

Characterization of the Earth's Surface and Atmosphere
from Multispectral and Hyperspectral Thermal Imagery

by
Erich D. Hernandez-Baquero

Submitted to the
Chester F. Carlson Center for Imaging Science
in partial fulfillment of the requirements
for the Doctor of Philosophy Degree
at the Rochester Institute of Technology

Abstract

The goal of this research was to develop a new approach to solve the inverse problem of thermal remote sensing of the Earth. This problem falls under a large class of inverse problems that are ill-conditioned because there are many more unknowns than observations. The approach is based on a multivariate analysis technique known as Canonical Correlation Analysis (CCA). By collecting two ensembles of observations, it is possible to find the latent dimensionality where the data are maximally correlated. This produces a reduced and orthogonal space where the problem is not ill-conditioned. In this research, CCA was used to extract atmospheric physical parameters such as temperature and water vapor profiles from multispectral and hyperspectral thermal imagery. CCA was also used to infer atmospheric optical properties such as spectral transmission, upwelled radiance, and downwelled radiance. These properties were used to compensate images for atmospheric effects and retrieve surface temperature and emissivity. Results obtained from MODTRAN simulations, the MODerate resolution Imaging Spectrometer (MODIS) Airborne Sensor (MAS), and the MODIS and Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) (MASTER) airborne sensor show that it is feasible to retrieve land surface temperature and emissivity with 1.0 °K and 0.01 accuracies, respectively.

Characterization of the Earth's Surface and Atmosphere
from Multispectral and Hyperspectral Thermal Imagery

by

Erich D. Hernandez-Baquero

B.S., Physics, U.S. Air Force Academy, 1994

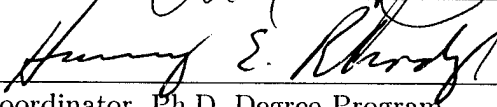
A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy
in the Chester F. Carlson Center for Imaging Science
Rochester Institute of Technology

2000

Signature of the Author



Accepted by


Coordinator, Ph.D. Degree Program

6/9/2000

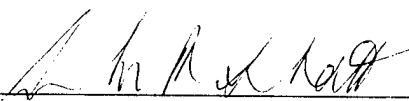
Date

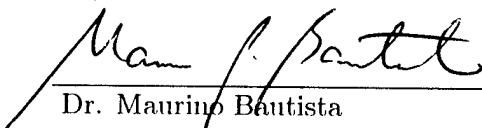
CENTER FOR IMAGING SCIENCE
ROCHESTER INSTITUTE OF TECHNOLOGY
ROCHESTER, NEW YORK

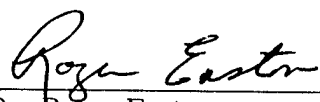
CERTIFICATE OF APPROVAL

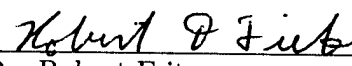
Ph.D. DEGREE DISSERTATION

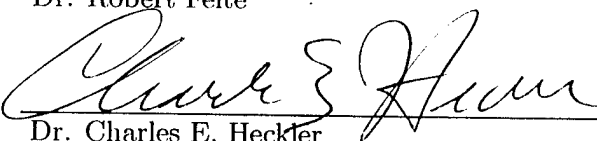
The Ph.D. Degree Dissertation of Erich D. Hernandez-Baquero
has been examined and approved by the
dissertation committee as satisfactory for the
dissertation required for the
Ph.D. degree in Imaging Science

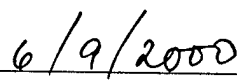

Dr. John R. Schott, Dissertation Advisor


Dr. Maurino Bautista


Dr. Roger Easton


Dr. Robert Feite


Dr. Charles E. Heckler


Date

DISSERTATION RELEASE PERMISSION
ROCHESTER INSTITUTE OF TECHNOLOGY
CENTER FOR IMAGING SCIENCE

Title of Dissertation:

**Characterization of the Earth's Surface and Atmosphere
from Multispectral and Hyperspectral Thermal Imagery**

I, Erich D. Hernandez-Baquero, hereby grant permission to Wallace Memorial Library of R.I.T. to reproduce my thesis in whole or in part. Any reproduction will not be for commercial use or profit.

Signature



6/9/2000

Date

Characterization of the Earth's Surface and Atmosphere
from Multispectral and Hyperspectral Thermal Imagery

by
Erich D. Hernandez-Baquero

Submitted to the
Chester F. Carlson Center for Imaging Science
in partial fulfillment of the requirements
for the Doctor of Philosophy Degree
at the Rochester Institute of Technology

Abstract

The goal of this research was to develop a new approach to solve the inverse problem of thermal remote sensing of the Earth. This problem falls under a large class of inverse problems that are ill-conditioned because there are many more unknowns than observations. The approach is based on a multivariate analysis technique known as Canonical Correlation Analysis (CCA). By collecting two ensembles of observations, it is possible to find the latent dimensionality where the data are maximally correlated. This produces a reduced and orthogonal space where the problem is not ill-conditioned. In this research, CCA was used to extract atmospheric physical parameters such as temperature and water vapor profiles from multispectral and hyperspectral thermal imagery. CCA was also used to infer atmospheric optical properties such as spectral transmission, upwelled radiance, and downwelled radiance. These properties were used to compensate images for atmospheric effects and retrieve surface temperature and emissivity. Results obtained from MODTRAN simulations, the MODerate resolution Imaging Spectrometer (MODIS) Airborne Sensor (MAS), and the MODIS and Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) (MASTER) airborne sensor show that it is feasible to retrieve land surface temperature and emissivity with 1.0 °K and 0.01 accuracies, respectively.

Acknowledgements

I am indebted to many people for their guidance, support, and kindness, without which this work would not have been possible. I am particularly grateful for my advisor, Dr. John Schott who inspired me to pursue this research topic and helped me transition to the Ph.D program. Also instrumental in this transition was Lt. Col. Ronald Fairbanks, whose leadership I admire. I am privileged to have a professional and personal relationship with him. I also thank Col. Patneaud and the U.S. Air Force for giving me the opportunity to stay at RIT and further my education.

I appreciate the friendship and encouragement of Maj. Charles Daly, who helped me “get my bearings” when I first arrived to RIT. His legendary sense of humor and intellectual prowess made my experience here that much more enjoyable. My thanks also go to all of the students and staff at the Center for Imaging Science who, at one point or another, lend a helping hand or insightful comment.

My best regards go to the members of my committee for their guidance and review. In particular, I thank Dr. Charles Heckler for his eagerness to help and for always finding the time to share his expertise in multivariate statistics. His dynamic personality and vast knowledge helped me tremendously as I endeavored to understand the material.

I offer my deepest appreciation to the members of the RIT faculty who motivated me through my studies. Dr. Harvey Rhody was always willing to meet with me and offered wisdom, guidance, and support when I most needed it. The dynamic teaching style and candid personality of Dr. Dan Lawrence motivated me to learn more about multivariate statistics. Indirectly, his instruction led to the chosen approach in this research.

Finally, I would like to acknowledge the scientists and engineers who shared many of the data sets used in this work. Many thanks to Simon Hook and Frank Palluconi from the Jet Propulsion Laboratory for supplying the Railroad Valley MASTER images and all other associated ground truth data. I also thank Dan Zhou from NASA LARC and Liam Gumley from SSEC for supplying the NAST-I and SSEC radiosonde data, respectively.

This work is dedicated to...

*...my parents, Raul and Niurka,
for always believing in me,
and teaching me to do the same.*

*...my sons, Kurtis and Nathan,
for bringing so much pride
and joy into my life.*

*...and my wife, Kimberly,
for being my soulmate
and always loving me.*

Contents

List of Figures	x
List of Tables	xiii
1 Introduction	1
1.1 Notation	10
1.2 Scope and Limitations	11
1.2.1 Atmospheric Characterization	11
1.2.2 Temperature and Emissivity Estimation	12
2 Background	13
2.1 Atmospheric Radiation and Propagation	14
2.1.1 The Atmosphere	14
2.1.2 The Infrared Spectrum	16
2.1.3 The “Forward” Model	18
2.1.4 Model-Matching	32
2.1.5 Atmospheric Sounding Techniques	36
2.1.6 The In-Scene Atmospheric Compensation Method	59
2.2 Temperature and Emissivity	65
2.2.1 Split-Window Algorithms	66
2.2.2 Alpha-Derived Emissivities	68
2.2.3 Temperature and Emissivity Separation (TES) Algorithm	72
2.2.4 Other Methods	75
2.3 Summary and Discussion	76
3 Approach	81
3.1 Theoretical Basis and Development	83
3.1.1 Ordinary Least Squares and Principal Components	84
3.1.2 Canonical Correlation Analysis	86
3.1.3 Derivation of Canonical Correlation Analysis and Regression	90
3.1.4 Interpretation and Observations	97
3.2 Implementation	99
3.2.1 CCR Inverse Model	99

3.2.2	Temperature and Emissivity Separation	103
3.3	Test and Validation	116
3.3.1	Simulations	117
3.3.2	MODIS Airborne Simulator (MAS)	118
3.3.3	MASTER	119
3.3.4	SEBASS	119
3.3.5	Experiments	120
3.3.6	Comparative Studies	128
3.3.7	Validation of Linear Model	130
4	Results	131
4.1	Experiment #1	131
4.2	Experiment #2	142
4.3	Experiment #3	145
4.3.1	Number of Bands Determination	145
4.3.2	Band Selection	149
4.3.3	Testing and Validation of Band Configurations	155
4.4	Experiment #4	156
4.4.1	Atmospheric Parameter Retrievals	158
4.4.2	Surface Temperature and Emissivity Retrievals	164
4.4.3	Other Findings	170
4.5	Comparison of Multivariate Regression Methods	173
4.6	Comparisons to ISAC	176
4.7	Appropriateness of Linear Model	182
4.7.1	Canonical Variables and Correlations	183
4.7.2	Analysis of Residuals	187
4.7.3	The CCA Regression Model	188
5	Conclusions and Recommendations	192
5.1	Summary	192
5.2	Extensions of the CCA Approach	198
5.3	Outlook on the future	203
A	Water Vapor Units and Conversions	205
B	Singular Value Decomposition	210
C	Simultaneous Retrieval of Temperature and Concentration Profiles	213
D	Other Multivariate Regression Methods	220
D.1	Partial Least Squares	220
D.2	Maximum Redundancy	226
D.3	Implementation	229

E	ISAC Implementation and Validation	230
E.1	Least-Squares Maximum-Hit Method	230
E.2	Kolmogorov-Smirnov Method	231
E.2.1	Kolmogorov-Smirnov Two-Sided Statistic	232
E.2.2	Kolmogorov-Smirnov Probability Distribution	233
E.2.3	Kolmogorov-Smirnov Regression and ISAC	234
E.3	Normalized Regression	238
E.4	A Qualitative Validation	241
	Works Cited	247

List of Figures

1.1	Earth's Radiation Budget	2
1.2	Interaction between the Earth's surface and atmosphere	4
1.3	Illustration of a hypercube.	5
1.4	Schematic of infrared forward model.	6
2.1	Vertical temperature profile for the 1976 U.S. Standard Atmosphere.	15
2.2	Infrared Spectrum	16
2.3	Carbon Dioxide Seasonal and Yearly Variation	18
2.4	Radiation Propagation Paths	19
2.5	Radiation originating from solar and terrestrial temperatures	20
2.6	Integration path for a layered inhomogeneous atmosphere	24
2.7	Lorentz line shape for an absorption band.	27
2.8	Effects of spectral resolution on band transmission functions.	30
2.9	Schematic of downhill simplex regression method	34
2.10	MODTRAN vertical profiles.	37
2.11	Planck curves superimposed on atmospheric radiation.	39
2.12	Weighting function arising from pressure broadening phenomenology	44
2.13	Regularization γ effects.	50
2.14	Mapping of state space to measurement space.	52
2.15	Scatter plot for the determination of blackbody pixels	63
2.16	Plot of alpha-residuals and emissivity.	70
2.17	Bias due to Wien's approximation.	71
2.18	Empirical relationship between MMD and ϵ_{min}	74
3.1	Path model for ordinary least-squares.	85
3.2	Path model for principal components regression.	87
3.3	Path diagram for CCR	88
3.4	Plot of sum of squared error.	100
3.5	CCA implementation block diagram	101
3.6	Geographic coverage of radiosonde measurements.	103
3.7	Plot of data and regression fit of the original empirical relationship.	106
3.8	Residual plots for the original empirical relationship.	107
3.9	Spectral emissivity of extreme outliers.	108

3.10	Residual plots for new model	111
3.11	Lag-1 plot showing no correlation in the residuals.	111
3.12	Two-dimensional confidence region of nonlinear parameters.	114
3.13	Schematic of test and validation approach.	116
3.14	Spectral intervals for the SEBASS LWIR band.	120
3.15	MAS infrared image of Death Valley	124
3.16	MASTER band spectral response.	128
3.17	MASTER IR images	129
4.1	Schematic of design of Experiment # 1	132
4.2	CCR Results.	133
4.3	CCR Modes.	134
4.4	CCR Results for Temperature.	135
4.5	Observed radiances for two different surface temperatures	136
4.6	Scaling factor	138
4.7	Comparison of scaled and unscaled observed radiances	139
4.8	Biased and unscaled CCR Results.	140
4.9	Biased and Scaled CCR Results.	141
4.10	Schematic of design of Experiment # 2	143
4.11	Cross-validation results for MAS retrievals	145
4.12	Schematic of design of Experiment # 3	146
4.13	MWIR FSL error profiles	147
4.14	MWIR NAST-I error profiles	147
4.15	FSL MWIR canonical modes	150
4.16	NAST-I MWIR canonical modes	152
4.17	Canonical weights and loadings for MWIR emissive region	154
4.18	RMS error profiles for different band configurations.	156
4.19	Schematic of design of Experiment # 4	157
4.20	NAST-I canonical weights and loadings for MASTER	161
4.21	NAST-I canonical weights and loadings for high-resolution case	162
4.22	RMS Errors and retrievals of atmospheric spectra	163
4.23	Simulated MASTER emissivity retrievals	167
4.24	Emissivity retrievals from MASTER imagery	168
4.25	Simulated SEBASS Emissivity retrievals	169
4.26	Comparison of water vapor profile retrievals	171
4.27	Transmission spectra retrievals for SEBASS simulations	172
4.28	Comparison of transmission spectra RMS errors	175
4.29	Comparison of atmospheric spectra residuals	176
4.30	CCR and ISAC surface radiance retrievals over Lake Mead	179
4.31	ISAC _{NR} emissivity retrievals for Railroad Valley playa.	180
4.32	Surface-leaving radiance ISAC retrieval for Railroad Valley playa.	181
4.33	ISAC and CCR transmission spectra retrievals over Railroad Valley playa.	181
4.34	Matrix plot of canonical variables for MASTER	183
4.35	Matrix plot of canonical variables for SEBASS	184

4.36	Histogram and normal probability plots for canonical variables	186
4.37	Residual plots for T_s and CWV with repeats	188
4.38	Residual plots for T_s and CWV without repeats	189
4.39	Example regression data from Draper	190
5.1	Temperature maps and brightness temperature image of Lake Mead	197
5.2	Histogram of Lake Mead temperature maps	199
5.3	Diagram of the cascaded CCA approach	200
A.1	Water vapor saturation pressure	207
C.1	Taylor approximation of the Planck function.	216
D.1	Path model for partial least squares.	221
E.1	Noisy data and model line plot	235
E.2	Histogram and cumulative distribution	238
E.3	Scatter and residual plots for Band 46	240
E.4	Histogram and normal plots for Band 46	240
E.5	Temperature and emissivity maps of calibration panels	242
E.6	ISAC _{KS} transmission and upwelled radiance results	243
E.7	ISAC _{KS} and TES emissivity retrievals for panel E1	244
E.8	ISAC _{NR} and ISAC _{MH} transmission and upwelled radiance results	244
E.9	ISAC _{NR} and ISAC _{MH} TES emissivity retrievals for panel E1	245
E.10	Comparison of transmission spectra for ISAC NR and KS	246

List of Tables

2.1	Summary Table of Methods	78
3.1	JHU spectral library reference table.	104
3.2	Regression results using $g = 0.737$ and no repeats.	106
3.3	ANOVA results for the new fitted model.	109
3.4	Description of radiosonde used for CCR	123
3.5	MAS thermal bands for Death Valley collect.	125
3.6	MASTER thermal band centers	127
4.1	Summary results for experiment # 2	144
4.2	Summary Band-Selection Results	155
4.3	Atmospheric database description and statistics for Experiment #4	157
4.4	Temperature profiles and CWV retrievals for Lake Mead	158
4.5	Temperature profiles and CWV retrievals for Railroad/White River Valley	158
4.6	Errors in retrieved surface temperature for blackbody targets.	164
4.7	Errors in retrieved surface temperature for variable emissivity targets.	166
4.8	Comparison of multivariate methods for MASTER	173
4.9	Comparison of multivariate methods for SEBASS	173
4.11	Comparison of multivariate methods for MWIR medium-resolution	174
4.12	Comparison of multivariate methods for MWIR 5-band resolution	174
4.10	Comparison of multivariate methods for high resolution	174
4.13	Comparison of CCR and ISAC surface temperature retrievals with TES	177
4.14	Comparison of CCR and ISAC surface temperature retrievals with NEM	179
4.15	Canonical correlations and corresponding Kolmogorov-Smirnov statistics.	187

Chapter 1

Introduction

Man must rise above the Earth—to the top of the atmosphere and beyond—for only thus will he fully understand the world in which he lives.

Socrates (500 B.C.)

One of the greatest scientific endeavors of our time is understanding the Earth's ecosystem at a global scale. Recent global phenomena, such as El Niño, ozone depletion, and the much debated global warming, demonstrate the powerful effect climate can have on people and the economy. For example, the record-setting 1997-1998 El Niño caused the deaths of about 2100 people and resulted in more than 33 billion dollars (U.S.) worth of property damage (Suplee 1999). Accurate forecasting of these major events, and all associated weather phenomena, could potentially save people's lives. To this end, it is necessary to characterize the thermal radiation processes that govern the heat exchange between the surface and the atmosphere, which play a large role in climate.

The Earth's radiation budget illustrated in Figure 1.1 provides a framework for analyzing the distribution of energy and heat exchange that influences the environment. The principal source of energy is the Sun. The thermodynamic state of the atmosphere and the Earth's surface plays a key role in how the energy is distributed. The incoming energy is

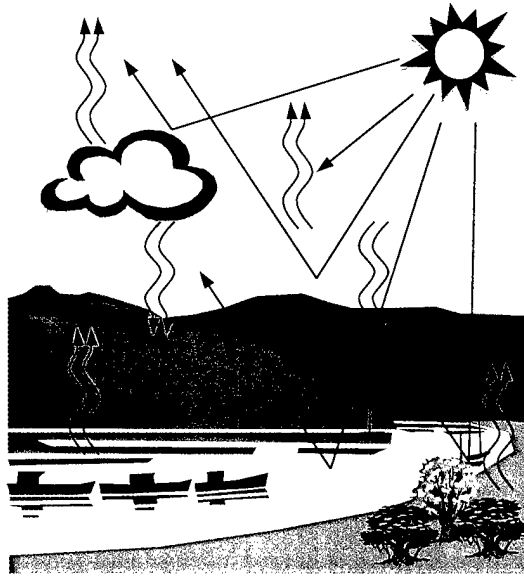


Figure 1.1: The Earth's radiation budget is the categorization of energy paths in the atmosphere. The blue arrows are solar radiation components. The red arrows are emission components due to temperature and emissivity.

either reflected or absorbed by the atmosphere and the surface. On average, about 70% of the solar energy is absorbed by the Earth (Seinfeld and Pandis 1998). This absorbed energy must be balanced by emission to maintain equilibrium. The distribution of energy and the coupling between the atmosphere and the Earth's surface directly influence the global climate.

Many years of research and atmospheric observations have led to the creation of General Circulation Models (or Global Climate Models-GCM). These models yield weather and climate predictions based on past observations and the current state of the atmosphere. The state of the atmosphere can be represented with several physical parameters such as vertical profiles of temperature and concentration of constituents. Because of the strong coupling between the atmosphere and the Earth's surface, surface temperature is also an important parameter to climate models. The accuracy of the models is driven by two major factors: 1) the resolution of the measurements (i.e., the temporal and spatial interval

between observations); and 2) the accuracy of the inputs. Remote sensing of the Earth from aircraft or satellites provides the synoptic view needed to satisfy the measurement resolution requirements. The accuracy of the model inputs depends largely on the quality of the system hardware and processing algorithms deployed with these remote sensing platforms. The goal for accuracy set forth by the scientific community is to retrieve land surface temperature (LST) within 1 °K and sea surface temperature (SST) within 0.3 °K (Wan 1999; Wan and Li 1997). To address this need, technological advances have resulted in imaging systems that capture an unprecedented amount of data with much higher fidelity than ever before. Consequently, specialized processing algorithms are needed to exploit the information content potentially contained in these large data sets.

