has been included as well as hydrometeors scattering effects which can be of importance for millimeter wave systems. Numerical simulations have been focused on atmospheric window frequencies for airborne or spaceborne instruments over various type of surfaces. Figures 3 and 4 illustrate these effects.

Millimeter wave passive signatures of terrains have also been evaluated. For low frequencies, stratified multilayer terrains have been dealt with using Njoku and Wilheit theory. Simulations have been achieved for the evaluation of soil moisture content in various soils (cultivated field, sands, ...) and compared to in-situ measurements [25]. The influence of the roughness of the surface has then be taken into account using small perturbation theory or physical optics depending on the ratio of the wavelength to the roughness scale. Figure 6 illustrates the effect of roughness for bare soils [26].

CERT-ONERA is currently participating to the development of a much global simulation code which is devoted to near real time simulation of the engagement of a missile with its sensor (IR, radiometer or radar) against terrestrial targets in their natural environment. The main part of the project is a scene generation software which must provide data time sequence to the sensor from a geographical information database. As far as the radiometric aspects are concerned, some features are somehow similar to the ARMSS simulation code developed by TRW [27]. All the backgrounds are modelled as deterministic geometric patches of statistical described features (emissivity and reflectivity distributions in the case of passive millimeter wave). For millimeter wave radiometry, targets are also described by a facetization. The core of the calculation is based on ray tracing methods. Atmospheric effects are also included. This simulation code is being developed with OKTAL Company for the French MoD and will make use of all the theoretical and experimental background acquired at CERT in the last ten years.

Others studies at CERT on microwave radiometry techniques have dealt with the concept of digital beam forming applied to radiometer systems and to interferometric radiometer systems [28]. Feasibility evaluations of landing and navigation aid for helicopter or airliners are also conducted.

ACTIVITIES AT DLR

At DLR the main activities in microwave radiometric imaging have been focused on the development of advanced systems, specially on the improvement of the spatial and radiometric resolution, as well as on signature measurements and their analysis.

DLR has built several radiometers systems in the last ten years, either for imaging capability validation at 90 GHz and 140 GHz, or for operational systems for the detection of oil spills at sea and also for synthetic aperture radiometric technique validation at 35 GHz.

A complete and detailed description of the modular 90/140 GHz system can be found in the literature [29]. The principal system configuration used with this airborne imager employs a sinusoidal mechanical scanning perpendicular to the flight track (see figure 6). This system has been specified for altitudes down to 150 m to obtain a complete ground coverage and to detect objects with a size of 2 m. One key specification is the theoretical data rate of about 10^4 samples/s for a temperature resolution of 1 K at a target temperature of 300 K. A second key specification is the temperature resolution of better than 0.4 K at an integration time of 1 ms, and could be realised only by the use of cooled front ends.

Various signature acquisition campaigns have been achieved with this instrument in order to evaluate possible applications in airborne reconnaissance, particularly for the detection of vehicles on roads, for the detection and classification of airports and airplanes and for the detection of high value targets (oil tanks, buildings..). Representative measurements results will be presented during the oral presentation.

For example [30], several flights at an altitude between 150 m and 200 m over a highway with moving personal cars and trucks have been done. In both cases (90 GHz and 140 GHz), the target/background contrast is very high, the size of vehicles clearly distinguishable. Similar flights over oil refineries allowed large oil tanks to be identified, as they reflect the cold sky temperature. Other structures from the refineries have been recognised but with less contrasts. Parked aircrafts in front of airports buildings have also been imaged. The metal roof of buildings implies a colder temperature than for aircraft, whereas the surrounding runway and vegetation exhibit warmer temperatures.

Apart from that experimental work on radiometric signature, much efforts have also been given towards radiometric data processing techniques. One of the most important part of image analysis is to detect and to extract edges from an image because considerable information lies in contours. Many objects can be recognised by their shape, even though relatively low spatial resolution of radiometer systems. Various edge detection techniques have been evaluated at DLR (Sobel, Marr-Hildreth, Cany edge detector, ...) from what it was concluded that a method of edge focusing was found to be the most efficient [31]. Figure 7 is an example of the results obtained with that method. Other processing techniques useful for improving the information content of the radiometric images, such as false colour enhancement, smoothing, histogram modification have also been demonstrated [32]. Finally antenna mainbeam deconvolution techniques have been studied and implemented in order to improve the spatial resolution [33].

In order to improve a bit more the spatial resolution, synthetic aperture imaging systems are very promising and could give rise to more effective instrument with higher flight altitude, larger swath widths where mechanical scanning could be replaced by signal processing on larger unfilled aperture. DLR activity on that subject has been dealing with the design of an experimental measurement

system at 35 GHz in order to identify and optimise critical parts of this new imaging method. A first step has been to build up a test equipment which a sketch is shown at figure 8 . The basic idea is to move antennas along a tee structure to change the baseline lengths, then sampling up the spatial frequency spectrum of the image. The real image is then obtained by Fourier transform. Various aspects of the problem have been assessed such as the realisation of analog correlator, the test of calibration methods, the comparison of reconstruction algorithm, ... Experimental measurements have also been performed and have demonstrated very good results and improved spatial resolution [34]. The application of such a concept to an airborne context still remains to be done, with some optimisation and compromises to be found on complexity, number of elements, IF bandwidth, integration time, computational load, and so on.

CONCLUSIONS

Whereas the interest for millimeter wave radiometer seemed to have been less pronounced at the end of the 80's (as compared to the beginning of the 80's), recent advances since the beginning of the 90's have now put a hot spot again on that technique. MMIC technology at millimeter wave (specially W-band) is now allowing direct detection receiver configuration, that is perfectly suited to focal plane array imagers. Synthetic aperture as interferometric processing can also yield specific improvements either alone by itself, or combined with new receiver technology.

The main goal of these developments is the improvement of spatial resolution and radiometric sensitivity without sacrifying the overall frame time to build an image. In that way, millimeter wave radiometry will likely be able to fulfil most of the requirements for imaging applications and will really demonstrate all the potential advantages inherent to its passive and all-weather nature. Apart from satellite remote sensing applications that are already running, the most likely applications areas will probably deal with civilian and military navigation and landing aids imaging systems, as well as battlefield surveillance sensors and missile seekers passive sensors for dual mode operation. The NATO/RSG 20 measurement campaign MIMEX (Millimeter wave Imaging Experiment) that is planned in 1996-1997 will experiment, apart from active millimeter wave radars, some millimeter wave radiometers. CERT-ONERA and DLR-HF will anyway be active in the field in the following years.

REFERENCES

[1] ULABY F.T., MOORE R.K. and FUNG A.K. : "Microwave Remote Sensing Active and Passive", vol. 1, Addison-Wesley Publishing Company, 1981

[2] WALLEZ C., COMBES P.F., LEMORTON J. : "Model for the simulation of atmospheric effects on targets to background passive signatures", AGARD CP 542, May 1993, Paper 31

[3] APPLEBY R., ANDERTON R.N., LETTINGTON A.H. : "An optical feed for high resolution passive millimeter wave imaging system", IEE Colloquium on Millimeter and submillimeter wave heterodyne receivers, London, Oct. 1992, pp 15/1-4

[4] F. BARBALISCIA et al. : "OPEX Reference Book on Radiometry and Meteorological Measurements", ESA WPP-083, 2nd Workshop of the OLYMPUS Propagation Experiments, Noordwijk, Nov. 1994

[5] Special Issue on the Defence Meteorological Satellite Program (DMSP) : "Calibration and Validation of the Special Sensor Microwave Imager (SSM/I)", IEEE Trans. Geoscience and Remote Sensing, Vol. 28 n°5, Sept. 1990

[6] PARRISH A. : "Millimeter wave Environmental Remote Sensing of Earth's Atmosphere", Microwave Journal, Dec. 1992, pp. 24-34

[7] M. FILY, BENOIST J.P. : "Image scale statistical study of Scanning Multichannel Microwave Radiometer (SMMR) data over Antartica", Journal of Glaciology, vol.37 n° 125, 1991, pp. 129-139

[8] J.P. WIGNERON et al. : "Simultaneous retrievals of Soil moisture and Vegetation Biomass from multiconfiguration Passive Microwave Measurements", IGARSS'95 Conference Proceedings, Florence, July 1995

[9] APPLEBY R. : "Passive millimeter wave imaging : seeing in very poor visibility" SPIE vol 2463 Synthetic Vision for Vehicle Guidance and Control, April 1995, pp. 10-19

[10] WILSON W.J., IBBOTT A.C. : "Millimeter wave Imaging Sensor Data Evaluation", JPL Publication 87-16, May 1987

[11] HARROP J.D., STUMP R., TETI J.J. : "Surface Navy Application of millimeter wave sensors. Vol 1 Design Trade off Study", NSWC TR-3748 Report, Nov. 1977

[12] FRITZ J. : "Groupe de Recherches OTAN RSG-20. Applications militaires de l'imagerie millimétrique", Rapport ONERA/RCS n° 710595/SAT2, 1994, Classified

[13] WANG H. et al. : "A high performance W-band monolithic pseudomorphic InGaAs HEMT LNA", IEEE MTT/S Digest, June 1991, p. 943

[14] LO D.C.W. et al. : "A monolithic W-band High gain LNA/Detector for Millimeter wave Radiometric Imaging Applications", IEEE MTT/S Digest, Orlando, May 1995, pp. 1117-1120

[15] HUGUENIN G.R. : "A millimeter wave focal plane array imager", SPIE Intern. Conference on Millimeter and Submillimeter waves and application, Conference Digest, 1994, p. 300-301