Another major scientific effort focuses on remote sensing of the Earth's surface for geological applications. Of great interest is the monitoring of volcanic activity, pollution, vegetation health, urban and agricultural development, etc. For these applications, accurate thermography of the Earth's surface is also required. An additional requirement is the accurate estimation of the surface spectral emissivity. The emissivity serves as a *signature* that may be used to identify materials. Spectral classification algorithms use this information to generate thematic maps from remotely sensed images. Frequently, this involves the allocation of pixels to a particular material class (e.g., deciduous forest).

Yet another application where accurate thermography and emissivity retrievals are important is in tactical surveillance. Future military conflicts will rely heavily on intelligence gathered via remote sensing measurements. Knowing the temperature of targets can provide information about its current thermodynamic state. The emissivity provides a method for accurate identification. Both measurements can be used to identify targets that need to be engaged or to provide battle damage assessments.

All of these applications benefit from observations made in the thermal infrared region of the electromagnetic spectrum. Passive acquisition of thermal radiation provides daytime and nighttime monitoring capability without the need for external illumination sources.

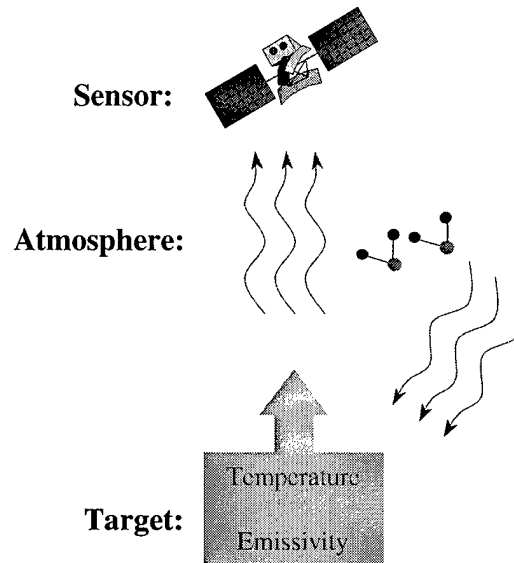


Figure 1.2: The radiation reaching the sensor is the result of a complex interaction between the Earth's surface and the atmosphere.

This increases the number of observations available for a given location on the Earth. Unfortunately, there is a strong coupling between the thermal emission of the surface and the atmosphere. In addition, the amount of atmospheric radiation observed by a sensor strongly depends on the emissivity of the Earth's surface. Finally, the surface emission is the result of a nonlinear interaction between temperature and emissivity. Figure 1.2 illustrates this problem. To satisfy the requirements of many applications, the effects of the atmosphere *and* the surface must be estimated and separated. Over the years, this has proven to be very difficult, particularly over land surfaces where temperature and emissivity vary and there is considerable spatial heterogeneity (Prata et al. 1995).

The advent of thermal *multispectral* and *hyperspectral* technology provides reason for hope. Modern multispectral sensors typically provide relatively high spatial resolution imagery with moderate spectral resolution (i.e., a few broad spectral bands). Hyperspectral sensors acquire images with high spatial *and* spectral resolution. The resulting image is often called a "hypercube" as illustrated in Figure 1.3. The sensor is responsive to incom-

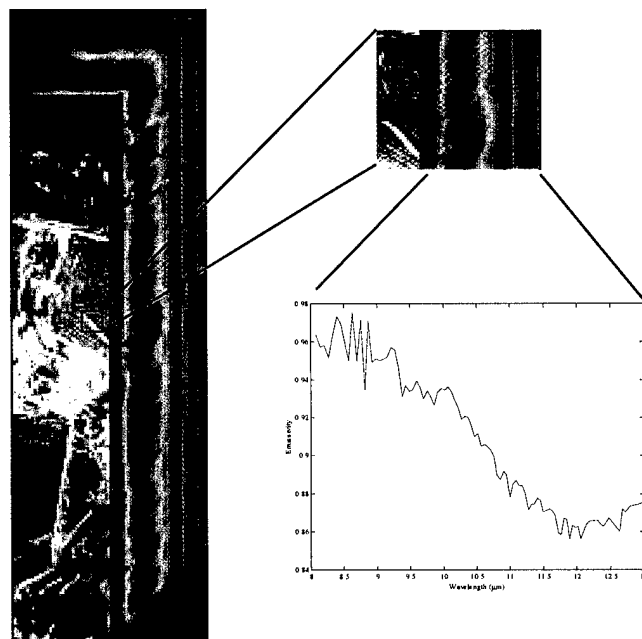


Figure 1.3: Illustration of a hypercube.

ing radiation over discrete wavelength bands. Thus, an image is generated for each sensor band and stacked over one another to form the hypercube. Each pixel can then be used to obtain the observed spectral radiation at a given location in the image. The large amount of spectral and spatial data contained in these images may provide enough information to characterize the atmosphere adequately, thus allowing compensation for its effects and determination of surface emission components. The efficient manipulation of these data, however, presents its own challenges. In addition, the nature of radiative transfer makes it impossible to exactly determine the parameters of interest based on the observational data alone. Each spectral and spatial measurement is the result of a complex interaction of many parameters.

One approach to solve this problem is to use a *forward model* of radiative transfer. Figure 1.4 is a schematic showing the flow of input and output parameters from a generic physical model. The observed radiance is shown by itself because it is the only parameter

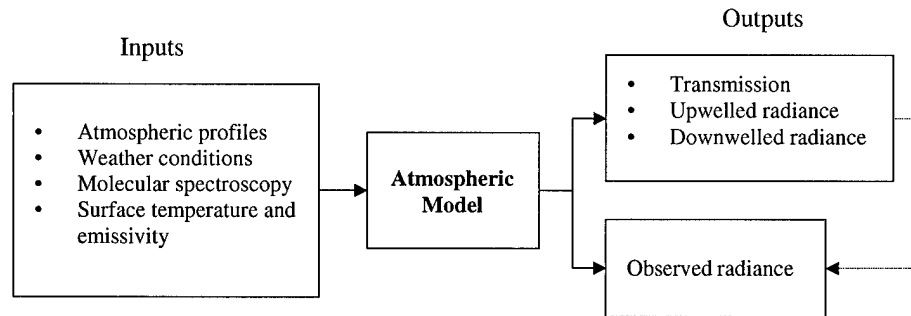


Figure 1.4: Schematic of infrared forward model.

measured by a remote sensor. The atmospheric effects may be characterized by assuming certain input parameters are known. These inputs may be obtained from ancillary data such as radiosonde measurements. Radiosondes are balloon-borne sensors launched at approximately the same time and place as the image acquisition. The radiosondes measure temperature, pressure, wind speed, and humidity during the balloon's ascent. This results in vertical profiles that can be used to model the optical properties of the intervening atmosphere between the sensor and the ground. Unfortunately, radiosondes are susceptible to drift during their ascent and may not accurately represent the actual composition of the atmosphere for a given column of air. Furthermore, the logistics of successfully launching a coincident radiosonde for every remote sensing acquisition over the planet is impractical. One approach that uses the forward model without ancillary data is to dynamically change the input parameters until the difference between the model output and the observed radiance is minimized based on some criterion. This "model matching" approach requires a good parameterization of the inputs that must be developed *a priori* for each specific application. Also, unless properly constrained, this approach may not converge or may lead to unrealistic atmospheric parameters.

Another approach is to reverse the arrows of the forward model diagram. This *inverse* problem is said to be *ill-conditioned* or *ill-posed* because there are more unknowns than observations and, therefore, no exact solution. The best that we can do is to obtain the best

or most likely solution. Several methods for doing this have been developed. Atmospheric *sounding* techniques retrieve vertical profiles of temperature and constituent concentrations from multi-angle or spectroscopic observations (Houghton 1984; King 1956; Kaplan 1959). The profiles are inferred directly from the sensor radiance. Unfortunately, the sounding approach usually needs many narrow spectral bands along an absorption feature, making it difficult to observe the Earth's surface. Also, it is increasingly difficult to maintain an adequate signal-to-noise ratio (SNR) as the spectral resolution increases because fewer photons are detected in each narrow band. To compensate, the detector elements must be large resulting in very low spatial resolution imagery. Finally, numerical algorithms for *physical* sounding often need a good initial estimate of the solution. Other approaches exist and are discussed in Chapter 2.

Even if the atmospheric effects are properly compensated, the problem of temperature and emissivity separation still remains. The problem is complicated because the surface radiation is the product of the temperature and the emissivity. For a given radiance measurement, an infinite number of temperature and emissivity solutions exist. Therefore, the accuracy of the temperature measurement is directly related to the accuracy of the estimate of the target emissivity. Research has shown that the emissivity must be known with an accuracy of 0.02 or less to obtain adequate estimates of temperature (Wan and Li 1997). When measuring temperature over water, the task is simplified because the infrared emissivity of water is well known and spectrally flat. Thus, it is possible to measure the temperature accurately with a radiometer that has a limited number of broad bands. Because of this, operational systems measuring ocean temperature have been successful for years. However, when applying these algorithms to land the result is not the same. Prata et al. (1995), present an excellent review of the complications that arise from measuring LST as opposed to sea surface temperatures (SST). The most popular algorithm for temperature estimation is the split-window technique. Prata et al. demonstrated that although several variations of the technique are feasible for LST estimation, they are seriously limited due to

lack of knowledge of the atmosphere and surface emissivity. The latter confounds the former because the emissivity of land objects vary considerably with wavelength. In addition, the emissivities in a land image are spatially heterogeneous. That means that unless the emissivity effects are considered, the atmosphere will appear to be changing spatially and may be represented inaccurately by a less sophisticated algorithm. Finally, the emissivity varies with the angle of incidence formed by the sensor-target geometry. Unless *a priori* knowledge of the target exists, the emissivity of a particular pixel in a remotely sensed image is unknown. Empirical solutions based on laboratory data have been implemented in order to take the emissivity effect into account. These results, however, are coarse in the spectral sense and rely on often-violated assumptions.

As mentioned previously, many applications require accurate estimates of the emissivity for target identification. Classification algorithms use a library of spectral curves that correspond to each material of interest. The measured emissivity is compared to the curves in the library and the curve that leads to the best match is selected. The problem is complicated by the fact that the measurement is limited by the sensor's spectral *and* spatial resolution. Low spectral resolution limits the absorption features that can be observed by the sensor. Thus, a narrow feature that uniquely distinguishes two similar materials may not be detected. If the spatial resolution is limited, each pixel will consist of a mix of "pure" surface components called *end-members* (Sabol et al. 1992). Thus, it would be necessary to perform some type of "unmixing." Clearly, the amount of mixing can be reduced when high spatial resolution is available. If the target size is bigger than the ground sampled distance (GSD) then the pixel *effective* emissivity is sufficient. The requirement for accurate *absolute* emissivity values is less stringent for classification algorithms. In general, classification algorithms depend only on the relative spectral features present in the retrieved emissivity. However, they do require that the atmospheric features be removed from the derived emissivities.

The goal of this research is to develop a technique that solves the inverse problem by finding the best or most probable solution. This is done by considering both atmospheric and surface effects on the radiance reaching the sensor. Using a multivariate analysis technique known as Canonical Correlation Analysis (CCA), it is possible to develop a *unified* approach that optimally predicts surface and atmospheric parameters. *Unified* is emphasized because the approach accounts for the joint effects of the surface and the atmosphere on the observed radiation. To do this, the surface and atmospheric parameters are varied and the outputs from the forward model are recorded. CCA is then used to build a linear “inverse model” from the database of model runs. The model is based on the *latent* or *inherent* relationships between the observed radiance and the parameters of interest. These latent relationships exist in a lower-dimensional orthogonal space making the problem more tractable and better conditioned. This orthogonality property aids the separability of surface and atmospheric effects. The inverse model can be designed to directly predict surface temperatures, or to predict transmission, upwelled radiance, and downwelled radiance for atmospheric compensation. To obtain spectral emissivity estimates, the resulting surface-leaving radiance is processed with a variation of the ASTER Temperature and Emissivity Separation (TES) algorithm. CCA can also be used to retrieve atmospheric physical parameters such as temperature and water vapor profiles.

Several databases were built using MODTRAN as the forward model. The databases included variations in atmospheric profiles, elevation, time of day, date, geographical coordinates, surface temperature, and surface emissivity. The CCA inverse model was then applied to the MODTRAN observations and the results compared to the model inputs. This was done at various spectral configurations and resolutions. Also, the inverse model was tested with MAS and MASTER thermal imagery. The results show that it is feasible to retrieve the surface temperature and emissivity with high accuracy using multispectral or hyperspectral sensors. The results also confirm that higher spectral resolution, assuming constant signal-to-noise ratio, can lead to better estimates of atmospheric and surface

parameters. A pragmatic question then follows: how many sensor bands are necessary and where should these bands be placed? The CCA approach provides insight into this problem through the analysis of the “pathways” that map the observed data and the latent canonical space. Results of a case study performed in the Midwave Infrared (MWIR) region of the spectrum show how it is possible to determine the least number and placement of bands for the retrieval of temperature and water vapor vertical profiles with reasonable accuracy.

1.1 Notation

I have made an effort to be consistent with the notation used in this dissertation. I have also tried to conform to standard notation used in the literature. Unfortunately, these two goals are often contradictory and there are bound to be some inconsistencies. However, certain general guidelines are followed:

- Vectors are bold face and lower case (e.g., $\mathbf{x}$) and assumed to be column vectors unless otherwise noted.
- Matrices are bold face and upper case (e.g., $\mathbf{A}$). Multivariate matrices are column-oriented so that the variables are defined along the column dimensions and the number of observations along the rows.
- Estimates of a variable are noted with the $\hat{}$ symbol (e.g., $\hat{y}$ is an estimate of y).
- A “known” variable that is used to estimate or predict another parameter is generically represented by x . The parameter to be estimated (i.e., the predictand) is y . It is important to emphasize this because the role of the physical and optical parameters change depending on whether the analysis is in the context of a forward or inverse model. Consequently, the values that the generic variables x and y represent also change.

1.2 Scope and Limitations

This section identifies related technical areas that are not covered by this research. The section is included for the benefit of researchers who already have a background in thermal remote sensing and are looking for work done on a particular issue. There are many complications that will not be addressed here. These are delineated in terms of the two major tasks: (1) atmospheric characterization and, (2) temperature and emissivity estimation.

1.2.1 Atmospheric Characterization

The layered-structure of the atmosphere is assumed to be in local thermodynamic equilibrium. This limits the characterization of the atmosphere between the surface and an altitude of about 100 km (Houghton 1984). When extending these techniques to spaceborne sensors, the breakdown of thermodynamic equilibrium must be addressed if the interest is in the characterization of the atmosphere. This limitation is irrelevant if the goal is the characterization of the Earth's surface.

The effect of clouds on the observed radiation are not considered in this research. The geometrical aspects of accounting for significant cloud cover are complicated and would detract from the main purpose of this research. The automated identification of clouds and the compensation of their effects is a topic of considerable research. The test imagery used in this research were acquired on relatively clear days, thus ensuring that the presence of clouds did not affect the analysis.

Geometrical effects due to the view angle of the sensor are also not characterized. As the viewing angle of the sensor changes, the length of the propagation path also changes. However, the CCA inverse models were all defined to a nominal sensor altitude and assume a nadir-viewing configuration. In an operational environment, the viewing geometry should be determined first and then incorporated into the design of the CCA inverse model.

It is assumed that the atmospheric constituent absorption characteristics and band models implemented by the radiative transfer code are true and accurate. Alternatively, any errors in the physical forward model will translate to errors in the inverse model.

1.2.2 Temperature and Emissivity Estimation

Viewing angle can also affect the apparent reflectance of a surface target. It was not the goal of this research to characterize the bidirectional reflectance distribution function (BRDF) of surface emissivities. In general, reflectances in the LWIR region of the spectrum tend to be Lambertian so this is not a gross assumption. However, there are instances when a particular material will have a significant specular component. These cases are not specifically addressed here but should be considered in an operational setting. In addition, a complete characterization of heterogeneity effects due to surface geometry (i.e., orientation of the targets with respect to the sensor viewing geometry) will not be presented.

Adjacency effects are also not included. Here, “adjacency” refers to radiation emitted or reflected by surrounding targets that enter the sensor’s Field-of-View (FOV). Typically, this is a negligible effect in the LWIR. However, if the effect is expected to be significant in a particular scene or application, then contextual information about the target’s surroundings must be included in the analysis.

Finally, it is not the intent of this research to address issues involving the extent of spectral mixing given the spatial resolution of a given sensor. This is more of a concern when implementing target detection or classification algorithms. Thus, only *effective* temperatures and emissivities were derived for a given Ground Instantaneous Field of View (GIFOV). Any errors due to spectral mixing are grouped into the overall error in the estimated parameters.

Chapter 2

Background

The important thing in science is not so much to obtain new facts as to discover new ways of thinking about them....

Sir William Bragg

A science is any discipline in which the fool of this generation can go beyond the point reached by the genius of the last generation.

Max Gluckman

No scientific endeavor should be undertaken without first exploring past research. The literature research presented in this dissertation summarizes previous work on the characterization of the Earth's surface and atmosphere from remote sensing platforms. Previous work has largely resided in two separate communities: atmospheric physicists and Earth scientists. Here, "Earth scientists" are researchers with a focus on remote sensing of the Earth's surface. This includes geologists, ecologists, environmentalists, etc. Atmospheric physicists focus on remote sensing of the atmosphere and are interested in its physical chemistry and effects on climate. Generally, both fields have been developed independently. This is unfortunate, because the strong coupling of atmospheric and surface radiation effects demands a unified and comprehensive analytical approach. At the very least, we should

attempt to minimize the duplication of research efforts. One of the goals of this research was to find a common ground and develop a technique that would be useful in both fields.

In this chapter, the underlying theory of this research is presented in two major sections. The first section covers the physics of atmospheric radiation and propagation relevant to remote sensing. It starts with a general overview of mathematical models for the *forward* propagation of energy. Based on these models, three methods for developing an *inverse* model are described: model-matching, atmospheric sounding, and the In-Scene Atmospheric Compensation (ISAC) algorithm. The goal of these methods is to compensate for the atmospheric effects to obtain a good estimate of the surface radiance. The second section covers current algorithms available to separate surface temperature and emissivity from the surface radiance. The chapter ends with a general discussion that compares the advantages and disadvantages of all of the reviewed techniques. The discussion provides a framework for the development of the approach used in this research (see Chapter 3).

2.1 Atmospheric Radiation and Propagation

2.1.1 The Atmosphere

The state of a static atmosphere is described in terms of the parameters of the ideal gas law:

$$P = nk_B T \quad (2.1)$$

where P is the atmospheric pressure, n is the number density, k_B is Boltzmann's constant, and T is the temperature. The atmosphere is commonly represented as a stack of layers, each having uniform temperature, composition, mixing, or ionization distribution. For the purpose of this research, we are concerned with the temperature distribution characteristics of the atmosphere, which happens to be the primary method for classification. In this scheme, a layer is labeled with the suffix "sphere" and a boundary with the suffix "pause". Figure 2.1 illustrates this scheme. The lowest layer is the troposphere which extends from

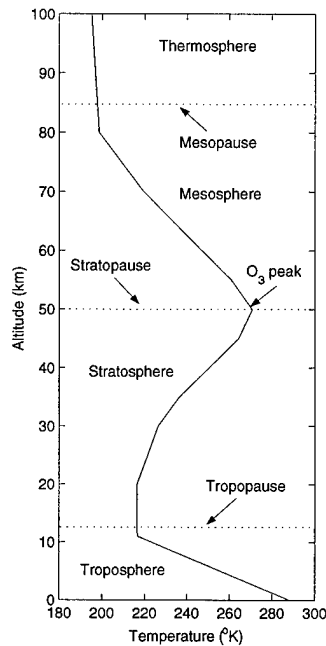


Figure 2.1: Vertical temperature profile for the 1976 U.S. Standard Atmosphere.

the Earth's surface to an altitude of about 10-12 km where it is bounded by the tropopause. In the troposphere, the temperature decreases at a rate of approximately $10^{\circ}\text{K}/\text{km}$ or less. This is true everywhere except just over the surface of the Earth where diurnal cycles cause a temperature *inversion*. When this occurs, temperature increases with height over the first kilometer or so and then decreases with altitude at a steady rate. This change in temperature with altitude is known as the *lapse rate* (Saucier 1989). The region above the tropopause is the stratosphere, which exhibits a negative lapse rate (i.e., the temperature actually increases with height) up to a maximum at about 50 km due to ozone heating. This region is bounded by the stratopause. The next region is the mesosphere where the temperature decreases with altitude until reaching the mesopause at about 85 km. The temperature at the mesopause is the coldest in the atmosphere: approximately 180°K . The last region is the thermosphere, where the temperature increases dramatically due to heating from direct solar ultraviolet radiation. The temperature reaches a maximum of about 1000°K and then levels off into an isothermal state (Hargreaves 1992).

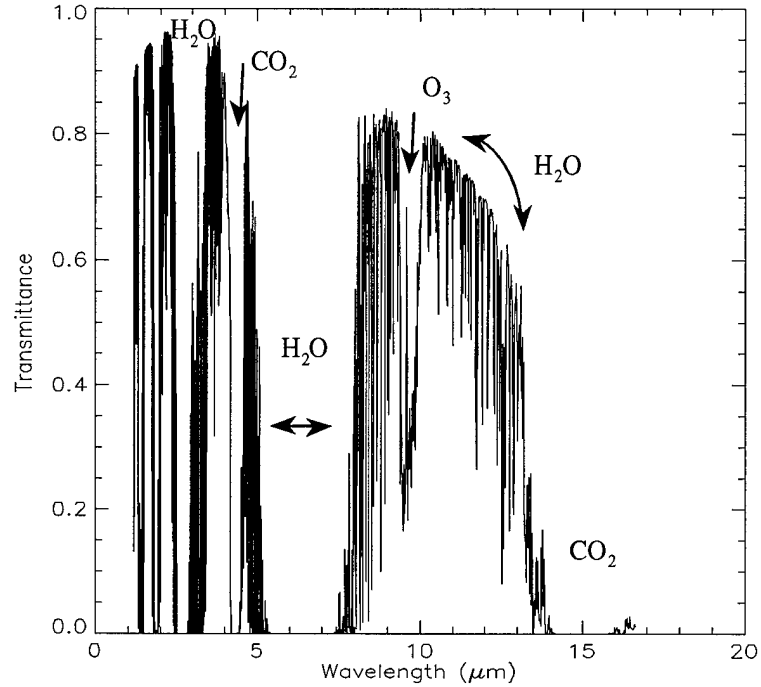


Figure 2.2: Infrared Spectrum

2.1.2 The Infrared Spectrum

To characterize the effects of the atmosphere on the propagation and emission of infrared radiation, it is necessary to understand the characteristics of the infrared spectrum. Figure 2.2 shows the prominent infrared features of the atmospheric emission. The continuum and absorption features of the spectrum result from contributions of the species present in the atmosphere. The most prominent features of the spectrum are driven by the amount of water vapor, carbon dioxide, and ozone (Liou 1980). These gases are key players in the greenhouse effect observed in the atmosphere. Other gases also absorb thermal radiation, such as methane, sulfur dioxide, nitrous oxide, and carbon monoxide. For the purposes of this research, we shall concentrate on the primary absorbers.

Water Vapor

Water vapor is the most influential atmospheric absorber due to its wide spectral coverage and concentration. On average, the concentration of water vapor is greatest in the troposphere and drops off considerably above an altitude of 12 km. One of the greatest challenges in dealing with water vapor, however, is that its concentration deviates considerably from this average depending on time and location (Smith 1993). Water vapor exhibits strong vibrational-rotational bands centered around 1.4 μm , 1.9 μm , 2.7 μm , and 6.3 μm (in wavenumbers, 7143 cm^{-1} , 5263 cm^{-1} , 3704 cm^{-1} , and 1587 cm^{-1}). Most of the infrared radiation between 20 μm to about 1 mm ($< 500 \text{ cm}^{-1}$) is absorbed by the rotational state of water vapor (Houghton and Smith 1966).

The “atmospheric window” between the 6.3 μm band and the rotational band is key to thermal remote sensing. This is because the peak values of the Planck curves derived from surface and tropospheric temperatures are located within this spectral range. Although the atmosphere is highly transmissive in this region, about 10% of the energy is absorbed by the water vapor continuum. Furthermore, weak high-J lines from the 6.3 μm and the rotational bands are superimposed on the continuum. These lines are due to transitions between molecular energy levels defined by the angular momentum quantum number (Goody and Yung 1989). The exact nature of the continuum and line absorption characteristics of water vapor are still a matter of much research and debate (Prata et al. 1995; Goody and Yung 1989; Wan and Li 1997).

Carbon Dioxide

In contrast to water vapor, carbon dioxide is uniformly concentrated up to an altitude of 80 km with little spatial or temporal variation (Smith 1993). However, CO_2 concentration increases in the spring and decreases in the late summer/early fall. There is also a yearly increasing trend in CO_2 as shown in Figure 2.3 (Keeling and Whorf 1999). CO_2 has two strong fundamental vibrational bands at 4.3 μm and 15 μm (Houghton and Smith 1966).

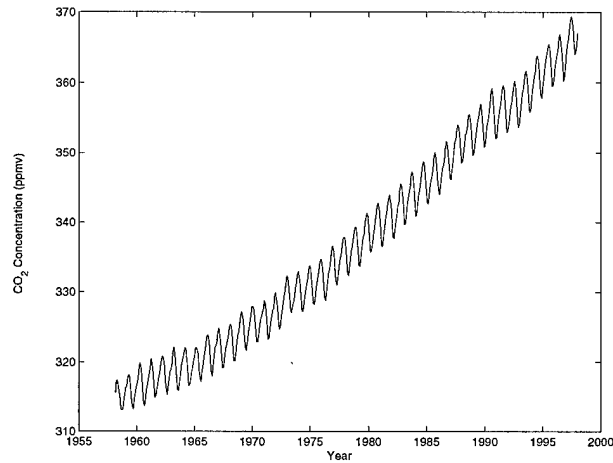


Figure 2.3: Carbon Dioxide Seasonal and Yearly Variation

The continuum effects of CO₂ are negligible in the thermal region of the spectrum. There are, however, several weak lines in the atmospheric window bounded by the water vapor and the 15 μm CO₂ bands (Smith 1993).

Ozone

The concentration of ozone has been of considerable interest due to the theorized depletion near the poles. These studies typically deal with ozone's absorption of ultraviolet radiation. Ozone is also of great importance in the thermal region of the spectrum. Unlike water vapor and carbon dioxide, it is present mostly in the stratosphere and has a strong absorption band at 9.6 μm (Houghton and Smith 1966).

2.1.3 The "Forward" Model

The mathematical description of the radiation propagation and emission processes that lead to the radiance observed by a remote sensing platform is often referred to as the "forward" model. The model is deterministic and the term "forward" is used to differentiate it from "inverse" models that are not exact and require the use of inference. The basic premise

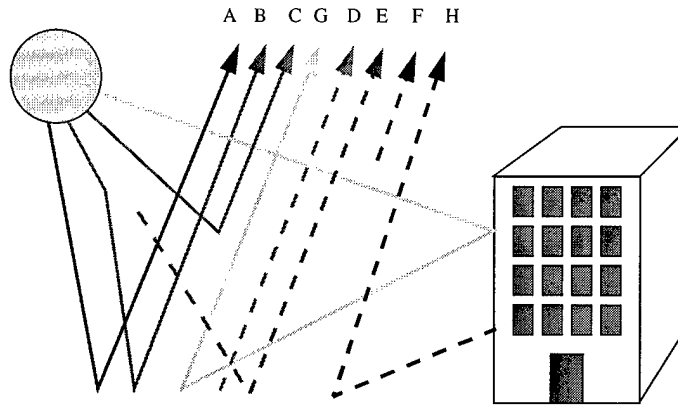


Figure 2.4: Radiation propagation paths for reflective and thermal regions (Courtesy of Kevin Ayer and DIRS Lab).

is that the observed radiance is a function of the scene in view and the composition and thermodynamic state of the intervening atmosphere.

The radiometric formulation of radiative transfer is fairly standard across the literature with the exception of the notation used. I will mostly use the notation introduced by Schott (1997). The radiance reaching a sensor is the sum of the contributions from different propagation paths as shown in Figure 2.4. The propagation paths depend on the region of the electromagnetic spectrum of the radiation. In the reflective (or solar) region, the dominant paths are: (A) direct sunlight hits a target and reflects, (B) sunlight scatters in the atmosphere and reaches a target and is then reflected, (C) sunlight scatters in the atmosphere and reaches the sensor, and (G) sunlight reflects off the background, reaches the target, and is then reflected. In the emissive (or thermal) region, the dominant paths are: (D) thermal photons emitted by a target reach the sensor, (E) thermal radiation from the atmosphere reaches the target and is reflected, (F) thermal photons from the atmosphere reach the sensor, and (H) thermal photons from the background reach a target and are reflected. In this research, we shall look at radiation between $4\text{ }\mu\text{m}$ and $15\text{ }\mu\text{m}$ so that we can neglect the reflective region components of radiative transfer. This separation is well

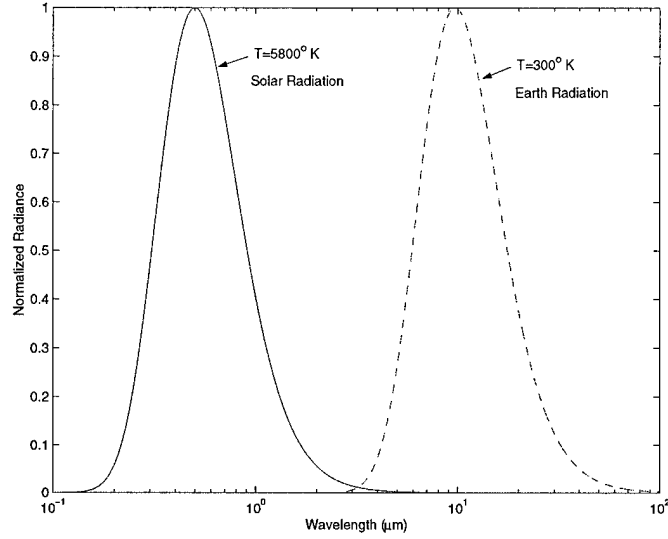


Figure 2.5: Radiation originating from solar and terrestrial temperatures

illustrated in Figure 2.5. The left curve is the radiation associated with solar temperatures while the curve on the right is associated with terrestrial temperatures. Thus, for certain Midwave Infrared (MWIR) and all Longwave Infrared (LWIR) cases, the radiation at the sensor is simply due to the path radiation emitted by the atmosphere and the surface. If the surface has a low emissivity, the reflectivity is large and it is necessary to include the reflected downwelled term (Jun 1994). The impact of the reflected downwelled radiance depends on the contrast of temperature between the sky and the surface of the earth.

The relationship between the emissivity and the reflectivity is obtained from Kirchhoff's Law, which states that the spectral absorption of an opaque object at thermal equilibrium is the same as the spectral emissivity. Thus,

$$\varepsilon + r = 1 \quad (2.2)$$

where ε is the emissivity and r is the reflectivity. The emissivity determines how much radiation will be given off by an object given its current temperature. It is really a measure of the radiation emitted by a particular object compared to a blackbody. The blackbody

radiation distribution is described by Planck's Law:

$$L_{BB}(T) = \frac{c_1}{\lambda^5 [e^{c_2/\lambda T} - 1]} \quad \frac{\mu\text{W}}{\text{cm}^2 \cdot \text{sr} \cdot \mu\text{m}} \quad (2.3)$$

where $c_1 = 1.19104 \times 10^{10} \mu\text{W} \cdot \mu\text{m}^4 / \text{cm}^2 \cdot \text{sr}$ and $c_2 = 1.43877 \times 10^4 \mu\text{m} \cdot \text{K}$ are the Planck radiation constants. The radiation units are often referred by the shorthand microflicks (μf). The temperature T is specified in Kelvin and wavelengths in μm . The amount of spectral radiation emitted by an object is the product of the emissivity and Planck's blackbody function.

The infrared spectral radiance reaching the sensor is then described by the radiative transfer equation

$$L(\lambda) = \tau(\lambda) \varepsilon(\lambda) L_{BB}(\lambda, T) + \tau(\lambda) [1 - \varepsilon(\lambda)] L_d(\lambda) + L_u(\lambda) \quad (2.4)$$

where $\tau(\lambda)$ is the transmission along the upwelling path (i.e., the atmosphere between the target and the sensor), $L_d(\lambda)$ is the downwelling radiance from the sky, $L_u(\lambda)$ is the upwelling radiance, and $\varepsilon(\lambda)$ is the spectral emissivity. Hereafter, the notation will be simplified by dropping the explicit reference to wavelength λ except when needed for clarity. Equation (2.4) can be simplified by collecting terms containing τ . The spectral radiance reaching the sensor can then be defined as

$$\begin{aligned} L &= \tau [\varepsilon L_{BB}(T) + (1 - \varepsilon) L_d] + L_u \\ &= \tau L_s + L_u \end{aligned} \quad (2.5)$$

This simple linear form of the radiative transfer equation suffices for the development of the In-Scene Atmospheric Compensation (ISAC) algorithm (see section 2.1.6). However, in order to develop the framework for the atmospheric sounding and model-matching algorithms, it is necessary to further analyze the atmospheric radiation components of equation (2.4) (i.e., the upwelled and downwelled radiance terms). I shall do this by deriving the equation of radiative transfer from first principles.

Schwarzschild's Equation

The derivation of radiative transfer follows the approach presented by Goody and Yung (1989) and by Liou (1980) with the addition of comments and detailed steps. First, consider the simple case of a beam of *monochromatic* radiation propagating through a homogeneous layer of the atmosphere. At any given wavelength, the propagating radiation L will be attenuated by the intervening air mass depending on the concentration of constituents, their size, and the thickness of the layer. Thus, the change in radiation due to propagation in a layer of thickness ds is

$$dL = -mC_{ext}L ds = -\beta_{ext}L ds \quad (2.6)$$

where m is the effective number density of constituents in the layer [m^{-3}], C_{ext} is the effective extinction cross-section of the constituents [m^2], and β_{ext} is the extinction coefficient [m^{-1}], which may vary with altitude depending on vertical homogeneity. For a thin layer ds , it is reasonable to assume that the layer is homogeneous and that β_{ext} is constant. At this point, it is important to note that this formalism implicitly assumes that the extinction of radiation is only due to absorption and single scattering phenomena. Fortunately, multiple scattering effects are negligible in the infrared. The radiation propagation over an entire atmospheric column is obtained by integrating the effects of each layer. After rearranging and integrating, equation (2.6) yields

$$\int_0^z \frac{dL}{L} = \int_0^z -\beta_{ext}(s) ds \quad (2.7)$$

where z is the altitude. Because the integral is over the entire atmospheric column, the extinction coefficient dependence on altitude is explicitly shown. To be concise, this explicit dependence is dropped whenever the extinction coefficient is used in a differential equation. After integrating the left side the radiance at altitude z is

$$L(z) = L(0) \exp \left[- \int_0^z \beta_{ext}(s) ds \right] \quad (2.8)$$

This is the Beer-Lambert Law (Schott 1997; Liou 1980). The transmission is the ratio of the emerging radiation $L(z)$ and the input radiation $L(0)$:

$$\tau(z) = \frac{L(z)}{L(0)} = \exp \left[- \int_0^z \beta_{ext}(s) ds \right] = e^{-\delta(z)} \quad (2.9)$$

where $\delta(z)$ is the optical depth for the path of length z . The optical depth is a measure of the apparent thickness of a medium. Now consider the radiative transfer in the infrared region of the spectrum. The atmospheric extinction coefficient is approximately equal to the absorption coefficient assuming scattering is negligible at these wavelengths. Furthermore, each atmospheric layer will emit radiation proportional to that absorbed as dictated by Kirchhoff's Law. Adding the self-emission of the layer, eq. (2.6) becomes

$$dL = -\beta_{abs}L ds + \varepsilon L_{BB}(T_s) ds = -\beta_{abs}L ds + \beta_{abs}L_{BB}(T_s) ds \quad (2.10)$$

where T_s is the temperature for the layer ds . Rearranging we get

$$\frac{dL}{\beta_{abs} ds} = -L + L_{BB}(T_s) \quad (2.11)$$

This is Schwarzschild's equation of radiative transfer. In this case, the Planck function is the *source function* as described in the literature (Liou 1980; Goody and Yung 1989). To simplify the formulation, it is assumed that the radiation propagation is along a zenith angle of 0° (i.e., nadir viewing). Otherwise, a correction factor for the length of the propagation path would have to be introduced.

The total radiation propagation through the atmosphere is obtained by integrating the effects of each atmospheric layer. This gives rise to an integral form of Schwarzschild's equation. Figure 2.6 is a schematic of an atmosphere with length z . The height index z' denotes an arbitrary lower boundary layer. Figure 2.6 also shows that the reference origin for the optical depth is at the top of the atmosphere. This definition is typical in the literature and is due to its origins in the astronomical community (Stephens 1994). The optical depth is defined with respect to incoming radiation so that as the propagation path increases *downward*, the total optical depth increases. The upward arrows denote the

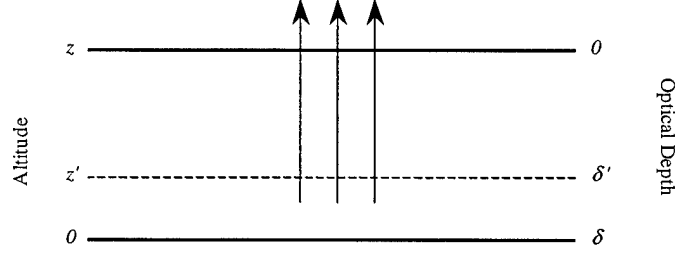


Figure 2.6: Integration path for a layered inhomogeneous atmosphere

integration path, which also corresponds to the outgoing radiation propagation of interest in remote sensing.

Recall from eq. (2.9) that the optical depth and the transmission are related exponentially. From this relationship and from the Beer-Lambert Law, the optical depth for a layer with lower boundary layer z' is

$$\delta(z', z) = \int_{z'}^z \beta_{abs}(s) ds = -\ln [\tau(z', z)] \quad (2.12)$$

Thus, a transmissive atmosphere has a small (thin) optical depth. In the limit, a completely transmissive atmosphere (i.e., $\tau = 1$) will have zero optical depth. A small change in optical depth is

$$d\delta(z', z) = d \left[\int_{z'}^z \beta_{abs}(s) ds \right] = -\beta_{abs} ds \quad (2.13)$$

The change in optical depth is negative because the optical depth decreases with altitude. Hereafter, the indices for δ are dropped and implied except when needed for clarity. Substituting eq. (2.13) into eq. (2.11) results in

$$-\frac{dL}{d\delta} = -L + L_{BB}(T_s) \quad (2.14)$$

and the problem reduces to solving a nonhomogeneous linear differential equation. First multiply through by $e^{-\delta}$ and $d\delta$ to get

$$-dL e^{-\delta} = -L e^{-\delta} d\delta + L_{BB}(T_s) e^{-\delta} d\delta \quad (2.15)$$

Rewriting in terms of transmission yields

$$-(dL)\tau = L(d\tau) + L_{BB}(T_s)e^{-\delta} d\delta \quad (2.16)$$

After rearranging, applying the chain rule, and substituting eq. (2.13), we obtain:

$$-[(dL)\tau + L(d\tau)] = -d[L\tau] = -L_{BB}(T_s)\beta_{abs}(s)e^{-\delta} ds \quad (2.17)$$

Now integrating both sides yields

$$\int_0^z d \left[L(z')e^{-\delta(z',z)} \right] = \int_0^z \beta_{abs}(s)L_{BB}(T_s)e^{-\delta(s,z)} ds \quad (2.18)$$

$$\Rightarrow L(z')e^{-\delta(z',z)} \Big|_0^z = \int_0^z \beta_{abs}(s)L_{BB}(T_s)e^{-\delta(s,z)} ds \quad (2.19)$$

Solving for the spectral radiance results in

$$L(z) = L(0)e^{-\delta(0,z)} + \int_0^z \beta_{abs}(s)L_{BB}(T_s)e^{-\delta(s,z)} ds \quad (2.20)$$

When applying this equation to a propagation path beginning at the surface ($z = 0$) of the Earth up to some distance z where the sensor is located, $L(0) = L_s$, and eq. (2.20) becomes identical to eq. (2.5) where

$$L_u = \int_0^z \beta_{abs}(s)L_{BB}(T_s)e^{-\delta(s,z)} ds \quad (2.21)$$

and $\tau = e^{-\delta(0,z)}$. For an atmosphere stratified in discrete layers, the upwelled radiance term becomes

$$L_u = \sum_{i=0}^N \beta_i L_{BB}(T_i)(\tau_{i+1} - \tau_i) \quad (2.22)$$

where there are N discrete layers and $\tau_{i+1} - \tau_i$ is the difference in the effective transmission of adjacent layers. This expression for upwelled radiance is common in the literature (Schott 1997).

The downwelled radiance term L_d in equation (2.5) can be derived in the same manner as the upwelled radiance except the propagation path is reversed. The integral form of the downwelled radiance is

$$L_d = \int_z^0 \beta_{abs}(s)L_{BB}(T_s)e^{-\delta(s,z)} ds \quad (2.23)$$

Again, the propagation path will be longer than z for off-nadir viewing and a correction factor must be included. Alternatively, we can think of z as being the propagation path length instead of the altitude.

Band Models and Absorption

While the radiative transfer model is fairly standard, the implementations of band models and line absorption features are not. The absorption coefficient was introduced in equation (2.10). However, the spectral structure of this parameter was only implied. This section addresses the implications of the spectral width of the absorption coefficient.

In reality, the shape of a molecular absorption feature has a finite width. Thus, the center position of the absorption feature, the strength of the absorption, and the shape of the band are all needed to fully characterize molecular absorption. In general, the shape of an absorption band is defined by a *shape factor* (Stephens 1994). The shape factor is a function of frequency that best matches laboratory spectra. Typically, the shape factor is based on some probability density function. The absorption coefficient is then represented as

$$\beta(\nu - \nu_0) = Sf(\nu - \nu_0) \quad (2.24)$$

where S is the strength of the absorption, f is the shape factor, ν is the radiation frequency, and ν_0 is the center frequency of the absorption band. The uncertainty principle causes a natural broadening of absorption features. In general, the broadening of line absorptions in the upper-atmosphere are attributed to this effect. In the lower atmosphere, the broadening of absorption features is due to pressure and Doppler effects resulting from collisions and molecular motion. These broadening effects are stronger than natural broadening and dominate atmospheric transmittance spectra. The understanding of these atmospheric effects is important for the interpretation of the atmospheric sounding scheme.