[16] HAAS R.W. et al. : "Multibeam millimeter wave focal plane array airborne imaging system", SPIE Intern. Conference on Millimeter and Submillimeter waves and Applications, 1994, pp. 302-311

[17] LO D.C.W. et al. : "A W-band direct detection radiometric imaging array", Proc. of IEEE Microwave and Millimeter wave Monolithic Circuits Symposium, San Diego, May 1994, pp 41-44

[18] SHOUCRI et al. : "A Passive Millimeter Wave Camera for Aircraft Landing in Low Visibility Conditions", IEEE AES Systems Magazine, May 1995, pp 37-42

[19] RUF C.S. et al. : "Interferometric Synthetic Aperture Microwave Radiometry for the Remote Sensing of the Earth", IEEE Trans. On Geoscience and Remote Sensing, vol. 26 n° 5, Sept. 1988, pp 597-611

[20] LE VINE D. et al. : "Status of ESTAR Validation : Results from Washita 92" IGARSS'94 Conference Proceedings, Pasadena, August 1994

[21] J. LEMORTON ; "Signatures radiométriques des navires en ondes millimétriques", Rapport Technique n° 1/1446/MO, April 1990, Classified

[22] C. WALLEZ, H. SAUVAGEOT, J. LEMORTON, P.F. COMBES : "Observation of supercooled water associated with updraft in stratiform ice cloud using microwave radiometer and Doppler radar", EGS Symposium, Grenoble, May 1994

[23] SHERJAL I. et al. : "Microwave Remote Sensing of snow from a cable car at Chamonix in the French Alps", IGARSS'95 Conference Proceedings, Florence, July 1995

[24] C. WALLEZ : "Absorption et diffusion des ondes électromagnétiques par les hydrométéores. Application à l'imagerie radiométrique en ondes millimétriques", PhD Thesis, Toulouse, April 1994

[25] F. COSTES et al. : "Microwave Radiometry on Bare Soils : comparison of various emission models of layered media with measurements", IGARSS'94 Conference Proceedings, Pasadena, August 1994

[26] F. COSTES : "Contribution à l'étude des signatures actives et passives micro-ondes des sols nus. Application à la télédétection", PhD Thesis, Toulouse, Dec. 1994

[27] HAUSS B.I. et al : "Advanced Radiometric and Interferometric Millimeter wave Scene Simulations", NASA Proc. of the Workshop of Augmented Visual Display Research, Dec 1993, pp133-162

[28] LEVEQUE H., J. LEMORTON : "Evaluation of Digital Beamforming Techniques applied to Microwave Radiometer systems", PIERS'94 Symposium, Noordwijk, July 1994

[29] VOVINKEL B. et al. : "Airborne Imaging System using a Cryogenic 90 GHz receiver", IEEE Trans. MTT, Vol. 29 n° 6, June 1987, pp. 535-541

[30] SUESS H. : "Airborne and ground based radiometric imaging between 32 GHz and 140 GHz for military applications", private communication

[31] SUESS H. : "Processing and analysis of radiometer measurements for airborne reconnaissance", SPIE Vol. 1341, Airborne Reconnaissance XIV, San Diego, July 1990, pp. 103-114

[32] WILSON W.J., SUESS H. : "Millimeter wave sensor image analysis", SPIE vol. 1156, Airborne Reconnaissance XIII, 1989, pp. 170-176

[33] SCHMID F. : "Erdbeobachtung mit dem multifrequenten SSM/I Radiometer bei horizontaler und vertikaler Polarisation", Studienarbeit Document, DLR, August 1992

[34] SPEICHL M., SUESS H. : "Theory and Design of an Experimental Aperture Synthesis Radiometer (ASR)", PIERS'94 Conference Proceedings, Noordwijk, July 1994

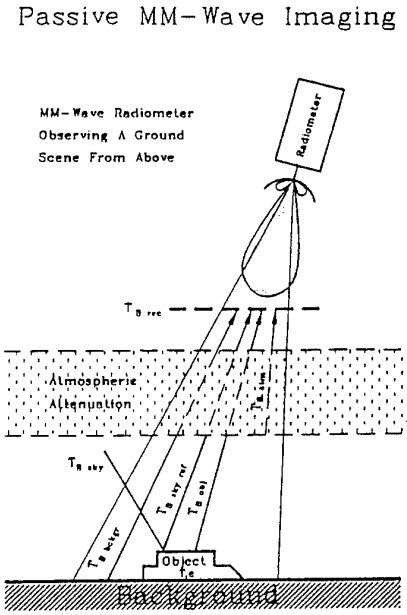


Fig.1 : Millimeter wave radiometer observing a ground scene from above

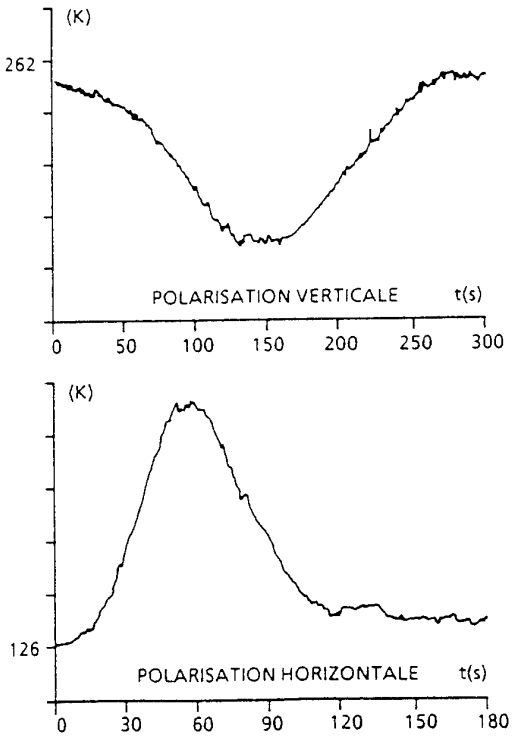


Fig. 2 : Radiometric temperatures measured when a ship goes through the radiometer mainbeam

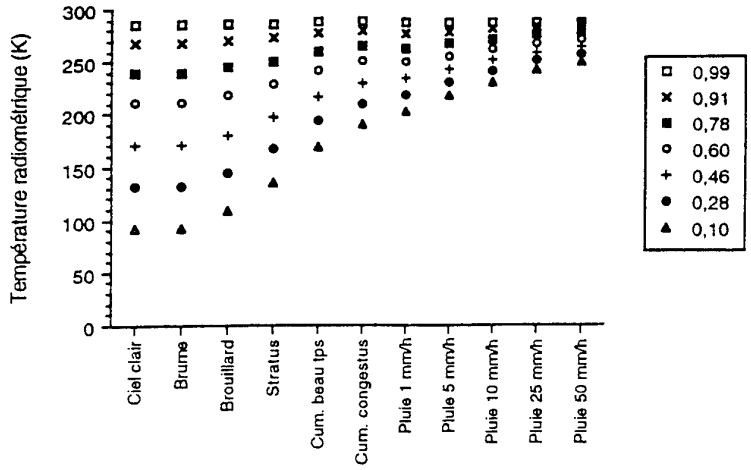


Fig. 3 : 94 GHz simulated apparent temperature on a 100 m altitude airborne radiometer for various meteorological conditions and various surface emissivities

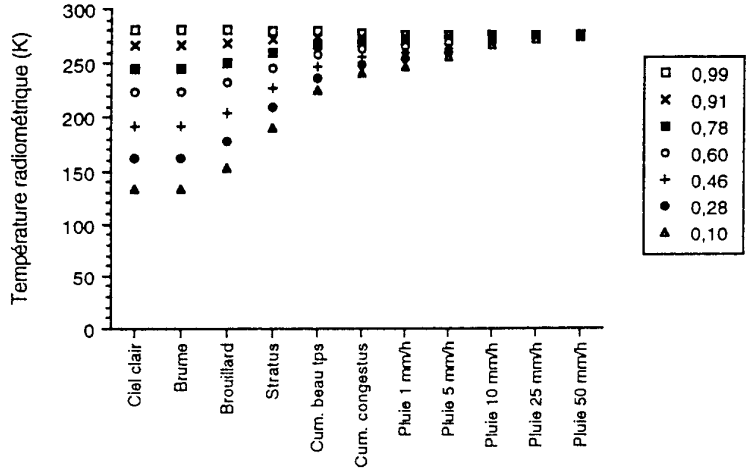


Fig. 4 : 94 GHz simulated apparent temperature on a satelliteborne radiometer for various meteorological conditions and various surface emissivities

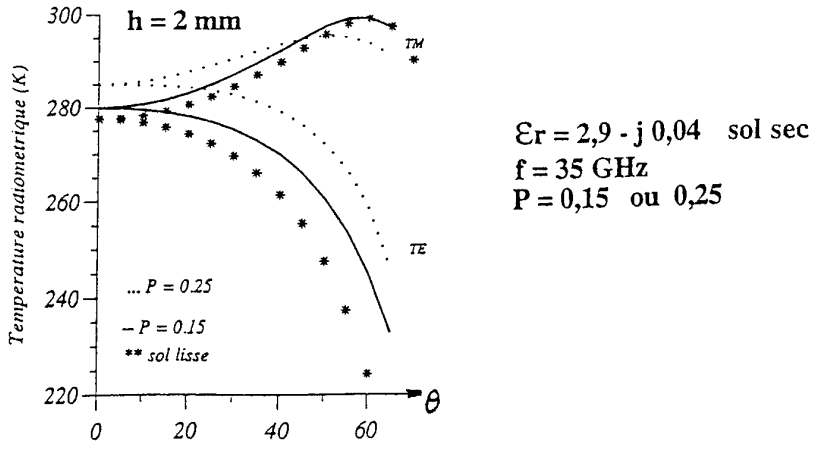


Fig. 5 : Effects of the surface roughness on radiometric temperature

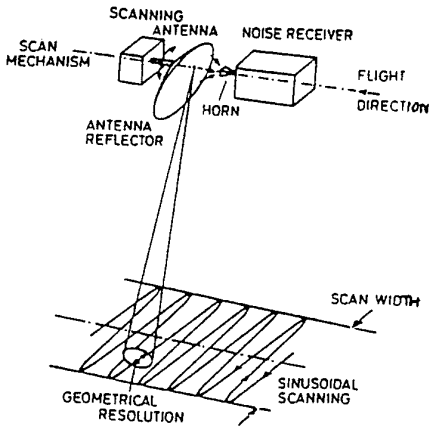


Fig.6 : System configuration with the DLR millimeter wave sensor mounted on an airborne platform and the scan on ground