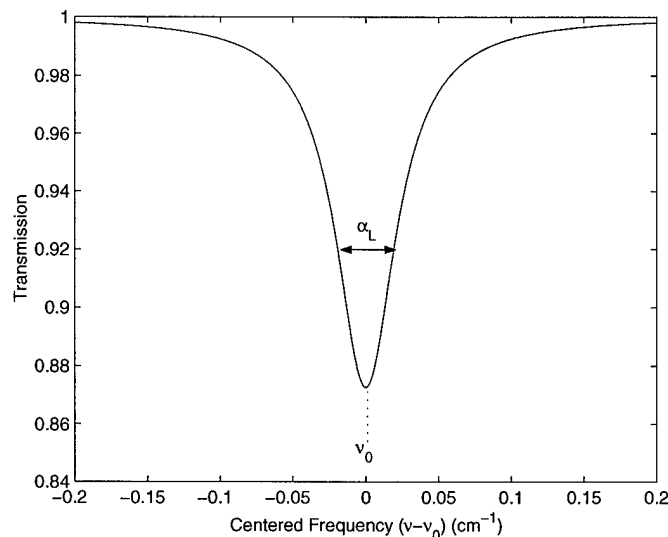


Figure 2.7: Lorentz line shape for an absorption band.

Pressure Broadening. A heuristic approach to the modelling of absorption broadening due to molecular collisions is based on a phase-shift model of molecular oscillatory motion. Consider a simple molecule represented by a dipole oscillator. Upon collision by another molecule, the phase of the oscillation is shifted randomly. When there are many collisions, the thermal energy associated with the molecular absorption and the electrical energy associated with the oscillation are equal. In other words, there is local thermodynamic equilibrium. A mathematical description of collision broadening based on this phase-shift model and Fourier analysis is

$$f_L(\nu - \nu_0) = \frac{\alpha_L}{\pi} \frac{1}{(\nu - \nu_0)^2 + \alpha_L^2} \quad (2.25)$$

where f_L is the Lorentz line shape factor and α_L is the Lorentz half width (Stephens 1994). The Lorentz shape factor can also be derived from the theory of absorption and dispersion (Liou 1980). An example absorption band is shown in Figure 2.7.

The Lorentz half width is $\alpha_L = 1/2\pi\bar{t}$ where $\bar{t}$ is the mean time between collisions (Houghton and Smith 1966; Stephens 1994). From kinetic theory, the half width of a spectral line can

be approximated as

$$\alpha_L \approx \alpha_{L,s} \frac{p}{p_s} \sqrt{\frac{T_s}{T}} \quad (2.26)$$

where $p_s = 1000$ millibar (mb) and $T_s = 273^\circ$ K are standard temperature and pressure (*stp*) and $\alpha_{L,s}$ is the half-width at *stp*. The relationship between the half width and the ambient pressure and temperature is crucial. As the pressure decreases relative to the standard pressure, the width of the Lorentzian curve decreases and the absorption at the center frequency increases. Conversely, as the pressure increases, the curve broadens and the absorption at the center frequency decreases. The line shape dependence on temperature is typically ignored, though this introduces some errors in the band model (Stephens 1994). Fortunately, the error is negligible for infrared radiation propagation because pressure changes are much larger than temperature changes (Liou 1980).

Doppler Broadening. When a molecule is moving either away from or towards an “observer”, this causes a Doppler frequency shift in the emitted radiation. The Doppler shift has a Gaussian distribution because of the range of velocities and the Central Limit Theorem. Therefore, the resulting Doppler broadening is also of the Gaussian form. The magnitude associated with Doppler broadening is much smaller than for pressure broadening in the lower atmosphere. In the upper atmosphere, the Doppler broadening effect is more dominant and appropriate changes to the band model must be made. The Gaussian shape factor for Doppler broadening depends on the half width which originates from Maxwell’s Laws and the distribution of molecular velocities. The shape factor is

$$f_D(\nu - \nu_0) = \frac{1}{\sqrt{\pi}\alpha_D} \exp \left[-\frac{(\nu - \nu_0)^2}{\alpha_D^2} \right] \quad (2.27)$$

where $\alpha_D = u_m \nu_0 / c$ is the Doppler half width, u_m is the RMS molecular velocity, and c is the speed of light. The average molecular velocity is

$$u_m = \sqrt{\frac{2k_B T}{m}} \quad (2.28)$$

where m is the molecular mass and k_B is Boltzmann's constant. The average velocity is obtained from the Maxwell distribution of velocities:

$$p(u) = \left(\frac{m}{2\pi k_B T} \right)^{1/2} \exp \left(-\frac{mu^2}{2k_B T} \right) \quad (2.29)$$

The half width depends on the square root of the ambient temperature. Thus, the temperature may be inferred from an accurate measurement of the absorption line if Doppler broadening is the dominant effect (Stephens 1994).

In the region where both Doppler and pressure broadening are equally important, the Voigt line shape is used (Stephens 1994). The Voigt shape factor is a convolution of the pressure broadening shape factor with the probability distribution of the molecular velocities (Goody and Yung 1989). That is,

$$f_V(\nu - \nu_0) = \int_{-\infty}^{+\infty} f_L \left(\nu - \nu_0 - \frac{u\nu_0}{c} \right) p(u) du \quad (2.30)$$

Band Transmission Functions At the beginning of this section, the theory of radiative transfer considered monochromatic radiation only. That is, the transmission was evaluated at a single wavelength. In practice, the radiation is measured over a spectral bandpass that is defined by the spectral response of the detector. This spectral response has the effect of "blurring" the monochromatic spectrum. The band transmission is then defined as

$$\mathcal{T}(z) = \left(\frac{1}{\nu_2 - \nu_1} \right) \int_{\nu_1}^{\nu_2} R(\nu) e^{-\int_0^z \beta(\nu, s) ds} d\nu \bigg/ \left(\frac{1}{\nu_2 - \nu_1} \right) \int_{\nu_1}^{\nu_2} R(\nu) d\nu \quad (2.31)$$

where $R(\nu)$ is the spectral response function of the detector, ν_1 and ν_2 define the bandpass of the detector, the exponential term in the numerator represents the monochromatic transmission, and the denominator is the normalization term (Liou 1980; Schott 1997; Stephens 1994). The constants in front of the integrals cancel and equation (2.31) simplifies to

$$\mathcal{T}(z) = \frac{\int_{\nu_1}^{\nu_2} R(\nu) e^{-\int_0^z \beta(\nu, s) ds} d\nu}{\int_{\nu_1}^{\nu_2} R(\nu) d\nu} \quad (2.32)$$

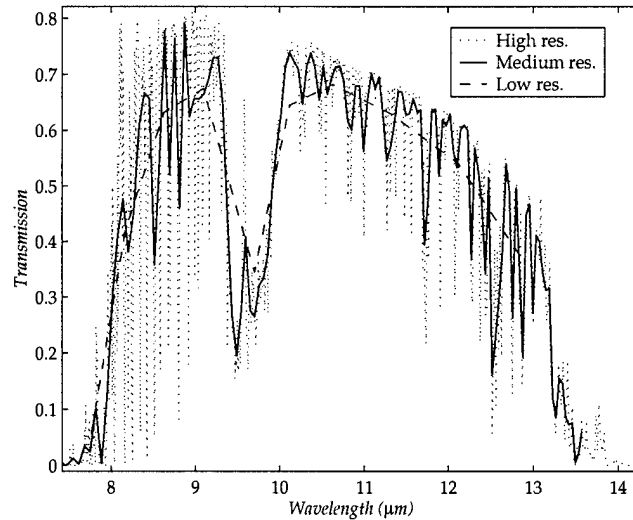


Figure 2.8: Effects of spectral resolution on band transmission functions.

Thus, the band transmission function is the convolution of the spectral response with the monochromatic transmission. Note that the absorption coefficient is defined in terms of a shape factor (see eq. (2.24)). Figure 2.8 is an example of how the band transmission function changes depending on the spectral resolution of the sensor. Strong line absorption features are still resolvable but have less intensity in the transition from high to medium resolution (1101 to 128 bands). Several features completely disappear in the low-resolution transmission function (10 bands).

Radiative Transfer Models

The radiative transfer depicted by eq. (2.4) is often implemented to some degree in computer models that use databases containing molecular and particulate absorption and scattering characteristics. Various models exist with different degrees of spectral resolution, number of atmospheric constituents, cloud models, etc. Computer models can be broken down into two major classes: line-by-line transmittance and band transmittance models.

Line-by-line Models Line-by-line models are very high resolution radiative transfer models that use a large database of molecular absorption and scattering measurements. The most widely used database is HITRAN96, which contains over 1 million absorption lines for 35 molecules (HITRAN 2000). The model FASCODE, developed by the U.S. Air Force, taps into this database to generate transmittance calculations at high spectral resolution. The radiation propagation is performed on a line-by-line basis so that the radiation is monochromatic and the Beer-Lambert Law holds. Unfortunately, this requires long computation times which are often impractical. Other “faster” line-by-line models have been built, such as PLOD and OPTRAN, which are derivatives of the GENLN2 model. These are usually tailored to a particular sensor and account for a smaller set of atmospheric constituents.

Band Transmittance Computer Models These models are designed to lower the resource cost of computing the radiative transfer of monochromatic radiation through an inhomogeneous path in the atmosphere. This is done by using band transmittances in the radiative transfer computation. The most widely used model is MODTRAN developed by the Air Force Research Laboratory (AFRL) (originally the Air Force Geophysics Laboratory). MODTRAN4 is the current version which includes significant improvements like updated cloud and water vapor models, spectral emissivity inputs, all-sky downwelling radiance calculations, etc. MODTRAN computes band transmittances using a statistical model which integrates Voigt absorption lines over a spectral range of 1 cm^{-1} . The band transmittance model is parameterized with pressure, temperature, absorption coefficient, and average line width and strength (Kneizys et al. 1996). MODTRAN uses the Curtis-Godson approximation for radiative transfer to determine an effective band absorption model for an inhomogeneous atmosphere. The approximation is based on the assumption that the effective absorption band can be represented by the integrated values of the band model parameters over the radiation path (Liou 1980; Berk et al. 1999). Thus, an effective Lorentz

band may be computed from a mean line strength and width given by

$$\bar{S} = \frac{1}{u} \int_0^u S(T) du \quad (2.33)$$

and

$$\bar{\alpha} = \frac{\int_0^u S(T) \alpha(p, T) du}{\int_0^u S(T) du} \quad (2.34)$$

where the integrals are over a path described by a variation of absorber amount du , $\bar{S}$ is the mean strength associated with a total absorber amount u , and $\bar{\alpha}$ is the mean line width that depends on pressure p and temperature T . This approximation is particularly useful in the infrared except when dealing with the $9.6 \mu\text{m}$ O_3 band. This is because ozone exists in larger concentrations at higher altitudes where the pressure is low (Liou 1980; Goody and Yung 1989).

2.1.4 Model-Matching

Now that we have a good understanding of the radiative process, we wish to find a way to use it so that the atmospheric radiation can be characterized and compensated. From the “Earth scientist” perspective, the atmosphere introduces structured noise that must be removed to properly analyze the Earth’s surface. To estimate the atmospheric effects, computer models such as those described in Section 2.1.3 are used to obtain the atmospheric spectra in the radiative transfer equation (2.4). Once these values are obtained, the surface radiance may be calculated. Unfortunately, this requires *a priori* knowledge of the atmospheric conditions.

Another approach that uses forward models is to dynamically vary the model inputs until the output best “matches” the observed radiance based on some criterion (e.g., least-squares). The solutions are obtained by iterations that are often nonlinear. To achieve convergence, it is necessary to apply physical constraints. In addition, a parametric scheme for the model inputs must be devised. These parameters can be thought of as knobs in a

machine that “tweak” the process until the desired output is obtained. Finally, a “good” estimate of the solution is required to initialize the algorithm. The atmospheric conditions used to match the observed spectra are then used in the radiative transfer equation to solve for the surface radiance.

Green et al. (1993) established this method for the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) to estimate aerosol content, pressure elevation, water vapor content, and surface reflectance using MODTRAN2. After estimation of these parameters, a “correction” vector can be generated to compensate the observed radiance for atmospheric effects and obtain surface radiance (Green et al. 1996).

The model-matching approach is normally implemented using a numerical nonlinear least-squares fitting technique. Green (1993) and Young (1998) used the downhill simplex regression method. This numerical technique works in an N -dimensional space spanned by the parameters in the model. It finds the minimum of a function (e.g., the squared error) with a geometrical *simplex* that expands and contracts in the parameter space until it converges to a minimum (Press et al. 1992; Sanders 1999). Figure 2.9 is a schematic showing the changes a three-dimensional simplex could take at each iteration. The advantage of this approach is that it does not need analytical function derivatives which may be too complex for high-dimensional parameter spaces. Sanders (1999) has successfully implemented this technique in the Interactive Data Language (IDL) using the `amoeba.pro` routine. One of the greatest challenges in the implementation is the presence of local minima in the parameter space. There is a tradeoff between rate of convergence and susceptibility to local minima which must be assessed for each particular situation. However, a search for a global minimum may be done by executing the algorithm again using the first solution as the initial estimate. The minimum is more likely to be global if the algorithm converges to the initial estimate (Sanders 1999). Also, recent numerical methods, such as *simulated surface annealing*, provide an effective methodology for the search of a global minimum (Press et al. 1992). Unfortunately, these techniques can be very computationally intensive.

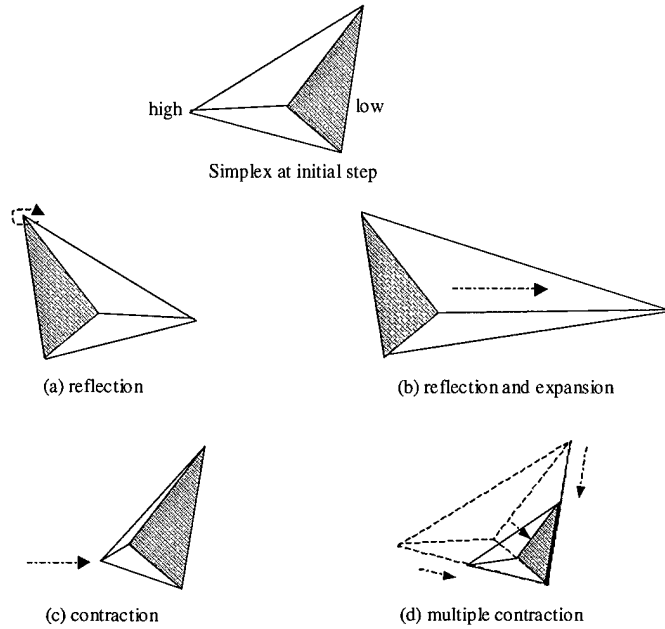


Figure 2.9: Schematic of downhill simplex regression method. The top diagram is the simplex at the beginning of the iteration. The simplex can then (a) reflect away from the high point, (b) reflect away from the high point and expand along one dimension to a new high point, (c) contract toward the low point along one dimension, and (d) contract toward the low point along all dimensions. A combination of these steps over several iterations leads to convergence (Press et al. 1992). (Reprinted with the permission of Cambridge University Press.)

Another advantage of the model-matching approach is the flexibility in the choice of model parameters. For example, Johnson and Young (1998) developed a model-matching technique for infrared applications using data from the Spatially-Enhanced Broadband Array Spectrograph System (SEBASS). The technique involves using the 1976 U.S. Standard Atmosphere in MODTRAN as an initial estimate of the atmosphere. The spectral transmittance of the Standard Atmosphere was varied in the regression by changing the column densities of water vapor and ozone. The relationship between the transmittance and the column densities was given by

$$\tau(\lambda) = \left[\frac{\tau_{total}(\lambda)}{\tau_{H_2O}(\lambda)\tau_{con}(\lambda)\tau_{O_3}(\lambda)} \right] \tau_{H_2O}^{\gamma_{H_2O}}(\lambda)\tau_{con}^{\gamma_{H_2O}^2}(\lambda)\tau_{O_3}^{\gamma_{O_3}}(\lambda) \quad (2.35)$$

where γ is the column density and the subscripts denote the contribution of ozone, water

vapor line absorption, and continuum absorption. This parameterization stems from assuming that the total optical depth is the sum of individual constituent contributions. This translates to the total transmission being a product of the individual contributions. Thus, the absorption follows the Beer-Lambert Law. The column density of water is squared when applied to the transmission contribution by the water continuum. This is done to scale the parameter down because of the relatively high transmission of the continuum compared to the water vapor and ozone line absorptions. The other active constituent in this region of the spectrum is carbon dioxide. However, the concentration is maintained fixed because CO₂ is well-mixed. Two more physical parameters were allowed to change in the regression: the surface temperature and effective atmospheric temperature.

A practical aspect of the regression technique was that the MODTRAN output radiance had to be resampled to match the spectral response of the SEBASS sensor. The spectra is resampled by averaging the MODTRAN output over the spectral response of the sensor defined by the full-width at half-max (FWHM) and a triangular detector function. Furthermore, there was a spectral misregistration between the SEBASS data and the MODTRAN output so that a spectral shift had to be introduced for some of the SEBASS bands. The spectral shift and resampling parameters were also adjusted in the regression.

The optimization criterion was to minimize the RMS difference between the modelled radiance and the SEBASS observed radiance. The reported case resulted in RMS differences of 0.01 μf . Despite the coarseness of this regression model, the resulting surface temperatures for four different cases were underestimated by only 2 °C. The “coarseness” of the model refers to the number of physical parameters that were changed to acquire a match between the observed and modelled radiances. Even for physically coarse models, the model-matching approach may require a parameter-rich scheme because of other effects (e.g., spectral misregistration).

Implicit in this discussion is the assumption that an appropriate radiative transfer model is available “online” during processing. That is, the model must be executed during

each iteration to calculate a new spectral radiance observation based on the current value of the model parameters. For hyperspectral image cubes, this can lead to prohibitive computational costs. One approach is to build look-up tables (LUT's) based on results obtained with past runs of the radiative transfer model. These LUT's are then indexed appropriately and an interpolation scheme is used to fit the LUT outputs to the observed radiances. This can significantly reduce run times and conditions the problem so that convergence can be more easily achieved (Sanders 1999).

2.1.5 Atmospheric Sounding Techniques

The theory of infrared radiation propagation and emission through the atmosphere was presented in section 2.1.3. In the present section, the concept, basic theory, and implementation of atmospheric sounding are presented. The atmospheric sounding approach is to directly *invert* the forward physical model. The term "sounding" is synonymous to "probing". The basic premise is that the spectral radiance reaching a sensor is a function of the temperature of the atmosphere and the concentration of constituents. If the constituent concentration is known, then the spectral radiance reaching a sensor is a function of the temperature of the atmosphere alone. Thus, it should be possible to infer the temperature of the atmosphere based on the observed radiation. The notion of using spaceborne infrared sensors for atmospheric sounding was first introduced by King in 1956. King's method for atmospheric sounding depended on the "limb-darkening" effect, and was thus based on the measurement of the atmospheric emission at different angles. In 1959, Kaplan demonstrated that the atmospheric temperature could be inferred from measurements of the atmospheric emission at different wavelengths. Since then, several radiometers and interferometers have been developed to do atmospheric sounding (Houghton 1984).

A big benefit of atmospheric sounding is that it provides a temperature profile of the atmosphere rather than only an effective atmospheric temperature. Thus, the vertical structure of the atmosphere can be defined and more accurate estimates of the atmospheric

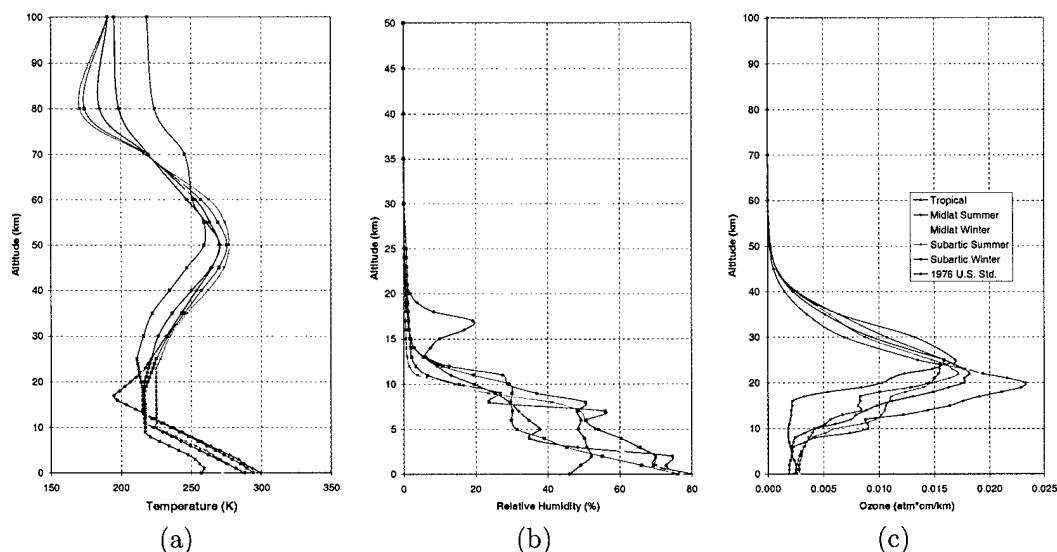


Figure 2.10: MODTRAN default vertical profiles for six model atmospheres: (a) temperature, (b) relative humidity, and (c) O_3 .

emission can be made. These vertical profiles are not only useful for climatology, but they can be used as initial estimates in a model-matching atmospheric compensation approach. However, the vertical resolution of the temperature profile is limited to the number of spectral measurements available. Historically, infrared sounders have consisted of moderate-resolution radiometers and high-resolution interferometers. The tradeoff between the two is a classical one: the higher the spectral resolution, the lower the spatial resolution. For atmospheric research, spectral resolution has traditionally been preferred. Therefore, the temperature profiles measured by previous sensors were averaged over several kilometers of a horizontal grid. While atmospheric temperatures may be constant over large areas, the same is not true for land surface temperatures (LST). The measurement of LST spatial variation is crucial for many applications. With the advent of modern hyperspectral sensors, it is becoming feasible to obtain reasonably high spectral *and* spatial resolution, thus overcoming this limitation.

In addition to obtaining the vertical temperature profile of the atmosphere, it is possible to infer the vertical concentration profiles of atmospheric constituents such as water vapor and ozone (see Figure 2.10 for typical profiles). As will be shown shortly, the determination of other atmospheric constituents is more complicated than obtaining the temperature profile. Because the temperature of the atmosphere and the constituent concentration are correlated, the solution of each profile must be obtained *simultaneously* or through the implementation of some iterative numerical method. High *spatial* resolution may also improve the retrieval of constituent concentration profiles because the spatial variation of concentration is higher than for atmospheric temperatures. The contextual information in the image may then be used to augment the spectral information and aid in the separation of the constituent concentration and temperature effects.

The determination of atmospheric temperature and constituent concentration profiles from “in-scene” data is extremely valuable. If the solutions to the inverse problem are physically meaningful, then compensation of the infrared hyperspectral imagery can be performed without using any ancillary data. Moreover, the compensation is likely to be more accurate. The compensation can also be done on a per-pixel basis such that the spatial heterogeneity of the atmosphere and the surface is accounted for in the solution. Finally, atmospheric sounding achieves the secondary goal of extracting information about the state of the atmosphere, which is also very useful. Unfortunately, accurate retrievals require many narrow spectral bands located in regions of high absorption. Therefore, considerable resources must be allocated to bands that are not useful for imaging the Earth’s surface.

The physical and mathematical framework of atmospheric sounding will be demonstrated with the retrieval of vertical temperature profiles. Practical issues with the implementation of a sounding approach are then discussed. Finally, the sounding concept is expanded to the retrieval of temperature and constituent concentrations.

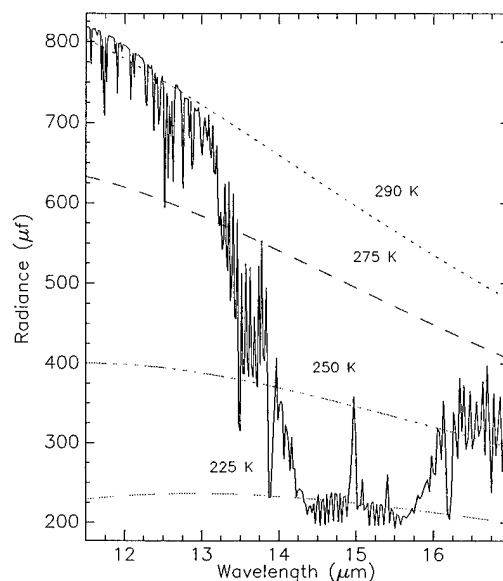


Figure 2.11: Planck curves superimposed on atmospheric radiation.

Temperature Profiles

The retrieval of temperature profiles from spectral radiance observations can be qualitatively understood by considering the emitted atmospheric radiation over a range of wavelengths associated with the absorption of a particular gas. For illustrative purposes, consider a strong absorption band where the transmission is zero at the center of the band. A good example is the $15\text{ }\mu\text{m}$ CO_2 absorption band shown in Figure 2.11. Any radiation reaching the sensor at the $15\text{ }\mu\text{m}$ wavelength must have originated in the upper layers of the atmosphere. This is because the transmission is effectively zero and all radiation from the Earth's surface is absorbed. The opposite extreme is to consider radiation at wavelengths with high transmission. At these wavelengths, very little of the radiation is coming from the atmosphere. Thus, the spectral measurement of radiance corresponds to radiation originating from different altitudes. As we move along the wing of the absorption band, the radiation reaching the sensor at each wavelength corresponds to a different altitude in the

atmosphere. If Planck curves are laid over a plot of the absorption band, then the points of intersection represent the temperature of the corresponding altitude. Four such curves are shown in Figure 2.11. These curves are calculated using eq. (2.3) for the given temperatures. The colder temperatures correspond to regions with large absorption. This makes sense if we recall that the temperature decreases with altitude in the troposphere. The sharp peak in the center of the absorption band is due to the temperature increase in the stratosphere (see Figure 2.1). This qualitative analysis demonstrates that a sensor with many narrow-band channels over the spectral range of the absorption band will provide the best vertical resolution of the estimated atmospheric profiles. The exact altitude to which each spectral band corresponds is obtained from the relationship between wavelength, transmission, optical depth, and altitude.

The vertical temperature profile is mathematically related to the observed spectral radiance by eq. (2.5). As shown in section 2.1.3, the observed spectral radiance is

$$L(\lambda) = \tau(\lambda)L_s(\lambda) + \int_0^z \beta_{abs}(s; \lambda)L_{BB}(T_s)e^{-\delta(s; \lambda)} ds \quad (2.36)$$

where the dependence on wavelength is explicitly noted. For now, it is assumed that the wavelength range is narrow enough that the Planck function is approximately independent of wavelength. Using the definition for optical depth in eq. (2.12), the radiative transfer equation can be rewritten as

$$L(\lambda) = \tau(\lambda)L_s(\lambda) - \int_\delta^0 L_{BB}(T_\delta)e^{-\delta} d\delta \quad (2.37)$$

Note that the sign convention for the integral follows from the definition of the optical depth. Subtracting the surface radiance contribution results in

$$\tilde{L}(\lambda) = L(\lambda) - \tau(\lambda)L_s(\lambda) = - \int_\delta^0 L_{BB}(T_\delta)e^{-\delta} d\delta \quad (2.38)$$

For cases when the atmosphere is optically thick (i.e., the transmission is effectively zero), the surface radiance term contributes very little to the total radiance reaching the sensor

and $\tilde{L} \approx L$. Otherwise, the effects of the surface emission must be considered. Rewriting the radiative transfer equation in terms of the transmission once again results in

$$\tilde{L}(\lambda) = \int_{\delta}^0 L_{BB}(T_{\delta}) \frac{\partial \tau(\delta, \lambda)}{\partial \delta} d\delta \quad (2.39)$$

Thus, the observed spectral radiance is related to the integral of the vertically distributed Planck function weighted by the derivative of the transmission. The transmission is noted as dependent on optical depth *and* wavelength even though the optical depth includes this wavelength dependence. This is done to separate the effects of altitude and wavelength. Thus, the optical depth refers to the vertical (path) dimension only. Because of its action on the Planck function, the derivative term in this equation is traditionally referred to as the *weighting function* and represented as $K(\lambda, \delta)$. From a linear systems perspective, we see that the resulting equation

$$\tilde{L}(\lambda) = \int_{\delta}^0 L_{BB}(T_{\delta}) K(\lambda, \delta) d\delta \quad (2.40)$$

is a *convolution*. Thus, the resulting spectral radiance is “blurred” by the weighting function. The extent of this blurring depends on the width of the weighting function. This concept will be described in more detail later in this section. Thus, if the weighting function is known, then it is theoretically possible to infer the vertical distribution of temperature in the atmosphere as defined by the Planck function.

Implementation

Certain issues must be considered to implement a realistic atmospheric sounding scheme. These include: assumptions about atmospheric physics, the nature of the inverse problem, and the spectral response of the sensor.

A. Atmospheric Physics. In the previous discussion, it was assumed that the weighting functions were known. Since the weighting functions are the vertical derivative of the spectral transmissions, it follows that the vertical concentration and absorption of the

atmospheric constituents are known. This is true only if a well-mixed gas with known concentration is used as a reference. Furthermore, it is necessary to evaluate the weighting functions over a spectral range where the absorption by the atmosphere is not caused by a combination of constituents. That way, the radiation effects can be isolated to the reference gas only. Although there are some isolated bands for CO_2 and O_3 in the infrared region of the spectrum, there is always some amount of water vapor absorption contribution throughout the infrared spectrum. This problem is complicated by the uncertainty in water vapor concentration. In addition, the exact nature of the water vapor continuum is not completely understood, as was discussed in section 2.1.2. Despite these effects, the 4.3 and 15 μm CO_2 bands have been successfully used for temperature profile retrievals (Houghton 1984).

An implicit assumption is that the atmosphere is in local thermodynamic equilibrium (LTE). This physical state allows the accurate depiction of an atmospheric layer with a homogeneous temperature field. Also, the population of excited molecular energy states follows the Boltzmann distribution when the layer is in LTE (Stephens 1994). As described in Section 2.1.3, this is a key parameter in the description of absorption band models. For the troposphere, LTE is a reasonable assumption because the atmosphere is dense and there are enough molecular collisions to ensure equilibrium. At higher altitudes, the molecular concentration decreases exponentially with respect to pressure depth and the thermodynamic state of the atmosphere must be considered more rigorously. Fortunately, this research focuses on the characterization of the lower atmosphere from which most of the atmospheric radiation originates.

Although the band model dependence on temperature is commonly ignored, the altitude dependence must be taken into account. Section 2.1.3 showed how molecular absorption bands are susceptible to pressure broadening. Because the weighting function depends on the absorption coefficient, it too is affected by pressure broadening. Using the

relationship given by eq. (2.9), the weighting function is

$$\frac{\partial}{\partial z} \tau(\lambda, z) = \beta(z; \lambda) \left[\int_0^z \beta(z; \lambda) dz \right] e^{-\int_0^z \beta(z; \lambda) dz} \quad (2.41)$$

Substituting the Lorentz line shape model in eqs. (2.24) and (2.25) for the absorption coefficient $\beta(z; \lambda)$ and changing the notation to refer to wavelengths instead of frequency results in

$$\beta(z; \lambda) = \frac{S(z)}{\pi} \frac{\alpha_L(z)}{(\lambda - \lambda_0)^2 + \alpha_L(z)^2} \quad (2.42)$$

where the strength and width of the absorption depend explicitly on altitude. Substituting eq. (2.26) for $\alpha_L(z)$ yields

$$\beta(z; \lambda) = \frac{\frac{S(z)}{\pi} \alpha_{L,s} \frac{p(z)}{p_s} \sqrt{\frac{T_s}{T(z)}}}{(\lambda - \lambda_0)^2 + \left[\alpha_{L,s} \frac{p(z)}{p_s} \sqrt{\frac{T_s}{T(z)}} \right]^2} \quad (2.43)$$

A weighting function created using this model is shown in Figure (2.12). The pressure broadening can lead to a sharpening of the weighting functions as shown. The solid line corresponds to the weighting function with respect to altitude. The dotted line is the same weighting function but plotted against $-\ln p(z)$ (with a bias added to it to bring it to the same scale). The temperature can cause either broadening or sharpening of the weighting functions (Goody and Yung 1989).

As shown previously, eq. (2.39) appears to be the convolution of the Planck function and the weighting function. However, this linear relationship arises from the assumption that the Planck function is independent of wavelength and the weighting function is independent of temperature. Neither of these are theoretically true. The Planck function becomes more dependent on wavelength as the atmospheric and surface temperature increases and as the spectral region of interest widens. For example, the Planck curve for 300 °K shown in Figure 2.5 clearly changes with wavelength over the extent of the CO₂ band while the 225 °K is relatively constant. Equation (2.43) clearly demonstrates that the weighting functions

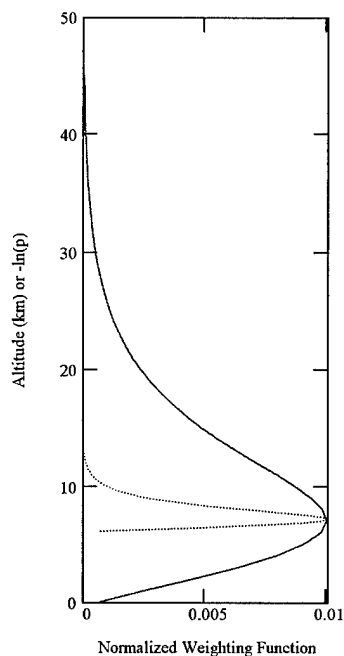


Figure 2.12: Weighting function arising from pressure broadening phenomenology

depend on temperature. Therefore, the Planck and weighting functions have a nonlinear relationship. In general, however, the Planck function varies more rapidly with altitude than with wavelength. Thus, it is often reasonable to disregard its dependence on wavelength. The omission of the weighting function's dependence on temperature, however, may be less appropriate (Goody and Yung 1989).

B. The Inverse Problem. Earlier, I stated that the temperature profile could be inferred from a set of known weighting functions and the measured spectral radiance. This seemingly simple task is complicated by the nature of the inverse problem. In its linear form, eq. (2.39) is a Fredholm integral equation of the first kind. The problem is complicated further by the nature of the weighting function. The weighting function is exponential and resembles the kernel in a Laplace transform. As the optical depth increases, the value of the weighting

function approaches zero. Thus, a large change in the vertical Planck distribution translates to a small change in the observed radiance. This leads to an unstable inverse solution because a small change in the observed radiance must be “amplified” to match a large change in the Planck function. Therefore, small errors can be amplified in the estimate of the temperature vertical profile (Twomey 1996). Also, the weighting function has the effect of “blurring” the Planck function distribution, which has the effect of destroying information. This can be illustrated by considering the limiting case where the weighting function is a Dirac delta function. In that case, the spectral radiance and the Planck function can be matched one-to-one and there is no blurring nor loss of information. As will be seen later, this is not the case and the sharpness of the weighting function will vary. Finally, we are restricted to a finite number of spectral samples. This leads to a finite number of vertical samples in the retrieved profiles. The act of taking a physical problem from the continuous domain to the discrete domain is often called *discretization*. This introduces further complexities to the inverse problem because the discrete profile is an estimate of the true continuous profile and therefore not exact. The result is that the inverse problem is ill-posed because the solutions are not unique. To solve the problem, it is necessary to constrain as many degrees-of-freedom as possible.

Solutions to eq. (2.39) involve the use of linear or nonlinear methods. These methods can also be either statistical or physical. The statistical approach assumes that enough *a priori* information about the statistics of the atmosphere are known to constrain the solution. Physical approaches are based entirely on the physics of the radiative transfer equation and are often constrained by the judicious use of an initial profile and by arriving to a solution iteratively (Goody and Yung 1989; Menzel and Gumley 1999). The goal of all of these approaches is to make the problem “well-posed”. The act of turning the ill-posed problem into a well-posed or solvable problem is often called *regularization*. The rest of this section will illustrate some regularization schemes. For now, it is sufficient to know that there are optimal and suboptimal algorithms used to implement linear and nonlinear

methods. The appropriate choice of implementation is largely governed by the application and available resources. The interested reader is referred to the large literature on numerical methods. A good starting point is *Numerical Recipes* (Press et al. 1992).

i. A Linear Approach. Equation (2.39) can be approximated as linear since the Planck function varies more with temperature (and indirectly with altitude) than with wavelength. The implications of this assumption will be illustrated shortly. The transmission can also be assumed to be independent of temperature, although this approximation is less accurate than the previous (Goody and Yung 1989). Finally, realizing that only discrete values of spectral radiance measurements are available, equation (2.39) can be rewritten as

$$\tilde{L}_i = \int_{\delta}^0 L_{BB_i}(T_{\delta}) K_i(\delta) d\delta \quad (2.44)$$

where i denotes the discrete spectral band. Using the approximation that the Planck function does not depend on wavelength, let $\tilde{L}_{BB} = L_{BB_i}$. Given this assumption, the altitude-dependent Planck function can be expressed as a sum of basis functions such that

$$\tilde{L}_{BB}(\delta) = \sum_{j=1}^M x_j f_j(\delta) \quad (2.45)$$

where $f_j(\delta)$ are the basis functions, x_j are unknown coefficients, and M defines the number of vertical layers. Conceptually, the simplest basis function is the **rect** function. In theory, any basis function can be used. In practice, however, the basis functions are chosen to satisfy some optimality criterion. Substituting eq. (2.45) into eq. (2.44) results in

$$\tilde{L}_i = \sum_{j=1}^M a_{ij} x_j \quad (2.46)$$

where

$$a_{ij} = \int_{\delta}^0 K_i(\delta) f_j(\delta) d\delta \quad (2.47)$$

This is the linear form of eq. (2.39). This equation can be represented in matrix notation as

$$\tilde{\mathbf{L}} = \mathbf{A} \mathbf{x} \quad (2.48)$$

where a_{ij} are the coefficients of the matrix $\mathbf{A}$, and $\mathbf{x}$ is a column vector. This can be expanded to

$$\begin{aligned}\tilde{L}_1 &= a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1M}x_M \\ \tilde{L}_2 &= a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \cdots + a_{2M}x_M \\ &\vdots \\ \tilde{L}_N &= a_{N1}x_1 + a_{N2}x_2 + a_{N3}x_3 + \cdots + a_{NM}x_M\end{aligned}\tag{2.49}$$

This expansion illustrates that the inverse problem can be reduced to the solution of a simultaneous set of equations. If the number of vertical layers and the number of spectral measurements are equal, and the N equations are linearly independent, then it is possible to solve for the coefficients x_j with a simple technique like Gauss-Jordan elimination. The x_j coefficients can then be used with eq. (2.45) to estimate the vertical profile. In matrix notation, the solution is

$$\mathbf{x} = \mathbf{A}^{-1}\tilde{\mathbf{L}}\tag{2.50}$$

or

$$x_j = \sum_{i=1}^N a_{ij}^{-1} \tilde{L}_i\tag{2.51}$$

where a_{ij}^{-1} are the coefficients of the matrix $\mathbf{A}^{-1}$. Substituting these values into eq. (2.45) results in

$$\tilde{L}_{BB}(\delta) = \sum_{j=1}^M \left[\sum_{i=1}^N a_{ij}^{-1} \tilde{L}_i \right] f_j(\delta)\tag{2.52}$$

Using the linear property of the summation operation

$$\tilde{L}_{BB}(\delta) = \sum_{i=1}^N \left[\sum_{j=1}^M a_{ij}^{-1} f_j(\delta) \right] \tilde{L}_i = \sum_{i=1}^N D_i(\delta) \tilde{L}_i\tag{2.53}$$

or in matrix notation

$$\tilde{L}_{BB}(\delta) = \mathbf{f}'(\delta)\mathbf{x} = \mathbf{f}'(\delta)\mathbf{A}^{-1}\mathbf{L} = \mathbf{D}(\delta)\mathbf{L}\tag{2.54}$$

The function $D_i(\delta)$ is commonly referred to as the *contribution function* since it defines the contribution of the spectral radiance measurement to the vertical distribution of the Planck

function. The contribution function also defines how much of the error in the spectral measurement contributes to the error in the estimate of the vertical profile of the Planck function. Assuming that the noise variances of each spectral measurement are equal and that they are statistically independent, the noise variance in the estimate of the Planck function is

$$\sigma_{BB}^2(\delta) = \sigma^2 \sum_{i=1}^N D_i^2(\delta) = \sigma^2 D^2(\delta) \quad (2.55)$$

The term $D^2(\delta)$ is called the *noise amplification factor*. The magnitude of this term is an indication of the conditioning of the matrix $\mathbf{A}$ (Goody and Yung 1989).

The approach that has just been outlined is often called the *direct* linear approach because the vertical profile is derived from a direct and linear operation. So far, we have assumed that the inverse of the matrix $\mathbf{A}$ exists. In practice, however, there are often less spectral measurements than desired vertical levels. This leads to an *underdetermined* system of equations and the matrix $\mathbf{A}$ is said to be *rank-deficient*. Even if the number of measurements is equal to the number of vertical levels, the radiance measurements are typically correlated so that not all of the rows of the matrix $\mathbf{A}$ are linearly independent. In this case, the matrix is *ill-conditioned*. One approach is to design a spectrograph so that there are more spectral radiance measurements than the number of desired vertical levels. This results in an *overdetermined* system of equations. All of these cases result in a matrix $\mathbf{A}$ that cannot be inverted directly.