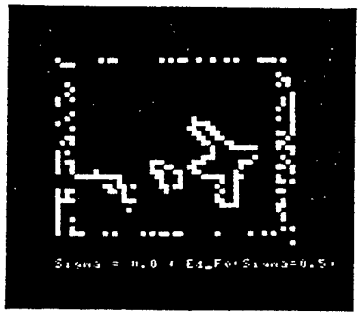


Fig. 7 : Result of the edge focusing method

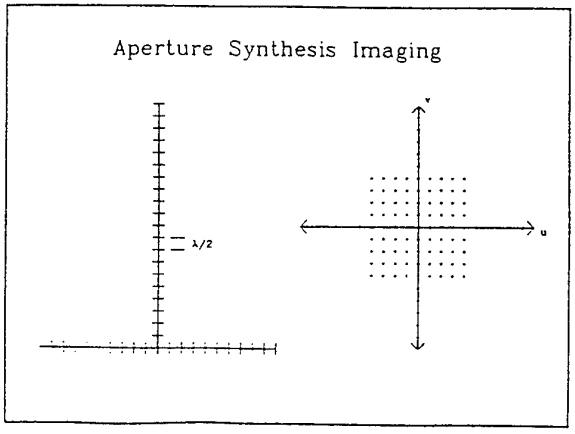


Fig 8. : Principal configuration of the ground based aperture synthesis imaging system.
Two antennas are moved in λ/2 steps along the tee in order to get a complete coverage of the uv plane

EXPECTATIONS FROM EMET AND OESEE GROUND LISTENING OF POSSIBLE E.M. EVENTS GENERATED BY THE TSS 1R

Robert D. Estes
Harvard-Smithsonian Center for Astrophysics, Cambridge,MA,02138,USA

Juan Sanmartin
Universidad Politecnica de Madrid, E.T.S.I. Aeronauticos, P.za Cardenal Cisneros 3, 28040 Madrid, E

Giorgio Tacconi
University of Genoa DIBE ,16145, and Consiglio Nazionale delle Ricerche, IAN,16169- Genoa , Italy

ABSTRACT - On 22nd February '96, the space mission STS 75 started ,from the NASA facilities at Cape Canaveral. Such a mission consists in the launch of the shuttle Columbia in order to carry out two experiments in the space: the TSS 1R (Tethered Satellite Sistem 1 Refliight) and the USMP (United States Microgravity Payload). The TSS 1R is a replica of a similar mission TSS 1 ('92) . The TSS space programme is a bilateral scientific cooperation between the USA space agency NASA (National Aeronautics and Space Agency) and the ASI (Italian Space Agency). The TSS 1R system consists on the shuttle Columbia which deploys, up-ward, by means a conducting tether 20 km long, a spherical satellite (1.5 mt diameter) containing scientific instrumentation. This system, orbiting at about 300 km from the Earth's surface, represents, presently, the largest experimental space structure. Due to its dimensions, flexibility and conducting properties of the tether, the system interacts, in a quite complex manner, wih the earth magnetic field and the ionospheric plasma, in a way that the total system behaves as an electromagnetic radiating antenna as well as an electric power generator. Twelve scientific experiments have been assessed by US and italian scientists in order to study the electrodynamic behaviour of the structure orbiting in the ionos phere. Two experiments have been prepared in the attempt to receive on the Earth's surface possible electromagnetic events radiated by the TSS 1R. The project EMET (Electro Magnetic Emissions from Tether) ,USA and the project OESEE (Observations on the Earth Surface of Electromagnetic Emissions) Italy, consist in a coordinated programme of passive detection of such possible e.m. emissions. This detection will supply the verification of some thoretical hypotheses on the electrodynamic interactions between the orbiting system, the Earth's magnetic field and the ionospheric plasma with two principal aims as the technological assesment of the system concept as well as a deeper knowledge of the ionosphere properties for future space applications. A theoretical model that keeps the peculiarities of tether emissionsis being developed for signal prediction at constant tether current. As a step previous to the calculation of the expected ground signal , the Alfven-wave signature left by the tether far back in the ionosphere has been determined. The scientific expectations from the combined effort to measure the entity of those perturbations will be outlined taking in to account the used ground track sensor systems.

1. Introduction

The EMET and OESEE observation programme , represents the *receiver* on the Earth's surface of a possible signal generated and irradiated by the TSS 1R orbiting system considered as the *transmitter.* The most promising Earth's surface positions for receiving stations are points on the ground track . The selected receiving sites are : Mona Island , Arecibo (Portorico) and Breebie Island (Australia) are under the EMET responsability, the other two stations , at Canary Islands (Spain) and Malindi (Kenya) are under the OESEE responsability. Based on previous theoretical studies, [1, 2, 3] an estimation on some measurable parameters of the expected signal at the ground station has been deduced. The generation/radiation

Paper presented at the AGARD SPP Symposium on "Remote Sensing: A Valuable Source of Information", held in Toulouse, France from 22-25 April 1996, and published in CP-582.

mechanisms are strongly dependent from time and space due to a variety of phenomenologies related to the Earth's magnetic distribution, to the ionospheric plasma, to the propagation boundaries conditions . In addition, the inherent electrodynamic interactions of the ionosphere with the orbiting huge and flexible structure, makes the modelling of the global behaviour extremely complicated and il-posed , with drastic reduction of the quality of the parametric estimations and predictions in terms of accuracy, precision, stability. In spite of the above considerations the theoretical study of modelling approach has produced several hints for the assessment of high quality e.m. detection systems suitable to be deployed at the selected receiving sites. A tethered system as a result of its interaction with the earth's magnetic field and with the plasma can radiate *spontaneous* radiations as well as *artificial* emissions generated on board of the system by means of planned modulating devices. Each of those possibilities requires an ad hoc detection/estimation strategy. Since these waves are the means by which current closure is achieved in the ionosphere , understanding them is of fundamental importance to understanding the basic physics of the system , Additionally, knowlwdge of the propagation characteristics of these waves is important for their possible applications to communication systems. In its basic form, an electrodynamic tether can be viewed as a single orbiting system with a long electrically conducting cable thais insulated from the ionosphere and stretches along the vertical with terminating masses at each end which are in contact with the ionosphere. As the system passes through the geomagnetic field , the system experiences an induced emf between the tether ends that is given by $\left(\bar{V}\times\bar{B}\right)\cdot\bar{L}/c$, where $\bar{V},\bar{B}$ and $\bar{L}$, are, rispectively, the system velocity, the geomagnetic field vector and the vector parallel to the tether whose magnitude is the tether length L. The motion induced emf produces a near-steady current flow in the tether as long as the terminal masses (the shuttle and the satellite) can exchange charge with the ambient plasma [4]. the terminal masses must provide good contact with the ionosphere or the system ability to draw the current will be restricted. If current collection is insufficient , most of the motion induced emf will go into charging the terminal masses to large potentials and diminish the intensity of electromagnetic emissions. Many studies have been done wich address the issue of wave excited by a conducting mass moving through a magnetized plasma . Some of these have specifically addressed the tether problem . Although it has been demonstrated that the tether of the kind employed for TSS will give rise to propagating electromagnetic waves , the production of waves which are intense enough to be observed by ground-based receiving systems has not been well established.

2. The Spontaneous Emission Model

Properly equipped satellites already in orbit might be used to measure waves left behind in the ionosphere by tethers as well as appropriate sensors sytems located at the Earth's surface on the ground track. A good theoretical representation of the plasma waves [5] associated with a tether hundreds of kilometers from the satellite might guide the planning of data collection and analysis, which could in turn validate the theory. The considered signature is that generated by a tether of length L , operating as truster or generator, and thus supporting a near steady current I . In addition to cold-plasma thory , can be assumed an infinite , uniforme ionosphere; constant, horizontal tether velocity $\bar{V}$ and a constant geomagnetic field, perpendicular to $\bar{V}$ but tilt at an angle θ with respect to the horizontal, Fig.1. As a steady current flows through the tether, by means of continual charge exchange with the ambient plasma at end contactors, circuit closure is accomplished by charge-carrying electromagnetic waves radiated to "infinity" by the passage of the system . The constraints of steady-state operation and negligible (thermal) damping restricts rdiation to 2

brabches of the cold-plasma dispersion relation: Fast Magnetosonic (Lower Hybrid range rather than whistler's) and Alfven branches . The "wing" structure of Alfven-wave packets , is considered (frequencies below the ion gyrofrequency Ω_i).

Electric and magnetic fields $\bar{E},\bar{B}$ in the Alfven pachet written in terms of a potential field [11],

$$\bar{E} = -\nabla_\perp \phi \ , \qquad \partial \bar{B} / \partial t = -c\nabla \wedge \left(\bar{1}_z \partial \phi / \partial z \right) \ ,$$

$$\phi(\bar{r}) = \frac{I}{\pi} \iiint \frac{d^3 k g(\bar{k}) e^{i\bar{k}\cdot\bar{r}}}{\omega k_\perp^2 \left[\varepsilon_1(\omega) - (ck_z/\omega)^2 \right]} \ , \quad (1)$$

$\underline{r} \equiv (x - Vt, y, z)$; $\omega = k_x V$ (steady state);

$\perp \equiv$ perpendicular to the geomagnetic field;

$$\varepsilon_1(\omega) = \frac{c^2 / V_A^2}{1 - \left(\omega / \Omega_i^2 \right)} \ , \qquad V_A \equiv \text{Alfven speed} \ ,$$

(2)

(diagonal component of cold-plasma dielectric tensor perpendicular to geomagnetic field)

$$g(\bar{k}) = \frac{-i}{2\pi I} \iiint d^3 r \nabla \cdot \bar{J}_{source} e^{-i\bar{k}\cdot\bar{r}} \qquad (3)$$

$\bar{J}_{source}(\bar{r})$ is current density in the tether; regions of net cgarge continually created in the plasma as it flows by the contactors, excite the waves. For (contactor length along x) << $V/\Omega_i \cong 36$ m,

$$g = \frac{1}{\pi} sin\left(\frac{L}{2} k_y' \right) \equiv \frac{1}{\pi} sin\left[\frac{L}{2} \left(k_y \cos\theta - k_Z sin\theta \right) \right]$$

(4)

Fields decay exponentially away from the planes $y = \pm 1/2\ L \cos\theta$.

Alfven wave-packets are called wings; they are disjoint if

$$L\cos\theta \ \rangle \ 4|x|^{1/3} (V/\Omega_i)^{2/3} \ , \text{ roughly} \qquad (5)$$

Leading front (of top wing, at $y = 1/2 \cos\theta$) occurs at

$$\frac{|x|}{z + 1/2 L sin\theta} \cong \frac{V}{V_A} \cong \frac{1}{40}$$

$$\frac{|x|}{z + 1/2 L sin\theta} \cong \frac{V}{V_A} \cong \frac{1}{40} \quad .(6)$$

Collision destroy the wing at a distance along the front , z ~ 600 km, or $x \cong zV/V_A \cong 15\ km$.

Front has Airy-function structure, Fig.2,

E_y

$$= \frac{V_A I}{c^2} \frac{\pi (2/3)^{1/3}}{|x|^{1/3} (V/\Omega_i)^{2/3}} \times$$

$$A_i(\zeta) \times sign(y - L/2\cos\theta) \qquad (7)$$

$$E_x = \frac{V_A I}{c^2} \ \frac{\pi (2/3)^{1/3}}{|x|^{1/3} (V/\Omega_i)^{2/3}} \times \ G_i(\zeta)$$

$$\text{with } \zeta \equiv \frac{x + (z + 1/2 L sin\theta) V/V_A}{|x|^{1/3} (V/\Omega_i)^{2/3}} \left(\frac{2}{3} \right)^{1/3} .$$

Front thikness grows as (distance along the front $)^{1/3}$; amplitude decays as inverse of thikness.

Frequencies in the range : $\omega/\Omega_i \le 0 \left[\left(V/|x|\Omega_i \right)^{1/3} \right]$ contribute to the front structure.

Behind the front,only wave components close to a single well defined frequency is found to radiate along each ray or direction. That frequency , ω_*, is given by :

$$\left(\frac{\omega_*}{\Omega_i}\right)^2 \equiv K_*^2 = 1-\left(\frac{V}{V_A}\ \frac{z+1/2\,L\sin\theta}{|x|}\right)^{2/3}$$

$$\left(\frac{\omega_*}{\Omega_i}\right)^2 \equiv K_*^2 = 1-\left(\frac{V}{V_A}\ \frac{z+1/2\,L\sin\theta}{|x|}\right)^{2/3} \quad (8)$$

for the top contactor. The region behind the front is defined by the condition : $\omega_* / \Omega_i >> \left(V / \Omega_i |x|\right)^{1/3}$.

The far fields behind the front of the top wing are :

$$E_y = \Lambda \times \cos\left(\frac{x\Omega_i}{V} k_*^3 + \frac{\pi}{4}\right) \times sign(y - L/2\cos\theta) \quad (9)$$

$$E_x = \Lambda \times sin\left(\frac{x\Omega_i}{V} k_*^3 + \frac{\pi}{4}\right)$$

with $\Lambda = \left(\frac{2\pi}{3}\right)^{1/2} \frac{V_A I}{c^2 k_*^{1/2}} \frac{1-k_*^2}{\left(|x|V/\Omega_i\right)^{1/2}}$

The decay as (distance)$^{-1/2}$.

3. The Observational Program

In 1986, the Harvard-Smithsonian .Astrophysical Observatory (SAO) was selected by the National Aeronautics and Space Administration (NASA) to investigate analitically and observationally using ground based systems the detectability of emissions by tthe electrodynamic tether . The title of this project was E.M. Emissions from Tether (EMET) [12]. Rice University is also associated to EMET with its experience in ground-based ionospheric observational systems. In 1987 the italian Space Agency approved the proposal of the University of Genoa for a similar program using coil and superconductive magnetometers (SQUID) in two sites on the ground track of the orbiting system. With the italian experiment the total of coordinated selected receiving locations where in Australia (Bribie Island), the Caribbean (Mona Island) ,the Canary Islands (Tenerife), and Kenya (Malindi) . EMET programme included also additional observations during the overflights of the TSS 1 system using the Arecibo (Puerto Rico) 430 MHz radar facility. In Fig. 3 are indicated the above mentioned coordinated receiving sites. The Arecibo facility is ideal for diagnosing the effects of the TSS on the ionosphere and providing meaningful results to the programs global data pool. The combination of different data during the overflights, namely the minimal closest points of approach of the TSS orbiting system to the receiving sites in cronological succession (Mona, Canary, Kenya, Australia) all in the ground track zone, will provide data to perform different signal processing strategies including space correlation techniques with redundancy of data. The selection of Kenya (Malindi) and Australian sites (Brebie Island) , followed the selection of Mona Island with its general location fixed by the condition that are situated in the opposite hemisphere and along the orbital ground track that connects them to Mona and Canary Islands within the same orbital revolution. For TSS 1R the shuttle will launch due east from Kennedy Space Center and overfly points on the ground up to maximun longitudinal extent of 28.5 degree north and south. The australian facility was established by by the University of Queensland, Brisbane , Australia. The Canary Island facility was established by the Istituto Astrofisico de Canarias, Universidad de La Laguna , Tenerife, Spain. Theoretical studies on the exploitation of correlated overflights have been performed to estimate the processing gain of such a spatial signal processing method, in particular for the Mona-Canary islands configuration [6].

In 1993 an other attempt to receive e.m.emissions by a tether was performed at Hawaii island. One OESEE triaxial magnetometer SQUID was located on the ground track of a NASA DELTA II SED system (Small Expendable Deployng System) For the experiment NASA Plasma Motor Generator , PMG '93. In tnat case the tether was long 500 m, and the on board generated current did not work at its maximum as expected. Anyhow some coincidences have been remarked between the tether current

structure and the received signal at the OESEE ground station. Such a result, has allowed some consequential considerations to formulate expectations for the successive tether experiment , namely the TSS 1R. The emf generated by the conducting tether during its flight in the ionosphere is expected to close its circuit through the ionospheric plasma. This ionospheric "closure" is a crucial argument for electrodynamic thether sistems. In Fig. 4 is pictorially represented this closure phenomenon, denominated peculiarly "phantom loop". Such a close current circuit has the associated magnetic moment expressed by the product of the current **I** in amps, by the loop area in square meters. From an estimation of the expexted magnetic moment of the TSS 1R '96 results at least two orders higher than that of PMG'93 tethered system [7]. Estimation of signal levels from reference [14] are shown in Figs. 5 and 6.

4. The Rationale of the Measurements

In principle, we should be in presence of a classical communication channel , with the signal radiated by the moving source , the TSS 1R, nearly "unknown or badly-known". A communication channel capacity depends mandatorily from the a-priori knowledge of both the transmitter and the receiver. In the present case, the "transmitter" is represented by the double interaction of the orbiting system with the ionospheric plasma and the earth magnetic field and is expected to have a not-well-defined multi-variate transfer function. The inherent lack of accurate informations about the transmitter must be, then, compensate, somehow by a rigorous definition of the receiving systemfeatures . The TSS radiating system is supposed to radiate fundamentally by two modes , the *spontaneous mode* (i.e. the generation of Alfven waves [9]) and the *stimulated mode* (the artificial modulation of tether current from board [8]).

Consequently, the OESEE and EMET measuremets can be considered rather a "listening" than a strictly "detection" operation. In the attempt to characterize and process the interferent noise structure, a particular attention have been devoted to a critical interpretation of the recorded data.

For both the radiating modes , the "physical" approach, namely the rigorous definition of the antenna transfer function , and the consequent structure of the radiated (and propagated) signal, is nearly unfeseable. Instead the "empirical" approach, namely the observation, on the Earth surface, and possibly estimate the received radiations emitted by a TSS will be, in any case adopted.