One approach is to use the Moore-Penrose pseudoinverse such that

$$\mathbf{A}^{-1} \approx (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}' = \mathbf{A}^\dagger \quad (2.56)$$

This satisfies the requirement of the inverse of a matrix since $\mathbf{A}^\dagger\mathbf{A} = \mathbf{I}$ where $\mathbf{I}$ is the identity matrix. Thus, the profile coefficients can be estimated by

$$\hat{\mathbf{x}} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\tilde{\mathbf{L}} \quad (2.57)$$

For an underdetermined matrix $\mathbf{A}$, the matrix $(\mathbf{A}'\mathbf{A})$ cannot be inverted directly because it has zero eigenvalues (and eigenvalues less than one actually become smaller!). However,

the inversion may still be possible if a technique such as the spectral decomposition of $\mathbf{A}'\mathbf{A}$ or the singular value decomposition (SVD) of $\mathbf{A}$ is used (see Appendix B). For an overdetermined matrix, this technique performs the inverse computation exactly, provided that there are no errors in the measured vector $\tilde{\mathbf{L}}$. In reality, measurements always have noise and $\hat{\mathbf{x}}$ is equal to the solution that minimizes

$$\left| \mathbf{A}\hat{\mathbf{x}} - \tilde{\mathbf{L}} \right|^2 \quad (2.58)$$

This is the *least-squares* solution because it minimizes the squared error in the estimated radiance. This regression is optimal in the sense that it minimizes the error in the spectral radiance. This does not imply that the solution optimally estimates $\mathbf{x}$. In fact, it is possible to obtain a better estimate of $\mathbf{x}$ if the least-squares criterion is appropriately constrained (Twomey 1996). One method is the Twomey-Tikhonov regularization which modifies eq. (2.57) to

$$\hat{\mathbf{x}} = [(\mathbf{A}'\mathbf{A}) + \gamma\mathbf{H}]^{-1}\mathbf{A}'\tilde{\mathbf{L}} \quad (2.59)$$

where γ is a constant and $\mathbf{H}$ is chosen so as to select the “best” solution of $\mathbf{x}$ given a specified constraint. One example of a constraint would be to select the smoothest solution as the most probable solution. Equation (2.59) arises from the minimization of

$$\left| \mathbf{A}\hat{\mathbf{x}} - \tilde{\mathbf{L}} \right|^2 + \gamma q(\hat{\mathbf{x}}) \quad (2.60)$$

where $\left| \mathbf{A}\hat{\mathbf{x}} - \tilde{\mathbf{L}} \right|^2$ corresponds to the least-squares criterion and $q(\hat{\mathbf{x}})$ is a measure of the smoothness of the solution $\hat{\mathbf{x}}$. For example, this measure may be obtained from the first or second “differences” (i.e., the discrete derivatives) of $\hat{\mathbf{x}}$ or from the variance of $\hat{\mathbf{x}}$. In the limit $\gamma \rightarrow \infty$, the solution would be based on the smoothness criterion only and would be independent of any measured values of $\tilde{\mathbf{L}}$. Alternatively, in the limit $\gamma \rightarrow 0$, the solution is based solely on the least-squares criterion. Thus, the γ factor allows “fine-tuning” of the solution based on the noise characteristics of $\tilde{\mathbf{L}}$ and the conditioning of $\mathbf{A}$.

A fundamental difficulty with this procedure is that γ will vary depending on the problem and must be determined *a priori*. This is due to the fact that we do not know

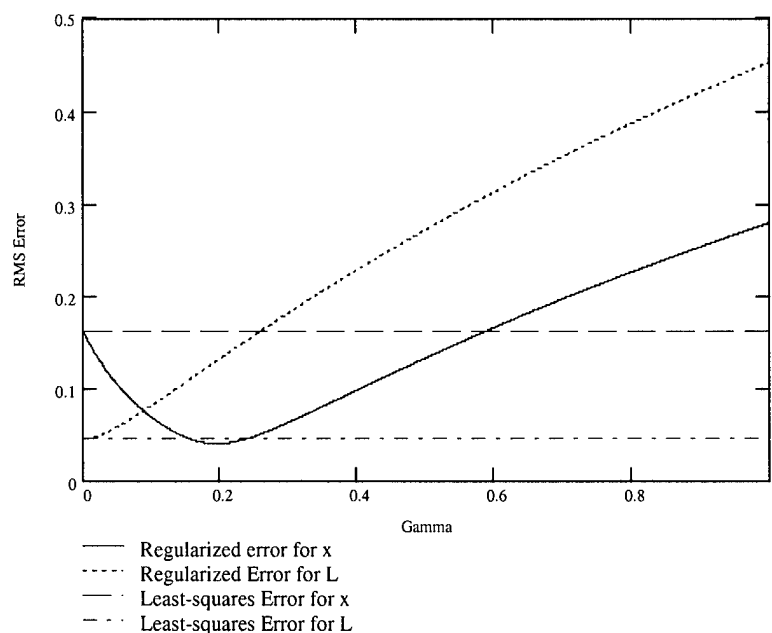


Figure 2.13: Effect of γ on residual error of estimates obtained from the regularized inverse.

the true vector $\mathbf{x}$ and therefore cannot compute the residual error $|\mathbf{x} - \hat{\mathbf{x}}|^2$. One solution is to compute the residual error for known values of $\mathbf{x}$ as a function of γ and find the γ that minimizes the error. Figure 2.13 shows hypothetical *regularized* root-mean-square (RMS) residual errors between the observed $\tilde{\mathbf{L}}$ and the estimated $\tilde{\mathbf{L}}$ and between the true $\mathbf{x}$ and $\hat{\mathbf{x}}$ as a function of γ . The problem was regularized by letting $q(\hat{\mathbf{x}}) = \text{var}(\hat{\mathbf{x}})$ (i.e., the constraint was to choose the solution with the minimum variance). These RMS curves are compared to the constant least-squares regression RMS for $\tilde{\mathbf{L}}$ and $\mathbf{x}$. When $\gamma = 0$, the regularized solution is equal to the least-squares solution. The plot indicates that $\gamma = 0.2$ is optimal for the estimation of $\mathbf{x}$. It also illustrates that the optimal value of γ cannot be determined from an analysis of the errors in $\tilde{\mathbf{L}}$. How well the estimated γ value does when it is applied to new observations depends on how well the measurement noise was characterized. Also, the optimal γ value may depend on the shape of the true $\mathbf{x}$.

One final comment should be made about the Twomey-Tikhonov regularization approach. The more ill-conditioned $\mathbf{A}$ is, the larger the γ value will have to be in order to

make the inverse solution stable. This means that less information in $\tilde{\mathbf{L}}$ is used to estimate $\mathbf{x}$. In certain cases, particularly when $\mathbf{A}$ is rank-deficient, another approach may be more suitable. For instance, a calculation of $(\mathbf{A}'\mathbf{A})^{-1}$ with a reconstruction based on a truncated SVD may lead to significantly better results. Another alternative is to use the SVD in the calculation of the Twomey-Tikhonov inverse.

The least-squares, Twomey-Tikhonov, and SVD solutions can result in lower noise amplification factors than the direct linear solution. In addition to implementing these techniques, the noise amplification factor may be reduced by the proper selection of the basis functions $f_j(\delta)$ used to represent the Planck vertical profile. According to Goody and Yung (1989), the noise amplification factor is minimized when the weighting functions are used as the set of basis functions, assuming the noise in the observations is identical and independently distributed (iid). Consequently, this set of basis functions is planned for implementation in the sounding algorithms for MODIS (Menzel and Gumley 1999).

ii. A Statistical Approach. In the previous section, a linear approach for solving equation (2.48) was introduced. This approach assumed that the radiative transfer was a linear transformation of a vector $\mathbf{x}$ to the observed value $\tilde{\mathbf{L}}$. To solve the inverse problem, it was found that a constrained least-squares solution might be able to handle errors in the measurement as well as ill-conditioning of the matrix $\mathbf{A}$. The choice of a constraint to the least-squares solution is arbitrary, and it was shown that using a minimum variance constraint can lead to improvements in the retrieval of $\mathbf{x}$. There was, however, no particular reason based on the *physics* of the problem for establishing this constraint; rather, it originated from the desire to avoid numerical instabilities in the solution due to measurement errors. It may be more appropriate to choose a constraint based on *a priori* knowledge of the atmosphere. One simple technique uses a climatological mean as a constraint in the Twomey-Tikhonov regularization method (Houghton 1984). Thus, the solution is based on a departure from the mean.

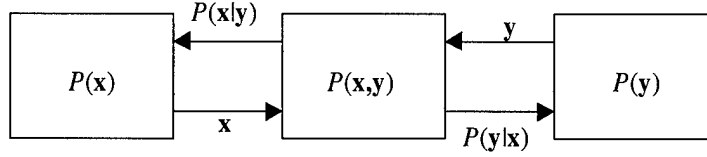


Figure 2.14: Mapping of state space to measurement space.

A more general use of statistics involves an understanding of the multivariate nature of the problem. Figure 2.14 shows a simplistic schematic of the problem at hand. The radiative transfer through the atmosphere represents a mechanism for *mapping* the thermodynamic state and character of the atmosphere to a set of spectral observations. The vector $\mathbf{x}$ is a multivariate random variable of size M , originating from a multivariate probability distribution $P(\mathbf{x})$. This is the *state* space and the vector $\mathbf{x}$ is a *source* vector. Similarly, the observation $\mathbf{y} = \tilde{\mathbf{L}}$ is a random variable in the *measurement* space having a multivariate probability distribution $P(\mathbf{y})$. The two are related by a joint multivariate probability distribution $P(\mathbf{x}, \mathbf{y})$. Bayes' rule provides the structure through which these distributions are related:

$$P(\mathbf{x})P(\mathbf{y}|\mathbf{x}) = P(\mathbf{x}, \mathbf{y}) = P(\mathbf{y})P(\mathbf{x}|\mathbf{y}) \quad (2.61)$$

where $P(\mathbf{x}|\mathbf{y})$ and $P(\mathbf{y}|\mathbf{x})$ are conditional probabilities that describe *a posteriori* knowledge of the variables $\mathbf{x}$ and $\mathbf{y}$. Thus, knowledge of the “source” distribution $P(\mathbf{x})$ and the joint distribution $P(\mathbf{x}, \mathbf{y})$ is all that is needed to describe the mapping. These values can also determine the amount of *information* that a measurement $\mathbf{y}$ carries about $\mathbf{x}$. That is, an optimal mapping from the measurement to the state space exists.

When a new observation $\mathbf{y}$ resulting from a state vector $\mathbf{x}$ is made, the *a posteriori* probability $P(\mathbf{x}|\mathbf{y})$ changes. One method for determining $\mathbf{x}$ from $\mathbf{y}$ is to use the *Maximum Likelihood* rule, which finds $\mathbf{x}$ such that $P(\mathbf{x}|\mathbf{y})$ is maximized. This is an intuitive approach and provides an optimality criterion. When the source and joint probability distributions

are symmetric, the maximum likelihood solution is the same as the expected value given by

$$\hat{\mathbf{x}} = E(\mathbf{x}|\mathbf{y}) = \int \mathbf{x} P(\mathbf{x}|\mathbf{y}) d\mathbf{x} \quad (2.62)$$

where $\hat{\mathbf{x}}$ is the most likely (and expected) estimate of the true $\mathbf{x}$ given the observation $\mathbf{y}$ (Rodgers 1998). The uncertainty in $\hat{\mathbf{x}}$ is obtained from the covariance matrix $\Sigma_{\hat{\mathbf{x}}}$, which is the expected value of the squared deviations from $\hat{\mathbf{x}}$:

$$\Sigma_{\hat{\mathbf{x}}} = E[(\mathbf{x} - \hat{\mathbf{x}})^2] = \int (\mathbf{x} - \hat{\mathbf{x}})^2 P(\mathbf{x}|\mathbf{y}) d\mathbf{x} \quad (2.63)$$

An analytical expression of $\hat{\mathbf{x}}$ can be obtained by assuming that the *a priori* distribution $P(\mathbf{x})$ is multivariate Gaussian:

$$P(\mathbf{x}) = \frac{1}{(2\pi)^{N/2} |\Sigma_{\mathbf{x}}|^{1/2}} \exp \left[-\frac{1}{2} (\mathbf{x} - \mu_{\mathbf{x}})' \Sigma_{\mathbf{x}}^{-1} (\mathbf{x} - \mu_{\mathbf{x}}) \right] \quad (2.64)$$

where $\mu_{\mathbf{x}}$ and $\Sigma_{\mathbf{x}}$ are *a priori* maximum likelihood estimates. The *a posteriori* probability for $\mathbf{y}$ is also assumed to be Gaussian distributed. This can be expressed in terms of the radiative transfer mapping established in equation (2.48) perturbed by some random error ϵ so that

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \epsilon \quad (2.65)$$

and

$$\ln P(\mathbf{y}|\mathbf{x}) = -\frac{1}{2} (\mathbf{y} - \mathbf{A}\mathbf{x})' \Sigma_{\epsilon}^{-1} (\mathbf{y} - \mathbf{A}\mathbf{x}) + \ln [(2\pi)^{N/2} |\Sigma_{\epsilon}|^{1/2}] \quad (2.66)$$

The second term on the right hand side of the equation is a normalization constant and does not provide any information about the how the state and measurement spaces are related; therefore it is omitted from the subsequent analysis. Solving equation (2.61) for $P(\mathbf{x}|\mathbf{y})$ and taking the natural logarithm results in

$$\ln P(\mathbf{x}|\mathbf{y}) = \ln P(\mathbf{y}|\mathbf{x}) + \ln P(\mathbf{x}) - \ln P(\mathbf{y}) \quad (2.67)$$

Because the observation $\mathbf{y}$ has been made, $P(\mathbf{y}) = 1$ and $\ln P(\mathbf{y}) = 0$. Substituting eqs. (2.64) and (2.66) into eq. (2.67) results in

$$\ln P(\mathbf{x}|\mathbf{y}) = -\frac{1}{2} (\mathbf{y} - \mathbf{A}\mathbf{x})' \Sigma_{\epsilon}^{-1} (\mathbf{y} - \mathbf{A}\mathbf{x}) - \frac{1}{2} (\mathbf{x} - \mu_{\mathbf{x}})' \Sigma_{\mathbf{x}}^{-1} (\mathbf{x} - \mu_{\mathbf{x}}) \quad (2.68)$$

This expression has the general quadratic form

$$-\frac{1}{2}(\mathbf{x} - \hat{\mathbf{x}})' \Sigma_{\hat{\mathbf{x}}}^{-1} (\mathbf{x} - \hat{\mathbf{x}}) \quad (2.69)$$

Thus, the *a posteriori* probability of $\mathbf{x}$ is Gaussian distributed because it is the result of a linear mapping of Gaussian-distributed random variables. Therefore, $\hat{\mathbf{x}}$ is the maximum likelihood estimate of $\mathbf{x}$. Rodgers (1998) shows that $\hat{\mathbf{x}}$ can be solved for in terms of the observation, weighting functions, and a priori statistics by equating the quadratic and linear terms in eqs. (2.68) and (2.69) such that

$$\hat{\mathbf{x}} = (\mathbf{A}' \Sigma_{\epsilon}^{-1} \mathbf{A} + \Sigma_{\mathbf{x}}^{-1})^{-1} (\mathbf{A}' \Sigma_{\epsilon}^{-1} \mathbf{y} + \Sigma_{\mathbf{x}}^{-1} \mu_{\mathbf{x}}) \quad (2.70)$$

This seemingly complex expression has a relatively simple interpretation. The first term $\mathbf{A}' \Sigma_{\epsilon}^{-1} \mathbf{A} + \Sigma_{\mathbf{x}}^{-1}$ is simply the sum of the variances due to the error in the observed value $\mathbf{y}$ and $\mu_{\mathbf{x}}$. This overall variance scales the second term, which is nothing more than the sum of the two random variables: the contribution to the estimate of $\mathbf{x}$ from $\mathbf{y}$ and $\mu_{\mathbf{x}}$ scaled by their appropriate variances. This result should be an improvement over other linear approaches because the multivariate statistics of the problem are used to obtain the most probable solution. The main drawback is that the form of the probability distributions are assumed to be Gaussian, which may not necessarily be a correct assumption.

Statistics may also be used to minimize the error made by selecting a particular $\mathbf{x}$ as $\hat{\mathbf{x}}$. This is done by constraining the solution to have minimum variance (i.e., squared error between the true and estimated $\mathbf{x}$). The variance of the solution is the maximum likelihood estimate $\Sigma_{\hat{\mathbf{x}}}$. This term can be minimized by setting the derivative with respect to $\hat{\mathbf{x}}$ equal to zero. When the joint probability distribution is symmetric, the estimate of $\mathbf{x}$ that leads to minimum variance is equal to the maximum likelihood estimate of $\mathbf{x}$ obtained with eq. (2.62).

The minimum variance concept may also be used to relate a set of measured $\mathbf{y}$ to a set of atmospheric states $\mathbf{x}$ empirically. This technique is particularly useful because it makes no assumptions about the shape of the probability distributions *and* does not require

rigorous characterization of the sensor and the atmosphere. Thus, the matrix $\mathbf{A}$ does not need to be created. This method, however, does require a large ensemble of representative data to exist. The ensemble may be generated from previous observations or simulations. The *a priori* expected values of $\mathbf{x}$ and $\mathbf{y}$ are estimated by calculating the means of the $\mathbf{x}$ and $\mathbf{y}$ ensembles. The resulting mean vectors $\hat{\mu}_{\mathbf{x}}$ and $\hat{\mu}_{\mathbf{y}}$ are averages over all of the observations in the ensembles. The method assumes that $\mathbf{x}$ and $\mathbf{y}$ may be related linearly by

$$\mathbf{x} - \mu_{\mathbf{x}} = \mathbf{D}(\mathbf{y} - \mu_{\mathbf{y}}) \quad (2.71)$$

where $\mathbf{x} - \mu_{\mathbf{x}}$ is a $q \times 1$ matrix, $\mathbf{D}$ is a $q \times p$ matrix, and $\mathbf{y} - \mu_{\mathbf{y}}$ is a $p \times 1$ matrix. If there are n observations in the ensemble, then the covariance between the two sets is

$$\Sigma_{\mathbf{xy}} = \frac{1}{n} \sum_{k=1}^n (\mathbf{x} - \mu_{\mathbf{x}})(\mathbf{y} - \mu_{\mathbf{y}})' \quad (2.72)$$

Assuming that the error in the measurement is additive, the covariance of the $\mathbf{y}$ observations is given by

$$\hat{\Sigma}_{\mathbf{yy}} = \Sigma_{\mathbf{yy}} + \Sigma_{\epsilon} \quad (2.73)$$

The matrix $\mathbf{D}$ that yields the minimum variation of $\mathbf{x}$ is

$$\mathbf{D} = \Sigma_{\mathbf{xy}} \hat{\Sigma}_{\mathbf{yy}}^{-1} = \mathbf{X}'\mathbf{Y}(\mathbf{Y}'\mathbf{Y})^{-1} \quad (2.74)$$

where $\mathbf{X}$ and $\mathbf{Y}$ are the $n \times q$ and $n \times p$ ensembles, respectively, and are mean-centered and scaled by the number of observations. Thus, $\mathbf{D}$ is the solution to the least-squares criterion. This may be more apparent by defining the coefficients in terms of $\mathbf{D}'$ such that

$$\beta = \mathbf{D}' = (\mathbf{Y}'\mathbf{Y})^{-1}\mathbf{Y}'\mathbf{X} \quad (2.75)$$

and

$$\hat{\mathbf{X}} = \mathbf{Y}\beta \quad (2.76)$$

The difference between this scheme and the one described in the context of direct linear solutions is that the regression is defined to predict $\mathbf{x}$ instead of $\tilde{\mathbf{L}}$.

There are no constraints to the relative dimensionality of the vectors $\mathbf{x}$ and $\mathbf{y}$. That is, p does not have to be equal to q . It might then be tempting to make $q \gg p$. However, this leads to ill-conditioning or rank-deficiency in the matrix β , which raises the practical issues associated with inverse problems discussed earlier in this section.

iii. A Nonlinear Approach. Nonlinear solutions to eq. (2.39) are typically iterative. The main advantage of this approach is that it does not introduce errors associated with linear radiative transfer assumptions. Furthermore, *a priori* knowledge of the statistics of the atmosphere is not required. Chahine's relaxation method is a good example of a nonlinear iterative solution (Goody and Yung 1989). It tends to converge quickly and is more likely to yield correct solutions given an appropriate set of weighting functions exists. The algorithm is based on approximating eq. (2.39) by noting that a typical weighting function has a maximum at some δ_i . Because the weighting function dominates at this value, equation (2.39) can be rewritten as

$$\tilde{L}_i \approx CK_i(\delta_i)L_{BB_i}(T_{\delta_i}) \quad (2.77)$$

where $\tilde{L}_i$ is the *fixed* observed radiance and C is an unknown constant that must be determined from empirical results. If an estimated temperature distribution profile is used to initialize the algorithm, the resulting estimated radiance field is given by

$$\tilde{L}_i^{(0)} \approx CK_i(\delta_i)L_{BB_i}(T_{\delta_i}^{(0)}) \quad (2.78)$$

where the superscript (0) represents the zeroth iteration. The vertical temperature profile of the first iteration is determined by the following ratio

$$\frac{L_{BB_i}(T_{\delta_i}^{(1)})}{L_{BB_i}(T_{\delta_i}^{(0)})} = \frac{\tilde{L}_i}{\tilde{L}_i^{(0)}} \quad (2.79)$$

The temperature profile obtained from the first iteration is an improved estimate that is used in the next iteration. The algorithm converges when the change between the temperature profiles between two iteration steps is less than some specified threshold (typically of the

same magnitude as the sensor noise). Chahine's method strongly depends on the sharpness of the weighting functions. Therefore, convergence is not always guaranteed (Goody and Yung 1989).