5. Preliminar OESEE Surveys and Site Selection

The receiving instrumentation , consisting in two tri-axial COIL magnetometers and two tri-axial SQUID systems suitable for installation have been assessed and deployed on the selected sites along the ground track of the two electrodynamic tether missions, that have been flown thus far. They where the TSS-1 of August 92 , and the DELTA II-PMG of June 1993. In reality the experimental activity started on October 89 with a test surveys on Tino island (Tirrenian sea) and on the selected site on top of the Tenerife Island of Canaries. Those preliminar experimental measurements where mandatory for the characterization of the local e.m. background natural and man-made noise. The two tri-axial COIL system have been supplied by the curtesy of NPGS , Monterey and NUSC New London (USA). The two SQUID sensors have been specially purchased by DIBE University of Genoa Italy, at the 2G Enterprise in California.In addition to the knowledge of the noise structure the equipment was assessed and calibrated.

The first operation was the recording on the tri-axial components which after the processing showed clearly the Schuman modes, tipical of the natural background noise at ELF bands. 36 hours recording on the three axis where performed in order to verify the stationarity of the noise for a duration of time of the same order of the duration of the forseen TSS-1 mission. The use of a second receiving station at Malindi in Kenya was also considered in order to improve the

probability of detection , on the ground track, of the expexted signals from the orbiting system. Other two receiving stations have been installed in other locations Mona Island (Portorico) and Bribie Island (Australia) under the responsability of the project EMET (SAO and Rice University USA). At this stage was started , in the DIBE, an investigation on the possibility to treat the ELF noise by the Higher Order Spectral (HOS) analysis in order to estimate the non-Gaussian character of the natural E.M. noise at ELF bands verified , otherwise , also from other scientists [Bernstein & other]. This techniques are expected to increase the probability of detection of week signals in presence of non-Gaussian noise.

5.1. The ELF Sensor : Monopole and Dipole Configurations

The same site , at Canary Islands has been used three years later in 1992 for the deploying of the final set up, inclusive of two SQUID magnetometers expecially builded in California. Taking into account the ground track of the orbiting system, the receiving properties of both Monopole and Dipole configurations have been considered. The first considered was the monopole configuration which represents the simplest receiving antenna. The possibility to install a second separate receiving station on the same ground track was also considered . The advantage of the Dipole Configuration, (one sensor located at Canary Islands and an identical one at Mona island-Portorico or at Malindi-Kenya) , considering the dipole distances (5/6 Mm)and the involved ELF bands. An ad hoc investigation has been carried out with the results that a processing gain increase of, at least, 6 dB, and consequently a higher probability of detection of the possible radiated signal, can be expected.

This processing gain is due to both combined properties of the dipole configuration , namely the possibility to carry out "spatial processing" and the "transitting target detection " techniques (Fig. 1 and 2). The non stationarity of the expected signal during any overflight can be exploited by the transiting target detector which was studied by Nicolas [10] . In this case , a pole-zero model of the signal is assumed and the model parameters are estimated in a maximum-likelihood sense. A generalized likelihood ratio test was then used for detection where the estimate is considered as true value of the received wave form . This scheme constitutes the optimum receiver in the Bayes sense. The possible modulation offers the opportunity of placing the energy in a part of the spectrum that will be more favorable to a given detection scheme. The passage over two receiving stations on the ground track (i.e. Mona-Tenerife and or Tenerife-Malindi dipoles) will provide a situation that can be exploited for spatial processing purposes. This property of the expected modulated signal, namely its deterministic generation , and the operative propagation conditions, make feseable the spatial processing. For instance , the spacing between the "elements" of one pair of receiving stations (Mona-Tenerife or Tenerife-Malindi) ranges from 5 to 6 thousands Km, representing half-wave length approximatly at 30 Hz, and a quarter-wave length at 15 Hz. With the adequate boundaries conditions [6] any signal energy in this frequency band can be processed by standard spatial processing techniques. Since the noise can be expected to have a generally horizontal directionality, and when the TSS-1 (and TSS-1R) system is overhead, the directionality of the signal can be expected to be generally downward, this receiving antenna , can be used to reject some of the noise . If, on the other hand , there is energy only at the very low end of the spectrum , the two receiving station configuration , can be used in two other schemes : a **spatial coincidence detector** and **a modified matched filter method.** In the event that the energy, in the 15-30 Hz spectral region, reach the ground with a level impossible to detect, a distinct possibility given by the complex "signal generation" mechanism, it is reasonable to expect that at least some energy in the 0.1 - 3 Hz region should be present. In this event , it is proposed that a set of filters to be used to look for this energy while using the transiting-target-processor independently and then looking for

coincidences between the two receivers . Assuming the correlation time of the noise to be less than the travel time of the TSS-1 between the two receivers , this would provide a gain of approximatly 3 dB over a single receiver . More exactly the probability of detection is given by :

$$P_D = 2\,p_d - (p_d)^2 \qquad (10)$$

Where p_d is the probability of detection for a single receiver. The processing gain , from the above equation converges to 3 dB as p_d approaches zero.

5.2. The Hawaii Islands Experiment : The DELTA II - PMG.

In June 1993, during the PMG experiment , in an appropppriate location on the ground track at Hilo Hawaii, the OESEE receiving site used the single sensor configuration , consisting in a tri-axial SQUID magnetometer and peripheral electronics. The system was deployed to detect on the Earth surface, possible emissions from the DELTA II tether. A pre-processing monitoring system was also employed during the mission . The site showed a particular low level man-made noise. A multiple channels data acquisition system battery operated was used . This attempt to reveal at the closest points of approach the possible radiation emitted by the PMG payload carried into low Earth orbit on a DELTA II second stage . Our attention is devoted mainly to the analysis of the "listened" data at the Hilo OESEE receiving site during the flight (Fig. 8 and Fig. 9), and to compare them with the observed variations of some electrical parameters which represent the power source of the tethered antenna. The most significative appear to be the tether current and voltage profiles during continous Electrometer Reads [7].

During the DELTA II - PMG mission data continous recording where carried out by the OESEE equipment. Although the on line pre processing monitor system , no significative signal related with the overflight or any other phase of the mission has been observed. On the other hand the natural background noise at ELF band was clearly recognized in a short integration time by the presence of the Shuman modes . Such an integration time gave an estimation of the favourable conditions of the local man-made noise. The rational to perform an optimal detector scheme suggest a preliminar exaustive modelling of the radiating system as well as that of the propagation system . Unfortunatly, such a modelling was not carried out taking in to account the constraints and the boundary conditions of the PMG (or TSS). To make reliable prediction about the performances of any telecommunication channel , one essential quantity to be known is the radiating efficiency of the radiating antenna , which in this case is an electrodynamic tether orbiting in the ionosphere. In 1992, [2,3] stated a good compromise for the estimation of the radiated energy by a dipole antenna orbiting in the ionospheric plasma . The estimation of the radiating efficiency from a dipole antenna in the ionosphere has been deduced taking in to account the following conditions :

- $f = 9$ kHz radiating frequency;
- $r = 400$ km , distance from the transmitter in the ionosphere and the receiver on the earth surface;
- $V_o =$ 1 kVolt , the applied voltage to the antenna
- $E_\theta^r = 1.18 \times 10^{-5}$ V/m;
- $I_z(0) = 5.2$ Amp, current at the driving point of the antenna;

One estimation of the radiating efficiency is given by the ratio of the input power to the driving point of the antenna orbiting in the ionospheric plasma and the power received at the receiving site at a distance on the earth surface of about 400 Km. This ratio is

$$\eta = \frac{3.7 \times 10^{-13} \ [W/m^2]}{0.258 \times 10^{-8} \ [W/m^2]} = 1.43 \times 10^{-4} \qquad (11)$$

This estimation of such efficiency is relevant to the conditions above mentioned and is inclusive not only of the effective antenna efficiency but also of the inherent propagation

losses for reach the receiving site. In the case of the PMG must be noted that the driving voltage of the antenna is not a sinusoidal shape but a sort of square vave (wide band signal) with much lower frequency components , from the point of view of wavelength this should compensate the the assumption of 9 kHz.

The radiation efficiency , in the case of a vertical antenna with respect to the ground/air interface is represented by the ratio [2] :

$$\eta = \frac{P[a]}{P[a] + P[g]} \qquad (12)$$

with P[a] = power radiated in air and P[g] = power radiated in the ground. By analogy a linear antenna feeded by an asimmetrical signal can be considered as constituted of two half antennas with different electrical inpedences with consequent degradation of the radiating efficiency.

On the basis of the Schuman theory about the resonance cavity made up of the Earth surface with the Ionosphere, under the cut frequency of 2 KHz, the electromagnetic waves propagate in an almost transversal-magnetic (TEM) way, thus allowing the propagation to huge distances of perturbations due to atmospheric lightnings. The typical values for the first three frequencies are around 8, 14, and 21Hz.

6. Conclusions and Expectations from the TSS-1R (Feb.96)

The activity done by the EMET-OESEE teams since the first approach to the problem and after the various activities like sensors calibrations, configuration studies, noise characterization, signal processing some detection-estimation strategies have been assessed and they are ready to be used for the TSS-1R., for both the configurations : Single Receiving site or Two Receiving sites. The used fundamental techniques to approach this singular problem of Detection - Estimation are indeed those of the electrical communication disciplines, but the problem can be considered from an other, more physical, point of view . Substantially the scientific relevance of a success (the detection on the earth surface of radiations coming from the orbiting system) resides in the fact that the joint observations of the EMET and OESEE projects provide the proof that tether-generated currents have circuit closure that is remote (along the lines of the geomagnetic field) rath er than local. This is an issue of electrodynamics of the ionosphere that has generated substantial debates in this past few years.

REFERENCES

1. M.D.Grossi, E.C.Lorenzini and M.L.Cosmo, 1991, SEDS/Delta-II Electrodynamic Tether Experiments on the Generation and Radiation on E.M. Waves from ULF to VLF , SAO W.P. Rev. No. 3 , June.

2. R.W.P.King,M.Owens and T.Tsun Wu, Lateral Electromagnetic Waves, Springer Verlag 92.

3. P.R.Bannister,J.K.Harrison,C.C.Rupp, R.W.P.King, M.L.Cosmo, E.C.Lorenzini C.J.Dyer and M.D.Grossi, Orbiting Transmitter and Antenna for Spaceborne Communications at ELF/VLF to Submerged Submarines, AGARD Meeting Bruxelles 1992.

4. J.E.Mc Coy and M.D. Grossi , Tether Electrodynamics From The Plasma Motor-Generator (PMG) Mission of June 1993. ESA Itern. Round Table on Tethers in Space ,Nordwijk Sept. 1994.

5. A.Schroer , W. Seboldt, RMHD-Simulation od an Electrodynamic Tethered Satellite System in nthe Ionosphere, ESA International Round Table on Tethers in Space , Noordwijk Sept.1994 .

6. E.J.Sullivan, P.R.Bannister and G.Tacconi, A Possibility of Improving The Detection of Tethered E.M. Radiation by Spatial Processing 3rd International Conference on Tethers In Space, S. Francisco 16 May 1989.

7. D.Chlouber, R.J.Jost, J.E.Mc Coy and T.L. Wilson, Plasma Motor Generator (PMG) Deployment Profile and Tether Current Report. JSC-26697 Report June 1994

8. G. Tacconi, S. Dellepiane, C. Ottonello, S. Pagnan, L. Minna, M.D. Grossi, Campaigns of Ground Listening to the E.M. Emissions Expected From Spaceborne Electrodynamic Tether Systems, *Fourth International Conference on Tethers in Space*, Washington, April 10-14, 1995, pp. 949-963.

9. R.D. Estes, Alfven Waves From an Electrodynamic Tethered Satellite System, Journal of Geophysiscal Research, vol. 93 N.A2 , No6 , Feb. 1988.

10. P. Nicolas, E.J. Sullivan, G. Tacconi and A. Tiano, The TSS ground communication channel: the detection of TSS ULF/ELF/VLF emissions on the earth surface and sea bottom, Proceedings 11th Gretsi, France June 1987.

11. Sanmartin J.R., Martinez-Sanchez M., The radiation impedance of orbiting conductors, J. of Geophysical Research, Vol. 100(2), pp. 1677-1686, February 1, 1995.

12. Noble S.T., Gordon W.E., Estes R.D., Grossi M.D., A ground-based receiving system for the detection of signals from the tethered satellite system, *Fourth International Conference on Tethers in Space*, Washington, April 10-14, 1995.

13. Estes. R., Sanmartin J., Alfven wave signature from constant-current tethers, *Fourth International Conference on Tethers in Space*, Washington, April 10-14, 1995.

14. Pappert . ., Ionospheric e.m. propagation model, Report from NOSC, San Diego, Ca, USA, 1973.

Acknowledgements: The authors are grateful to Dr. M.D. Grossi (Harvard-Smithsonian Center for Astrophysics), L. Minna (Advanced Engineering Technology, Genoa), C. Ottonello (DIBE, Univ. of Genoa), S. Pagnan (IAN-CNR, Genoa), S. Noble and W.E. Gordon (Rice University), M. Sanchez (MIT, Cambridge, Ma), for their smart interactive contribution to extract clear meaningful hints for the future activity in this field.

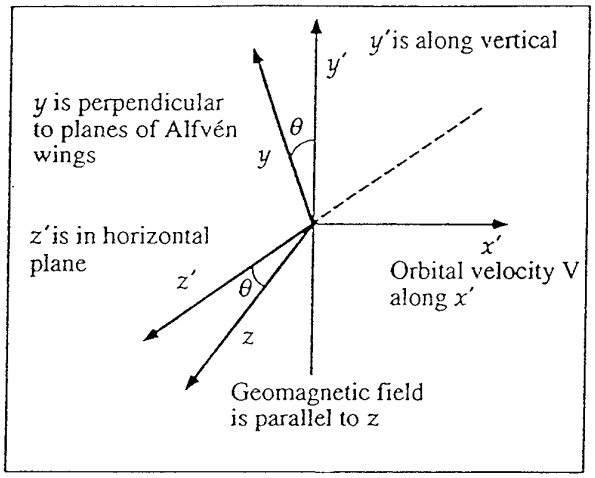


Figure 1: Co-ordinate system used in the analysis from [13].

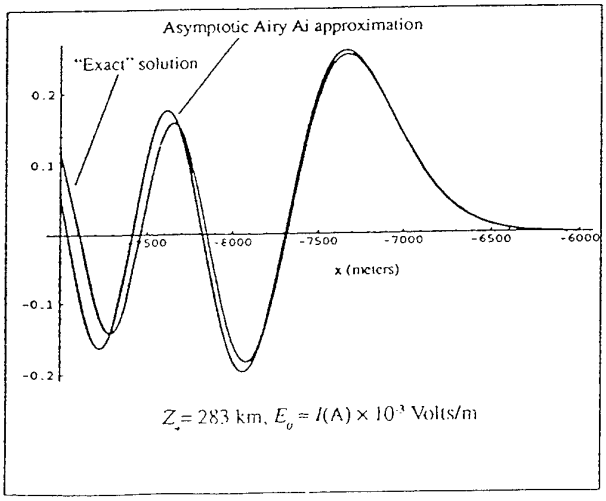


Figure 2: Comparison of asymptotic Alfven wing solution and numerically integrated exact solution of E_y/E_0 from [13].

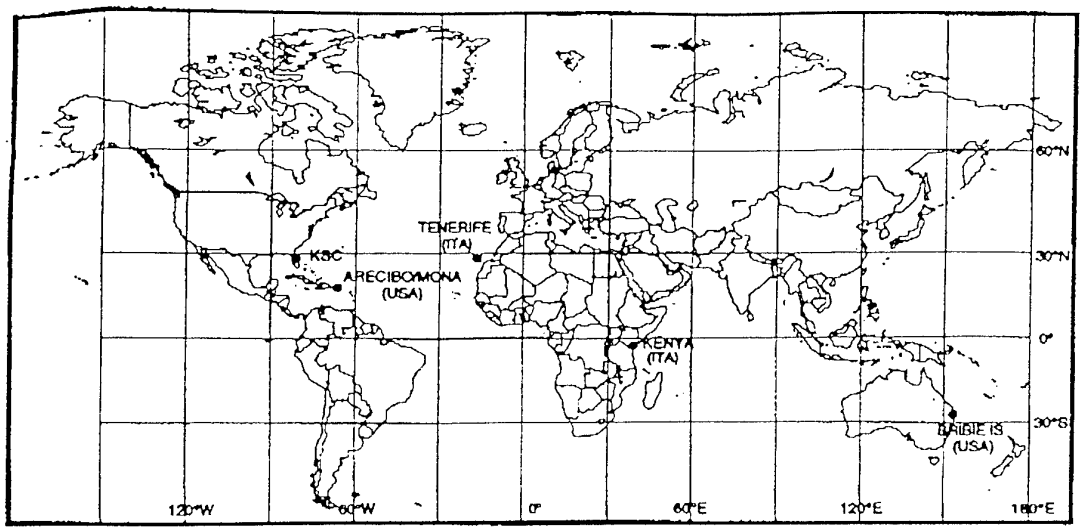
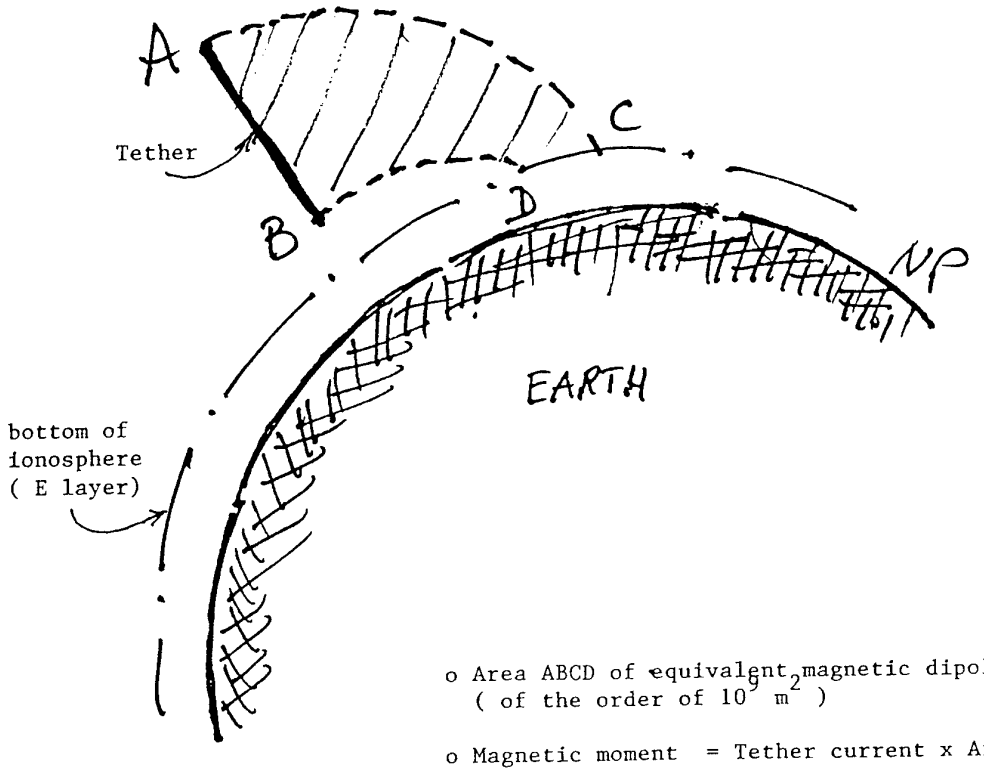


Figure 3: Map of ground stations



- Area ABCD of equivalent magnetic dipole (of the order of 10^9 m^2)
- Magnetic moment = Tether current x Area ABCD
- E.M. field intensities at the Earth surface can be computed according to Pappert, 1973, under the simplifying assumption of a magnetic dipole as a point source (Radio Science, Vol.8 n.6, pp 535-545).