The method described here is also known as *nonlinear relaxation*. Note that the "non-linear" operation described so far is the determination of a new atmospheric profile after each iteration. In order to account for the nonlinearities in the *radiative transfer*, new weighting functions must be generated using the new temperature profile estimate at each iteration step. The same is true for any nonlinear sounding method. For example, a nonlinear sounding method may consist of simply solving the linear inverse problem with some initial weighting functions, using the solutions to build new weighting functions, and solving the "new" linear inverse problem at the next iteration step. This would be done several times until convergence. As with any nonlinear technique, there must be a suitable set of constraints in order to obtain an adequate solution.

C. Sensor Response. So far, the development of the theory of atmospheric sounding assumed monochromatic radiation. Since realistic measurements are made with a detector that has a finite-width spectral response, it is necessary to use band transmission functions. Thus, equation (2.39) must be rewritten:

$$\tilde{L}_{\bar{\lambda}} = \int_0^{\delta'} L_{BB}(T_{\delta}) \frac{\partial \mathcal{T}_{\bar{\lambda}}(\delta)}{\partial \delta} d\delta \quad (2.80)$$

where the subscript $\bar{\lambda}$ denotes the spectral bandpass of the detector. For a detector with many channels, this spectral bandpass would be specified for each channel. This averaging of the transmission function broadens the calculated weighting function. Thus, the spectral response of the system will determine how absorption features will contribute to the observed radiance. This may actually be a desirable condition when a broad temperature sounding band (e.g., CO₂ 15 μm band) is overlapped with narrow absorption features due to other atmospheric constituents (e.g., water vapor). On the other hand, we have seen that the

sharpness of the weighting function may lead to a more numerically stable solution of the inverse problem. A prudent system design analysis should consider this tradeoff.

Water Vapor Profiles and Other Constituents

The determination of the concentration of water vapor and other constituents as a function of altitude is more complicated than obtaining just the temperature profile using known weighting functions. The total transmittance through the atmosphere is the product of the individual transmission contributions of each constituent and is represented by

$$\tau = \prod_i \tau_i \quad (2.81)$$

where τ_i is the transmission of the i^{th} constituent. In concept, we should be able to build weighting functions and infer an *effective* temperature profile for each constituent. This effective temperature profile can then be used to determine the constituent concentration profile. This logic fails, of course, because we cannot build weighting functions unless we know the constituent concentration profiles. However, we can use *initial estimates* of the weighting functions and use these to solve for an effective *change* or *perturbation* about the initial estimate of temperature. To do so requires a modification to eq. (2.39) to include the contributions from each constituent of interest (Huang 1989).

The modified radiative transfer equation results in a perturbed radiance observation about some initial estimate. The goal then is to relate this perturbation to a perturbation of the temperature and constituent profiles using “known” initial weighting function estimates and profiles. This equation is then used to solve for the temperature perturbation for all the constituents of interest. The solutions are *effective* temperature profiles for each constituent, which can then be related to the absorber concentration based on the hydrostatic equation. This results in a *simultaneous* retrieval of temperature and concentration profiles. A rigorous mathematical development is provided in Appendix C and follows that of Smith et al. (1991) with some minor additions and changes.

This method was implemented for the analysis of Visible Infrared Spin Scan Radiometer (VISSR) Atmospheric Sounder (VAS) data. This sensor is onboard the operational Geostationary Operational Environmental Satellites (GOES) (Hayden 1998). The same technique is used for the TIROS Operational Vertical Sounder (TOVS) (Houghton 1984). The technique will also be implemented with the planned Advanced Infrared Radiation Sounder (AIRS), which has an unprecedented high spectral resolution (Smith et al. 1991). Proper implementation requires the use of a line-by-line radiative transfer code for the Beer-Lambert Law and eq. (2.81) to hold. Smith et al. demonstrated that temperature retrieval errors for the AIRS sensor are expected to be around 1.5 °K RMS. Water vapor mixing ratio profiles have RMS errors between 0.2 and 0.5 g/kg and ozone retrievals have errors of about 3%.

2.1.6 The In-Scene Atmospheric Compensation Method

The ISAC algorithm was developed by Hackwell, Johnson, and Young of the Aerospace Corporation for the analysis of hyperspectral imagery from the Spatially-Enhanced Broadband Array Spectrograph System (SEBASS) (Johnson and Young 1998). The method makes use of eq. (2.5) and image statistics to define unscaled spectral transmission and upwelled radiance curves. The use of the image statistics assumes that there is no variation of the atmospheric parameters over the image. These curves are then scaled to “true” values by assuming a known value at a reference wavelength or by scaling to output spectral radiance curves from MODTRAN.

Unscaled Parameters

Consider the case of a blackbody imaged by a remote sensor. From eq. (2.5) we find that there is no contribution from the reflected downwelled radiance because $\varepsilon = 1$. Thus, the surface-leaving radiance is solely defined by the Planck function and the equation of transfer

greatly simplifies to

$$L = \tau L_{BB}(T_s) + L_u \quad (2.82)$$

where T_s is the surface temperature. If the transmission and upwelled radiance are estimated, then the temperature of the blackbody can be determined by rearranging the previous equation, substituting in the Planck function, and solving for T_s :

$$T_s = c_2 \left[\lambda \left\{ \ln \left[\frac{c_1}{\lambda^5 \left(\frac{L - L_u}{\tau} \right)} \right] + 1 \right\} \right]^{-1} \quad (2.83)$$

where T_s is the temperature of the blackbody. For targets that are approximately blackbodies in the far infrared (such as water), the surface temperature retrieval problem is solved at this point. For targets that are not blackbodies, an approximation is necessary.

First, let the *brightness temperature* be the temperature that an object *appears* to have if it is assumed to be a blackbody. For this reason, the brightness temperature is also known as the *apparent* temperature. The brightness temperature can also be thought of as the temperature necessary for the Planck equation to generate a radiance equal to that observed by the sensor. The brightness temperature at the top of the atmosphere can be obtained by rewriting eq. (2.83):

$$T'(\lambda) = c_2 \left[\lambda \left\{ \ln \left[\frac{c_1}{\lambda^5 L(\lambda)} \right] + 1 \right\} \right]^{-1} \quad (2.84)$$

where $T'(\lambda)$ is the “at-sensor” brightness temperature and $L(\lambda)$ is the observed spectral radiance. The radiance reaching the sensor can then be described as

$$L(\lambda) = L_{BB}(\lambda, T') \quad (2.85)$$

Substitution into eq. (2.82) yields

$$L_{BB}(\lambda, T') = \tau(\lambda) L_{BB}(\lambda, T_s) + L_u(\lambda) \quad (2.86)$$

This relationship is exact for blackbodies. Unfortunately, objects in an infrared image are not always blackbodies. However, most natural objects have fairly high emissivities in the

LWIR so that it can be reasonably assumed that there will be several pixels corresponding to blackbodies at some reference wavelength λ_r , in any given image. An appropriate reference can be determined by finding the wavelength at which the brightness temperature is a maximum for the largest number of pixels in the image. The radiation at these pixels is then assumed to be from blackbody emitters. This approach is known as the “maximum hit” criterion for finding blackbody targets in a hyperspectral image. Now assume that k pixels are selected by this procedure. Equation (2.86) then becomes

$$L_{BB}(\lambda_r, T'_k) = \tau(\lambda_r) L_{BB}(\lambda_r, T_k) + L_u(\lambda_r) \quad (2.87)$$

This relationship is valid within the error resulting from estimating the blackbody pixels. However, there will be a bias $\epsilon(\lambda)$ introduced at wavelengths other than the reference due to emissivity effects. Thus, an approximation of equation (2.87) for all wavelengths can be expressed as

$$L_{BB}(\lambda, T'_k) = \tau(\lambda_r) L_{BB}(\lambda, T_k) + L_u(\lambda_r) + \epsilon(\lambda) \quad (2.88)$$

where the transmission and upwelled radiance values are evaluated at the reference wavelength only. Solving for the emitted surface radiation we get

$$L_{BB}(\lambda, T_k) = \frac{1}{\tau(\lambda_r)} [L_{BB}(\lambda, T'_k) - L_u(\lambda_r) - \epsilon(\lambda)] \quad (2.89)$$

This equation is an estimate of the true surface-leaving radiance. Thus, it is appropriate to evaluate $\tau(\lambda)$ and $L_u(\lambda)$ only at λ_r where the atmosphere is assumed to be least influential. This assumption stems from the fact that the largest number of pixels have a maximum brightness temperature at λ_r , implying that the atmosphere is most transmissive at λ_r . An implicit assumption is that the temperature of the Earth’s surface dominates the detected radiance. While this condition is not unrealistic, it can be easily violated. For instance, a large cloud area may dominate the scene and skew the reference wavelength.

Substituting eq. (2.89) into eq. (2.86), and using the relationship in eq. (2.85), results in

$$\begin{aligned}
L(\lambda, k) &= \tau(\lambda) \left[\frac{1}{\tau(\lambda_r)} [L_{BB}(\lambda, T'_k) - L_u(\lambda_r) - \epsilon(\lambda)] \right] + L_u(\lambda) \\
&= \frac{\tau(\lambda)}{\tau(\lambda_r)} L_{BB}(\lambda, T'_k) - \frac{\tau(\lambda)}{\tau(\lambda_r)} [L_u(\lambda_r) + \epsilon(\lambda)] + L_u(\lambda) \\
&= \tau'(\lambda) L_{BB}(\lambda, T'_k) + L'_u(\lambda)
\end{aligned} \tag{2.90}$$

where

$$\tau'(\lambda) = \frac{\tau(\lambda)}{\tau(\lambda_r)} \tag{2.91}$$

and

$$L'_u(\lambda) = \frac{\tau(\lambda)}{\tau(\lambda_r)} [L_u(\lambda_r) + \epsilon(\lambda)] + L_u(\lambda) \tag{2.92}$$

Equation (2.90) is similar to the simple radiative transfer equation for blackbody targets. The main difference is that the transmission and upwelled radiance values are now unscaled or biased by emissivity effects and residual errors in the procedure.

At this point, it is necessary to emphasize we have been using the maximum brightness temperature at the reference wavelength as the best estimate of the surface brightness temperature. This value changes for each pixel but it is constant *with respect to wavelength*. Consider the relationship between the observed radiance and the surface-leaving radiance at some arbitrary wavelength λ_n . If the atmosphere was perfectly transmissive, then there would be a straight “one-to-one” mapping of the surface-leaving radiance and the radiance measured at the sensor. (In fact, this is the case for λ_r because of the way the relationship was defined.) Any deviations from this mapping are due to atmospheric effects. Thus, the unscaled parameters may be obtained from a scatter plot of the observed radiance and the estimated surface-leaving radiance values. The slope and intercept are then $\tau'(\lambda_n)$ and $L'_u(\lambda_n)$ respectively (Figure 2.15). To find the complete spectrum of unscaled transmission

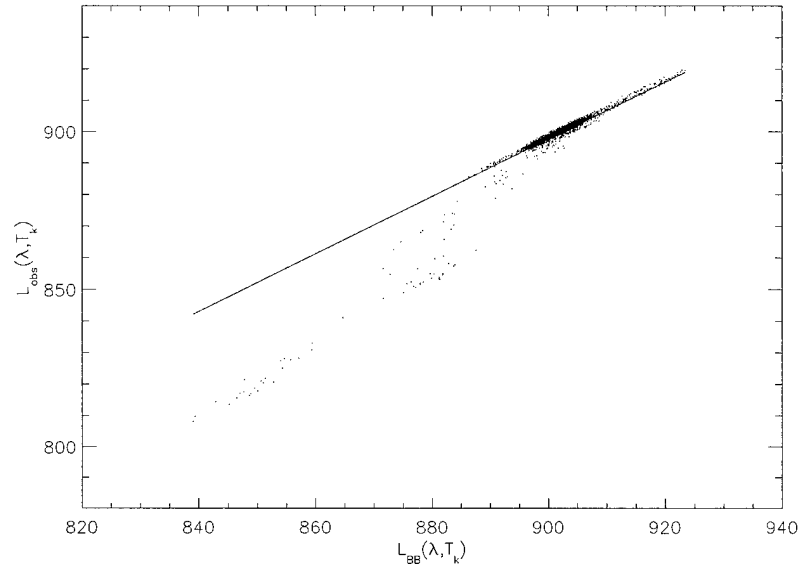


Figure 2.15: Scatter plot for the determination of blackbody pixels

and upwelled radiance, the slope and intercept must be found for all of the bands of the sensor. The slope and intercept may be calculated using a standard least-squares regression. Finally, note that a larger range of surface-leaving radiance values leads to a more accurate estimation of the unscaled parameters. The spread is dictated by the surface temperature distribution of a given scene and by the atmospheric transmission.

An alternative approach is to construct a scatter plot of *all* of the pixels in the image. The Kolmogorov-Smirnov goodness-of-fit test can then be used to fit a line across the top of the scatter of points. The basic premise behind this method is that for a given brightness temperature and wavelength, the pixels that generate the largest values of observed radiance are the most likely to be blackbodies. The unscaled parameters are then the slope and intercept of this line. The advantage of this method is that the sensor noise can be accounted for by fine-tuning the test statistic. Furthermore, there will be a correspondingly larger spread of temperature values because all pixels are used. This results in a more robust line fit. The drawback is that the atmospheric spectra retrievals are more prone to errors induced by surface emissivity effects.

Scaled Parameters

Although the unscaled parameters are representative of the spectral structure of the actual atmospheric parameters, it is necessary to scale the parameters to true values when accurate radiometric studies are needed. Johnson and Young (1998) present several methods. The simplest assumes that the transmission is known at a given wavelength. The unscaled transmission is then scaled to these values. The upwelled radiance can then be found by using the scaled transmission spectrum and by assuming the true effective atmospheric temperature is accurately estimated from the unscaled upwelled radiance. This relationship is derived from the approximation

$$L'_u(\lambda) \approx [1 - \tau'(\lambda)]L_{BB}(\lambda, T'_a) \quad (2.93)$$

where T'_a is an effective atmospheric temperature based on the unscaled parameters. A similar relationship holds with the *scaled* parameters so that

$$L_u(\lambda) \approx [1 - \tau(\lambda)]L_{BB}(\lambda, T_a) \quad (2.94)$$

where T_a is the effective atmospheric temperature derived from the scaled parameters. Johnson and Young claim that numerical analysis can be used to show that $T'_a \approx T_a$ so that setting eqs. (2.93) and (2.94) equal yields

$$\frac{L_u(\lambda)}{1 - \tau(\lambda)} = \frac{L'_u(\lambda)}{1 - \tau'(\lambda)} \quad (2.95)$$

This equation can then be used to solve for L_u . However, it is numerically unstable because it requires division by $1 - \tau'(\lambda)$, which approaches zero in regions of interest where the atmosphere is highly transmissive—particularly when using the unscaled transmission.

A similar approach requires that both the true transmission and upwelled radiance be known at a wavelength λ_0 . The relationship between the scaled and unscaled parameters is built by equating eqs. (2.90) and (2.86). This results in

$$\tau'(\lambda)L_{BB}(\lambda, T') + L'_u(\lambda) = \tau(\lambda)L_{BB}(\lambda, T) + L_u(\lambda) \quad (2.96)$$

Substituting λ for λ_0 and solving for T yields

$$T = T_b \left\{ \lambda_0, \left[\frac{1}{\tau(\lambda_0)} \tau'(\lambda_0) L_{BB}(\lambda_0, T') + L'_u(\lambda_0) - L_u(\lambda_0) \right] \right\} \quad (2.97)$$

Two observed brightness temperatures T'_1 and T'_2 can then be chosen and substituted into this equation to yield two independent estimates of the “true” temperatures T_1 and T_2 . This assumes that $\tau(\lambda_0)$ is not close to zero. These “true” temperatures can then be reinserted into eq. (2.96) to form a linear system of two equations. These can then be used to solve for the scaled parameters $\tau(\lambda)$ and $L_u(\lambda)$.

Another technique involves the use of MODTRAN to calculate transmission and upwelled radiance values at a given wavelength and then scaling the parameters to these values. The wavelength used for this procedure is that for which the unscaled upwelled radiance is a minimum. This reduces errors associated with transmission through the atmosphere and downwelled radiance. Another method, also involving MODTRAN, is to use the relationship between the slope of the water vapor continuum and the strength of the absorption in the vicinity of the 11.7 μm water absorption band. The hypothesis was that this relationship was correlated to certain parameters such as surface temperature, effective atmospheric temperature, and the transmission for that band. Several look-up-tables were generated varying these parameters, which were then used as inputs into MODTRAN. The MODTRAN-predicted radiances were then related to the input parameters. This relationship was then used to infer these input parameters from the observed data. The estimated parameters were then processed with MODTRAN to obtain estimates of the atmospheric transmission, upwelled radiance, and downwelled radiance. This approach is similar to the model-matching techniques described in Section 2.1.4

2.2 Temperature and Emissivity

The surface temperature and emissivity are important parameters in several applications. In this section, current techniques used for the separation of temperature and emissivity

effects are presented. These techniques rely on successful compensation of atmospheric effects. Thus, errors in the estimate of surface radiance will strongly impact the ability to accurately measure surface temperature and emissivity. Another challenge with this problem is that the spectral surface radiance consists of N measurements, while the unknown spectral emissivity and temperature add up to $N + 1$ unknowns. Therefore, this is an underdetermined mathematical problem. This section covers three separate techniques for the measurement of surface temperature and emissivity given that the atmosphere is well characterized and an estimate of surface radiance exists.

2.2.1 Split-Window Algorithms

The split-window method is the most popular algorithm for measuring surface temperature from a remote platform. The main reason for its popularity is that it is simple to implement and consistently yields reasonably accurate measurements. The technique was originally developed for determination of Sea Surface Temperatures (SST). For this application, the split-window method works very well since the emissivity of the water is well-known. However, complications arise from its implementation over land, especially for arid land surfaces that may contain materials with emissivities that have substantial spectral contrast (Pieters and Englert 1993). For these situations, the estimate of surface temperature could have substantial errors. A great deal of effort has been directed at modifying existing split-window algorithms to compensate for emissivity effects. Unfortunately, the lack of knowledge and measurements of emissivities at the required spatial scales limits the utility of this algorithm for land surface temperature measurements (Prata et al. 1995; Goita and Royer 1997; Caselles et al. 1997a). Still, it is valuable to understand the basic theory behind this technique since it may be extended to applications in which more than two spectral bands are available. Finally, the split-window algorithm is a suitable baseline method for cases where a limited number of bands are available.

The main assumption of the split-window technique is that the ratio of radiances measured in two spectral bands is proportional to the ratio of the absorption coefficients for

the same bands. The theoretical basis for this technique is summarized from that presented by Schott (1997). The split-window algorithm works best when the target is a blackbody. Using this simplified case, the linear representation of the radiative transfer equation can be written as

$$L(\lambda) = \tau(\lambda)L_{BB}(\lambda, T_s) + [1 - \tau(\lambda)]L_{BB}(\lambda, T_a) \quad (2.98)$$

where $L_{BB}(T_a)$ is the radiance emitted by the atmosphere for an effective atmospheric temperature T_a . The second term on the right side of the equation is equivalent to the upwelled radiance L_u .

The transmission can be represented with a first-order Maclaurin series. From eq. (2.9), the series expansion is

$$\tau(z) = e^{-\beta_{ext}z} \approx 1 - \beta_{ext}z \quad (2.99)$$

Substituting this equation into eq. (2.98) results in

$$L(\lambda) = L_{BB}(\lambda, T_s) - [L_{BB}(\lambda, T_s) - L_{BB}(\lambda, T_a)]\beta_{ext}z \quad (2.100)$$

This equation can be rewritten in terms of the brightness temperatures to yield an approximate relationship

$$T_i = T_s - [T_s - T_a]\beta_{ext_i}z \quad (2.101)$$

where T_i is the brightness temperature at the i^{th} sensor band, T_s is the surface temperature, and β_{ext_i} is the extinction coefficient in the i^{th} spectral bandpass. Thus, the surface temperature is the intercept of the regression line through the scatter plot of T_i against $\beta_{ext_i}z$. A practical consideration is that the extinction coefficients used over the range of spectral bands must sufficiently vary to allow an accurate linear regression fit. Thus, the sensor must have spectral bands located along the wings of atmospheric absorption features or where the continuum changes rapidly. This is the same requirement needed for atmospheric sounding. When there are only two windows available, the surface temperature can be obtained from eq. (2.101):

$$T_s = \frac{T_1\beta_{ext_2} - T_2\beta_{ext_1}}{\beta_{ext_2} - \beta_{ext_1}} \quad (2.102)$$

where T_1 and T_2 are the brightness temperatures for the selected windows. Recall that so far we have neglected to consider the case where the surface is not a blackbody radiator. The effects of emissivity are difficult to compensate. To do this, eq. (2.102) can be generalized and represented in terms of three coefficients that are derived empirically for a set of atmospheric and surface conditions (Prata et al. 1995). The general form of the split-window temperature is then

$$T_s = aT_1 + bT_2 + c \quad (2.103)$$

The coefficients a , b , and c typically depend on emissivity and transmission effects. Because the true emissivity values are not known, the accuracy of the split-window temperature is limited.