(from PAPPERT 73)

Figure 4: Phantom Loop

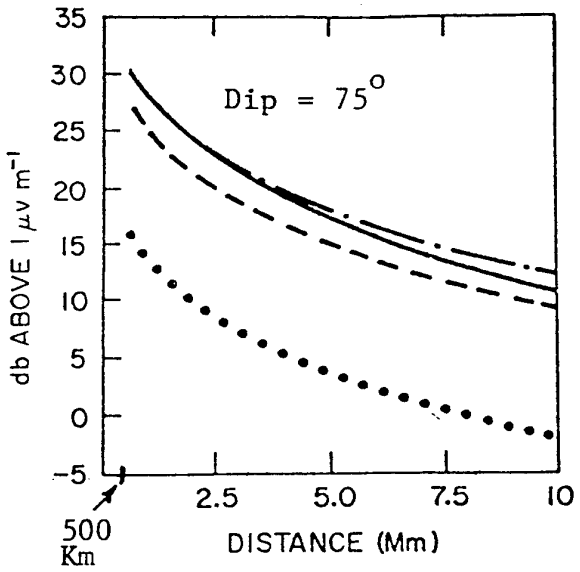


Figure 5: Signals levels versus distance. Signal levels for electric-dipole sources referred to a current moment of 3.18×10^4 amp.-m. Signal levels for magnetic dipole sources referred to a current loop of 2.02×10^{10} vamp.-m^2. The azimuth is 90°, the dip is 75°; and the frequency is 75 Hz. Legend: ---- ground-based electric dipole, end fire, $\sigma=10^{-4}$ mho/m; ●●● vertical electric dipole, 500 Km; _____ horizontal electric dipole, broadside and end fire, 500 Km; -.-.- horizontal magnetic dipole, broadside and end fire, 500 Km. (Pappert 73)

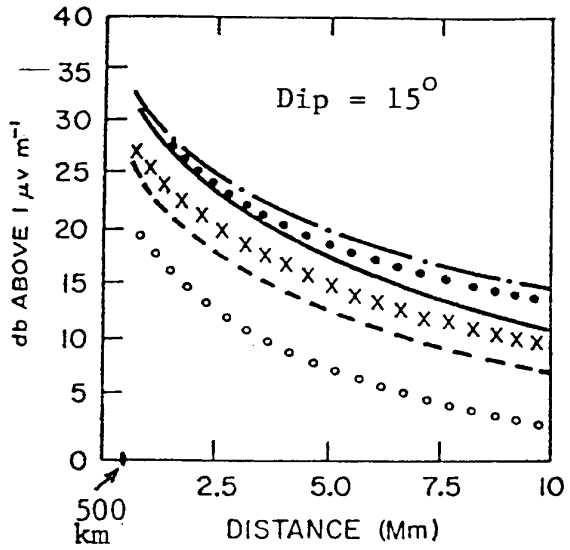


Figure 6: Signals levels versus distance. Signal levels for electric-dipole sources referred to a current moment of 3.18×10^6 amp.-m. Signal levels for magnetic dipole sources referred to a current loop of 2.02×10^{10} vamp.-m^2. The azimuth is 90°, the dip is 15°; and the frequency is 75 Hz. Legend: ---- ground-based electric dipole end fire, $\sigma=10^{-4}$ mho/m; ●●● vertical electric dipole, 500 Km; _____ horizontal electric dipole, end fire, 500 Km; °°°°°° horizontal electric dipole, broadside, 500 Km; × × × horizontal magnetic dipole, endfire 500 Km -.-.-.- horizontal magnetic dipole, broadside, 500 Km.. (Pappert 73)

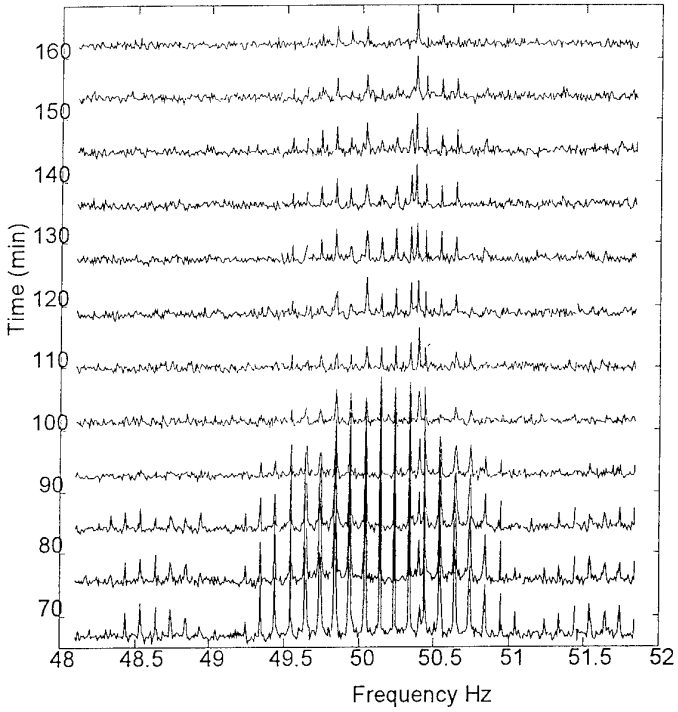


Figure 7: Variation of Power Spectra over Time. PMG data from M.E.T 1:06 to M.E.T. 2:41.

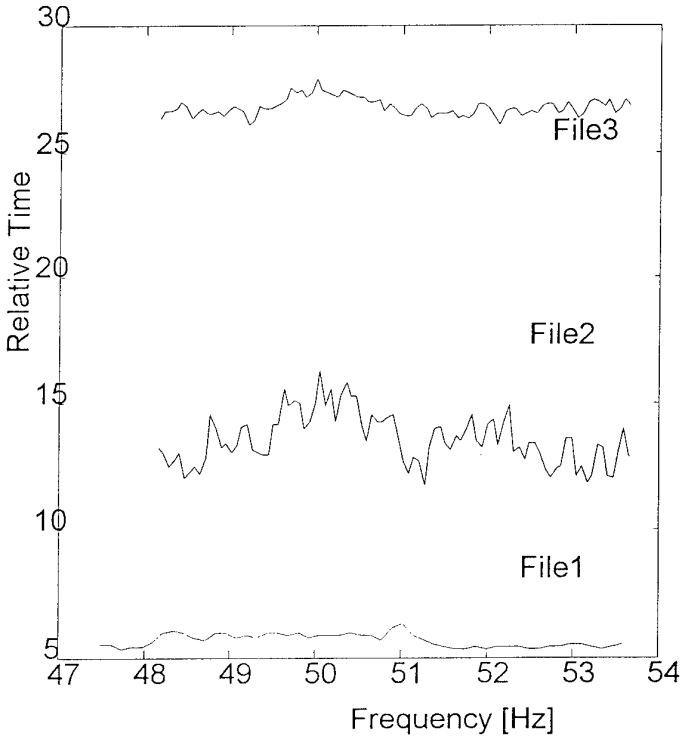


Figure 8: Variation of Power Spectra over Time. PMG noise-only data.

REPORT DOCUMENTATION PAGE

1. Recipient's Reference	2. Originator's Reference	3. Further Reference	4. Security Classification of Document
	AGARD-CP-582	ISBN 92-836-0032-0	UNCLASSIFIED/ UNLIMITED

5. Originator Advisory Group for Aerospace Research and Development
North Atlantic Treaty Organization
7 rue Ancelle, 92200 Neuilly-sur-Seine, France

6. Title
Remote Sensing: A Valuable Source of Information

7. Presented at/sponsored by
the SPP Panel Symposium held in Toulouse, France 22-25 April, 1996

8. Author(s)/Editor(s)	9. Date
Multiple	October 1996
10. Author's/Editor's Address	**11. Pages**
Multiple	444

12. Distribution Statement There are no restrictions on the distribution of this document. Information about the availability of this and other AGARD unclassified publications is given on the back cover.

13. Keywords/Descriptors

Remote sensing	Platforms
Synthetic aperture radar	Artificial satellites
Signal processing	NATO forces
Sensor characteristics	Military operations
Surveillance	Multispectral
Propagation	Data fusion

14. Abstract

This publication reports the unclassified papers presented at a specialists' meeting held by the Sensor and Propagation Panel at its Spring 1996 meeting.

The topics included:

— sensor systems;

— platforms;

— propagation effects;

— signal processing;

— military aspects (requirements, applications);

— military benefit of civil systems;

— (special) civil applications;

— special techniques;

— future developments.

AGARD

NATO OTAN

7 RUE ANCELLE • 92200 NEUILLY-SUR-SEINE

FRANCE

Télécopie (1)47.38.57.99 • Télex 610 176

DIFFUSION DES PUBLICATIONS

AGARD NON CLASSIFIEES

Aucun stock de publications n'a existé à AGARD. A partir de 1993, AGARD détiendra un stock limité des publications associées aux cycles de conférences et cours spéciaux ainsi que les AGARDographies et les rapports des groupes de travail, organisés et publiés à partir de 1993 inclus. Les demandes de renseignements doivent être adressées à AGARD par lettre ou par fax à l'adresse indiquée ci-dessus. *Veuillez ne pas téléphoner.* La diffusion initiale de toutes les publications de l'AGARD est effectuée auprès des pays membres de l'OTAN par l'intermédiaire des centres de distribution nationaux indiqués ci-dessous. Des exemplaires supplémentaires peuvent parfois être obtenus auprès de ces centres (à l'exception des Etats-Unis). Si vous souhaitez recevoir toutes les publications de l'AGARD, ou simplement celles qui concernent certains Panels, vous pouvez demander à être inclu sur la liste d'envoi de l'un de ces centres. Les publications de l'AGARD sont en vente auprès des agences indiquées ci-dessous, sous forme de photocopie ou de microfiche.

CENTRES DE DIFFUSION NATIONAUX

ALLEMAGNE
Fachinformationszentrum Karlsruhe
D-76344 Eggenstein-Leopoldshafen 2

BELGIQUE
Coordonnateur AGARD-VSL
Etat-major de la Force aérienne
Quartier Reine Elisabeth
Rue d'Evere, 1140 Bruxelles

CANADA
Directeur, Services d'information scientifique
Ministère de la Défense nationale
Ottawa, Ontario K1A 0K2

DANEMARK
Danish Defence Research Establishment
Ryvangs Allé 1
P.O. Box 2715
DK-2100 Copenhagen Ø

ESPAGNE
INTA (AGARD Publications)
Carretera de Torrejón a Ajalvir, Pk.4
28850 Torrejón de Ardoz - Madrid

ETATS-UNIS
NASA Goddard Space Flight Center
Code 230
Greenbelt, Maryland 20771

FRANCE
O.N.E.R.A. (Direction)
29, Avenue de la Division Leclerc
92322 Châtillon Cedex

GRECE
Hellenic Air Force
Air War College
Scientific and Technical Library
Dekelia Air Force Base
Dekelia, Athens TGA 1010

ISLANDE
Director of Aviation
c/o Flugrad
Reykjavik

ITALIE
Aeronautica Militare
Ufficio del Delegato Nazionale all'AGARD
Aeroporto Pratica di Mare
00040 Pomezia (Roma)

LUXEMBOURG
Voir Belgique

NORVEGE
Norwegian Defence Research Establishment
Attn: Biblioteket
P.O. Box 25
N-2007 Kjeller

PAYS-BAS
Netherlands Delegation to AGARD
National Aerospace Laboratory NLR
P.O. Box 90502
1006 BM Amsterdam

PORTUGAL
Estado Maior da Força Aérea
SDFA - Centro de Documentaçao
Alfragide
2700 Amadora

ROYAUME-UNI
Defence Research Information Centre
Kentigern House
65 Brown Street
Glasgow G2 8EX

TURQUIE
Millî Savunma Başkanliği (MSB)
ARGE Dairesi Başkanliği (MSB)
06650 Bakanliklar-Ankara

Le centre de distribution national des Etats-Unis ne détient PAS de stocks des publications de l'AGARD.
D'éventuelles demandes de photocopies doivent être formulées directement auprès du NASA Center for AeroSpace Information (CASI) à l'adresse ci-dessous. Toute notification de changement d'adresse doit être fait également auprès de CASI.

AGENCES DE VENTE

NASA Center for AeroSpace Information
(CASI)
800 Elkridge Landing Road
Linthicum Heights, MD 21090-2934
Etats-Unis

The British Library
Document Supply Division
Boston Spa, Wetherby
West Yorkshire LS23 7BQ
Royaume-Uni

Les demandes de microfiches ou de photocopies de documents AGARD (y compris les demandes faites auprès du CASI) doivent comporter la dénomination AGARD, ainsi que le numéro de série d'AGARD (par exemple AGARD-AG-315). Des informations analogues, telles que le titre et la date de publication sont souhaitables. Veuiller noter qu'il y a lieu de spécifier AGARD-R-nnn et AGARD-AR-nnn lors de la commande des rapports AGARD et des rapports consultatifs AGARD respectivement. Des références bibliographiques complètes ainsi que des résumés des publications AGARD figurent dans les journaux suivants:

Scientific and Technical Aerospace Reports (STAR)
publié par la NASA Scientific and Technical
Information Division
NASA Langley Research Center
Hampton, Virginia 23681-0001
Etats-Unis

Government Reports Announcements and Index (GRA&I)
publié par le National Technical Information Service
Springfield
Virginia 22161
Etats-Unis
(accessible également en mode interactif dans la base de données bibliographiques en ligne du NTIS, et sur CD-ROM)

Imprimé par le Groupe Communication Canada
45, boul. Sacré-Cœur, Hull (Québec), Canada K1A 0S7

AGARD

NATO OTAN

7 RUE ANCELLE • 92200 NEUILLY-SUR-SEINE

FRANCE

Telefax (1)47.38.57.99 • Telex 610 176

DISTRIBUTION OF UNCLASSIFIED

AGARD PUBLICATIONS

AGARD holds limited quantities of the publications that accompanied Lecture Series and Special Courses held in 1993 or later, and of AGARDographs and Working Group reports published from 1993 onward. For details, write or send a telefax to the address given above. *Please do not telephone.*

AGARD does not hold stocks of publications that accompanied earlier Lecture Series or Courses or of any other publications. Initial distribution of all AGARD publications is made to NATO nations through the National Distribution Centres listed below. Further copies are sometimes available from these centres (except in the United States). If you have a need to receive all AGARD publications, or just those relating to one or more specific AGARD Panels, they may be willing to include you (or your organisation) on their distribution list. AGARD publications may be purchased from the Sales Agencies listed below, in photocopy or microfiche form.

NATIONAL DISTRIBUTION CENTRES

BELGIUM
Coordonnateur AGARD — VSL
Etat-major de la Force aérienne
Quartier Reine Elisabeth
Rue d'Evere, 1140 Bruxelles

CANADA
Director Scientific Information Services
Dept of National Defence
Ottawa, Ontario K1A 0K2

DENMARK
Danish Defence Research Establishment
Ryvangs Allé 1
P.O. Box 2715
DK-2100 Copenhagen Ø

FRANCE
O.N.E.R.A. (Direction)
29 Avenue de la Division Leclerc
92322 Châtillon Cedex

GERMANY
Fachinformationszentrum Karlsruhe
D-76344 Eggenstein-Leopoldshafen 2

GREECE
Hellenic Air Force
Air War College
Scientific and Technical Library
Dekelia Air Force Base
Dekelia, Athens TGA 1010

ICELAND
Director of Aviation
c/o Flugrad
Reykjavik

ITALY
Aeronautica Militare
Ufficio del Delegato Nazionale all'AGARD
Aeroporto Pratica di Mare
00040 Pomezia (Roma)

LUXEMBOURG
See Belgium

NETHERLANDS
Netherlands Delegation to AGARD
National Aerospace Laboratory, NLR
P.O. Box 90502
1006 BM Amsterdam

NORWAY
Norwegian Defence Research Establishment
Attn: Biblioteket
P.O. Box 25
N-2007 Kjeller

PORTUGAL
Estado Maior da Força Aérea
SDFA - Centro de Documentaçao
Alfragide
2700 Amadora

SPAIN
INTA (AGARD Publications)
Carretera de Torrejón a Ajalvir, Pk.4
28850 Torrejón de Ardoz - Madrid

TURKEY
Millî Savunma Başkanliği (MSB)
ARGE Dairesi Başkanliği (MSB)
06650 Bakanliklar-Ankara

UNITED KINGDOM
Defence Research Information Centre
Kentigern House
65 Brown Street
Glasgow G2 8EX

UNITED STATES
NASA Goddard Space Flight Center
Code 230
Greenbelt, Maryland 20771

The United States National Distribution Centre does NOT hold stocks of AGARD publications.
Applications for copies should be made direct to the NASA Center for AeroSpace Information (CASI) at the address below.
Change of address requests should also go to CASI.

SALES AGENCIES

NASA Center for AeroSpace Information
(CASI)
800 Elkridge Landing Road
Linthicum Heights, MD 21090-2934
United States

The British Library
Document Supply Centre
Boston Spa, Wetherby
West Yorkshire LS23 7BQ
United Kingdom

Requests for microfiches or photocopies of AGARD documents (including requests to CASI) should include the word 'AGARD' and the AGARD serial number (for example AGARD-AG-315). Collateral information such as title and publication date is desirable. Note that AGARD Reports and Advisory Reports should be specified as AGARD-R-nnn and AGARD-AR-nnn, respectively. Full bibliographical references and abstracts of AGARD publications are given in the following journals:

Scientific and Technical Aerospace Reports (STAR)
published by NASA Scientific and Technical
Information Division
NASA Langley Research Center
Hampton, Virginia 23681-0001
United States

Government Reports Announcements and Index (GRA&I)
published by the National Technical Information Service
Springfield
Virginia 22161
United States
(also available online in the NTIS Bibliographic
Database or on CD-ROM)



Printed by Canada Communication Group
45 Sacré-Cœur Blvd., Hull (Québec), Canada K1A 0S7

ISBN 92-836-0032-0