2.2.2 Alpha-Derived Emissivities

This technique uses the concept of *alpha residuals* developed by Kealy and Gabell (Kealy and Hook 1993). The goal of this technique is to separate the contribution of temperature and emissivity to brightness by the use of Wien's approximation of the Planck equation. From this, a mathematical manipulation of the parameters yields a quantity—the alpha residual—that can be obtained from direct measurements and knowledge of the spectral response of the sensor. The alpha residual has a spectral shape associated with the actual emissivity. The actual emissivity can then be derived from the alpha-residual because the statistical properties of the alpha-residual are associated empirically with the statistical properties of the emissivity.

To derive the form of the alpha residual, we begin with Wien's approximation to the Planck function:

$$L_{BB}(\lambda, T) \approx \frac{c_1}{\lambda^5 \exp[c_2/\lambda T]} \quad (2.104)$$

where the coefficients c_1 and c_2 are the same as for the Planck function. The radiance from

an object is

$$L(\lambda, T) = \varepsilon(\lambda)L_{BB}(\lambda, T) \quad (2.105)$$

Now consider the case where there are N discrete spectral-radiance measurements indexed by j . After taking the natural logarithm of both sides and multiplying by λ_j we get

$$\lambda_j \ln(L_j) = \lambda_j \ln(\varepsilon_j) + \lambda_j \ln(c_1) - 5\lambda_j \ln(\lambda_j) - \frac{c_2}{T} \quad (2.106)$$

The expected value of this equation yields

$$E[\lambda_j \ln(L_j)] = E[\lambda_j \ln(\varepsilon_j)] + E[\lambda_j \ln(c_1)] - 5E[\lambda_j \ln(\lambda_j)] - \frac{c_2}{T} \quad (2.107)$$

and subtracting from eq. (2.106) we get

$$\begin{aligned} \lambda_j \ln(L_j) - E[\lambda_j \ln(L_j)] &= \lambda_j \ln(\varepsilon_j) - E[\lambda_j \ln(\varepsilon_j)] + \lambda_j \ln(c_1) - E[\lambda_j \ln(c_1)] \\ &\quad - 5\lambda_j \ln(\lambda_j) + 5E[\lambda_j \ln(\lambda_j)] \end{aligned} \quad (2.108)$$

which effectively removes the dependence on temperature. All of the terms on the right side that do not include the spectral emissivity are either constants or are known from the detector spectral response. Thus, these can be grouped into a band-dependent value K_j . After rewriting the previous equation, we obtain the alpha residual:

$$\alpha_j = \lambda_j \ln(L_j) - E[\lambda_j \ln(L_j)] + K_j = \lambda_j \ln(\varepsilon_j) - E[\lambda_j \ln(\varepsilon_j)] \quad (2.109)$$

Equation (2.109) shows that the alpha residuals can be directly calculated from our knowledge of the surface radiance and the detector's characteristics. From this equation, the emissivity is

$$\varepsilon_j = \exp \left[\frac{\alpha_j + E[\lambda_j \ln(\varepsilon_j)]}{\lambda_j} \right] \quad (2.110)$$

To estimate this parameter, it is necessary to estimate or approximate the expected-value term on the right side of the equation. Kealy and Hook (1993) report that the variance of the alpha residual is empirically related to the expected value term in eq. (2.110) by

$$E[\lambda_j \ln(\varepsilon_j)] = \bar{X} = c \left[\sigma_{\alpha_j}^2 \right]^{1/M} \quad (2.111)$$

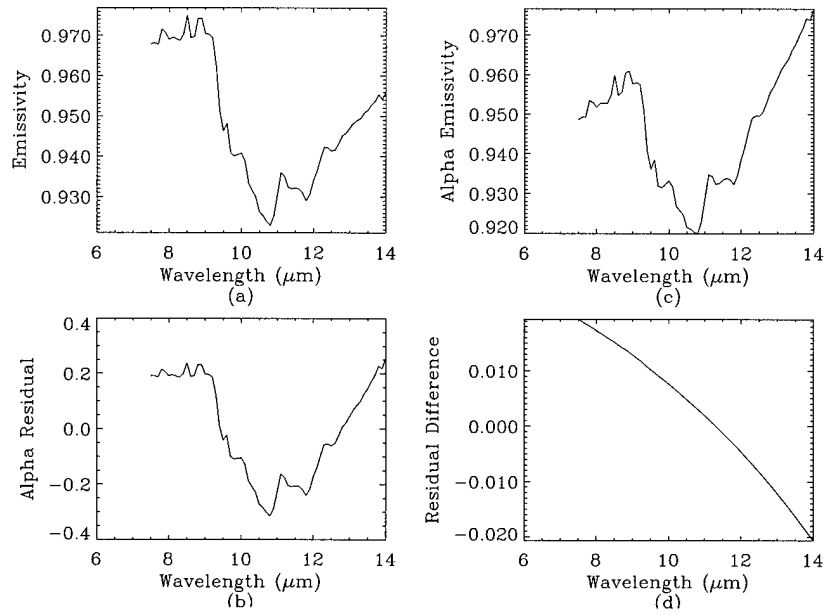


Figure 2.16: Plot of alpha-residuals and emissivity.

where the coefficients c and M are obtained from a nonlinear least-squares fit (2.110). The coefficients vary depending on the type of material (e.g., igneous rocks, soils, etc.) although the variation is not great and the classes are fairly broad. Also, the optimum coefficients will vary depending on the configuration of the sensor.

Figure 2.16 shows an example of an alpha residual and an alpha-derived emissivity. Subplot (a) is the emissivity of pine tree obtained from the John Hopkins University spectral library included with the Environment for Visualization (ENVI) package (Better Solutions Consulting LLC 1998). Subplot (b) is the alpha-residual based on a modeled radiance. The modeled radiance is simply the product of the emissivity and the Planck function at a temperature of 300 °K. Note the slight difference in “tilt”, which is due to bias introduced by Wien’s approximation to the Planck function. Subplot (c) is the alpha-derived emissivity. In this example, the expected value in eq. (2.110) used is the true expected value from the original emissivity curve. Therefore, any errors in the estimate emissivity are due to Wien’s

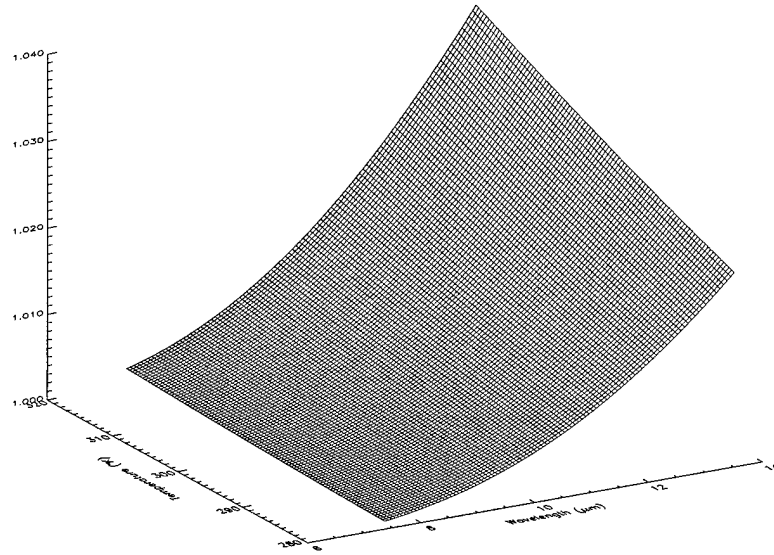


Figure 2.17: Bias due to Wien's approximation.

approximation. Subplot (d) is the difference between the true and alpha-derived emissivities and shows that the error due to Wien's approximation tends to bias the longer wavelengths more strongly. The magnitude of the error is 1-2 %, which is too high for a simple case with no other sources of error. Figure 2.17 is a plot of the ratio of the Planck blackbody function and Wien's approximation against temperature and wavelength. Although the wavelength dependence is known for a particular sensor, the temperature dependence is not unless some estimate of the surface temperature is used. If this is known, then an appropriate correction can be implemented. However, the goal of the alpha residual technique is to be independent of temperature. Thus, the implementation of a correction factor negates the entire basis of the alpha-residual technique.

2.2.3 Temperature and Emissivity Separation (TES) Algorithm

The TES algorithm has been developed to support the generation of standard data products for the ASTER sensor. The sensor is onboard the NASA Terra satellite that was launched in December, 1999. The algorithm is a combination of other techniques developed for the separation of temperature and emissivity effects. To summarize, the algorithm starts with an initial estimate of the maximum value in the emissivity spectrum and performs a series of iterations to find the best estimate of surface temperature and emissivity. The inputs required are the surface-leaving radiance and the total downwelled radiance. One of the key features of this algorithm is that it compensates for downwelled radiance. The next three sections briefly describe the main modules of the algorithm. Gillespie et al. (1999) provide an extensive description of the algorithm and its performance as implemented for ASTER. The algorithm has also been successfully implemented for the Thermal Infrared Multispectral Scanner (TIMS) sensor and tested extensively in the HAPEX-Sahel field campaign (Schmugge et al. 1997).

Normalized Emissivity Method (NEM) Module

The NEM module begins with an initial estimate of the maximum emissivity value. Since most materials have a high emissivity, the initial maximum emissivity ε_{max} is set to 0.96. This value is then used to compute an estimate of the surface radiance which includes the downwelled radiance term and is given as follows:

$$\hat{R}(\lambda) = L_s(\lambda) - (1 - \varepsilon_{max})L_d(\lambda) \quad (2.112)$$

where $\hat{R}(\lambda)$ is the interim estimate of the surface radiance due to emission only. In this formulation, it is assumed that the surface-leaving radiance was calculated without considering the downwelled radiance term. The NEM temperature is then defined as the largest brightness temperature associated with $\hat{R}(\lambda)$. This NEM temperature $\hat{T}$ is then inserted into the Planck function to get the blackbody radiance. The new spectral emissivity can

then be found by

$$\hat{\varepsilon}(\lambda) = \frac{\hat{R}(\lambda)}{L_{BB}(\hat{T})} \quad (2.113)$$

The entire spectral emissivity curve (instead of just ε_{max}) is then used with eq. (2.112) to calculate a new $\hat{R}(\lambda)$. This process is then repeated until $\hat{R}(\lambda)$ is less than or equal to some threshold. As implemented for ASTER, the algorithm uses the noise-equivalent radiance difference as the threshold. The final spectral emissivity and temperature are then used in the Ratio Module.

Ratio (RAT) Module

This is the simplest step of the algorithm. It merely involves calculating the average emissivity from the spectral curve obtained with the NEM module. The spectral emissivity is then scaled to this mean value giving

$$\beta(\lambda) = \frac{\varepsilon(\lambda)}{\bar{\varepsilon}} \quad (2.114)$$

Minimum-Maximum Difference (MMD) Module

The β spectrum from the RAT module can then be used to define an empirical relationship between the observed spectrum and the actual emissivity. The MMD is defined as

$$MMD = \max(\beta) - \min(\beta) \quad (2.115)$$

Using laboratory emissivity spectra, an empirical relationship between this MMD value and the minimum emissivity ε_{min} was found to be

$$\varepsilon_{min} = 0.994 - 0.687 \cdot MMD^{0.737} \quad (2.116)$$

where the minimum emissivity was chosen because it resulted in a higher correlation. In contrast to the empirical relationship found for the alpha-derived emissivity, this empirical relationship is valid for a relatively wide variety of target types as shown in Figure 2.18.

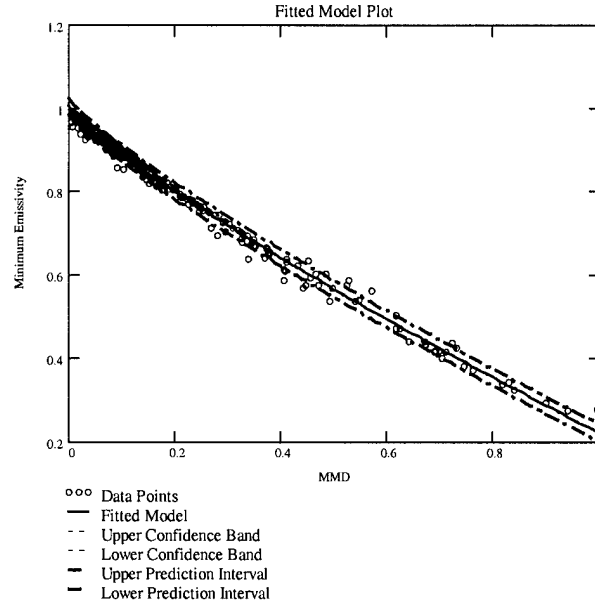


Figure 2.18: Determination of Empirical Relationship Between MMD and ε_{min} .

The TES emissivities are then calculated from this empirical relation such that

$$\varepsilon(\lambda) = \beta(\lambda) \frac{\varepsilon_{min}}{\min\{\beta(\lambda)\}} \quad (2.117)$$

The maximum of TES spectral emissivity and $\hat{R}$ are then used to calculate the TES temperature by using equation (2.84). Using the maximum emissivity minimizes errors introduced by ambiguities in the estimate of the downwelled radiance term.

Final TES Temperature and Emissivities

The TES temperature and emissivities calculated by the MMD module are then used as inputs to the NEM module for one single final pass through the NEM, RAT, and MMD modules. This final pass is non-iterative, and can lead to refinement of TES emissivities by as much as 0.01 (Gillespie et al. 1999). Testing for the ASTER sensor has shown that there is little gain in doing this final pass more than once (Gillespie et al. 1999).

2.2.4 Other Methods

There are many temperature and emissivity estimation techniques described in the literature. Caselles et al. (1997b) provide a good survey of these and make some comparisons. In addition to the survey, a geometric model for estimating emissivity is described which accounts for the heterogeneity of land surface. Although the inclusion of heterogeneity effects in emissivity calculations should theoretically yield better estimates, it introduces another unknown quantity to the problem; namely, the location and amount of heterogeneity that must be determined before an estimate of the emissivity can be made. Thus, the method relies on scene classification analysis which may introduce other errors and complexities.

Some techniques for simultaneous measurement of temperature and emissivity have been developed within the metallurgy community. These techniques were applied to measurements made with multispectral pyrometers (Hunter et al. 1985). Because the spectral shape of emissivities tend to be smooth, it is reasonable to parameterize the emissivities in terms of linear, polynomial, or exponential curves. The parameters representing the emissivity are then determined from a regression fit to observed radiances.

More recently, methods for estimating emissivities based on the smoothness criterion have been proposed. One method uses the decorrelation number as a measure of the smoothness of the emissivity curves (Borel 1998). By using an iterative algorithm with an initial temperature estimate, the smoothest emissivity curve is selected as the best estimate. Another method decomposes the spectral emissivity into a truncated Fourier series (Liang 1998). Only a small set of components need to be retained in the series because emissivities are smooth. The coefficients of the Fourier series are then solved by an iterative process with proper constraints. One potential improvement derived from this approach is that the set of unknown parameters needed to determine the emissivity can be considerably less than those required by traditional processing algorithms. In this case, it is possible to end up with an overdetermined system of equations which could potentially reduce the error in the estimate of the emissivity.

With the exception of the split-window algorithm, it has been assumed thus far that a relatively high spectral resolution sensor is available for the determination of temperature and emissivity. The general form of eq. (2.101) suggests that it may be possible to determine the temperature and emissivity to some degree of accuracy with a moderate-resolution multispectral sensor. One approach that can be used to overcome the limitation of having few spectral bands is the day-night method proposed by Becker and Li (1993) and recently planned for implementation for MODIS (Wan 1999; Wan and Li 1997). This method assumes that the platform is able to visit the same scene twice within a period of 24 hours. One collection is performed during the day while the other is done at night. This doubles the spectral measurements provided that the inherent surface properties do not change (i.e., emissivity). Furthermore, if the sensor has a spectral band in the Shortwave Infrared (SWIR) where both reflectance and emission are present during the day, the radiance can be compared directly to the emission at night to estimate the reflectance/emissivity. Certain practical issues must be considered when implementing this method. One is the directional property of emissivity for the surfaces being imaged. If a target has a specular character to it, the daytime radiance measurement will have a bias associated with it if the characteristics of the surface are not compensated. This problem is somewhat ameliorated by the fact that most materials are approximately Lambertian for SWIR wavelengths. Another complication is that the reflectance in the SWIR is a more dominant factor than in the thermal infrared. Thus, special considerations for downwelled radiance are critical. Finally, the algorithm assumes that the surface conditions of the scene do not change considerably between collection times and that the images can be spatially-registered accurately.

2.3 Summary and Discussion

This chapter described the fundamental theory of infrared radiation and propagation through the atmosphere. It also introduced the difficulties related to atmospheric effects and the combination of temperature and emissivity effects. Several techniques for atmospheric com-

pensation, as well as temperature and emissivity separation, were introduced. This section summarizes the methods discussed in this chapter with a discussion of the advantages and disadvantages of each approach. Table 2.1 highlights key points in this discussion. Common to all of the atmospheric compensation techniques is the inability to determine the reflected downwelled radiance component of the observed radiation. Common to all of the temperature and emissivity separation algorithms is the need for an empirical scaling relationship between biased estimates of the emissivity and the “true” emissivity.

The ISAC algorithm is attractive because it requires no ancillary or *a priori* information to obtain reasonable estimates of the surface-leaving radiance. It is also able to obtain scene-derived parameters without requiring the sensor to have high spectral calibration fidelity. However, it requires a wide spatial distribution of temperature and a stationary atmosphere across the scene. These two requirements are contradicting because a larger scene is needed to increase the spread of surface temperatures, which makes the assumption of a stationary atmosphere less appropriate. This is particularly the case for a scene without much thermal contrast and with high water vapor variability. Finally, it is difficult to retrieve *absolute* radiometric measurements because of the ambiguity in the atmospheric parameter retrievals. To scale the parameters appropriately requires the use of ancillary or *a priori* information, thus negating the algorithm’s main advantage.

The statistical minimum-variance sounding algorithm is versatile, in the sense that it can be used to build a statistical relationship between any atmospheric parameter and the observed radiation. The method can also perform retrievals on a per-pixel basis, thus addressing spatial variability in the atmosphere. To build the relationships, it uses *a priori* knowledge—represented by atmospheric statistics—rather than image spatial statistics. Unfortunately, the accuracy of these statistics depends on the number of observations in the ensembles used to estimate them. Therefore, it may be necessary to build large ensembles, particularly for hyperspectral sensors. This becomes an issue when computational, time, and storage resources are limited. Also, as the spectral resolution of the sensor increases,

Method	Advantages	Disadvantages
ISAC	Simple; does not require accurate spectral calibration; completely in-scene method (unscaled parameters); fast	Assumes same atmosphere over spatial scale of image ; requires temperature spread for regression; requires ancillary information for scaling
Statistical Sounding	Does not require weighting functions; simple; numerically stable; versatile; handles spatially-varying atmosphere; fast	Not “physical”; requires large ensemble database to build correlation matrices; may suffer from rank-deficiency or ill-conditioning
Linear Sounding	Yields profiles for temperatures and constituents; handles spatially-varying atmosphere; “physics-based”	Requires weighting functions; does not account for nonlinearities in radiative transfer; requires many narrow bands in absorption regions
Nonlinear Sounding	Same as linear sounding + accounts for nonlinear radiative transfer	Requires weighting functions at each iteration step; not guaranteed to converge on a solution; needs good initial estimate
Model-Matching	Versatile; handles spatially-varying atmosphere; accounts for nonlinear radiative transfer	Not guaranteed to converge; computationally intensive; suboptimal estimation of atmospheric parameters
Split Window	Simple; does not require many spectral bands	Low accuracy; does not estimate surface emissivity
Alpha residuals	Simple; works well in classification applications; does not require estimate of surface temperature	Bias from Wien’s approximation; difficult to obtain accurate estimates of true emissivities
TES	Compensates for reflected downwelled radiance estimate; solves for temperature and emissivity simultaneously; works with many material types;	Assumes relationship between variability and true emissivity

Table 2.1: Summary Table of Atmospheric Correction and Temperature and Emissivity Methods

the ensembles may become more ill-conditioned. Finally, there is no guarantee that the relationship will work with observations that are not part of the *a priori* ensembles because the solutions may not be physical.

Linear sounding algorithms may be used to obtain atmospheric temperature and constituent profiles simultaneously. The solutions are based on a direct inversion of the radiative transfer equation and are therefore physical. Regularization methods can be implemented to make the inverse problem better conditioned. However, this method does not account for nonlinearities in the radiative transfer. It also requires the use of weighting functions, which are built with radiative transfer models and are based on specific atmospheric conditions. Thus, the weighting functions could be inaccurate if the actual conditions are much different than those used in the model. Perhaps of greater concern is that the weighting functions are very “sensor-specific” and may not yield accurate profiles if the sensor is not designed to optimize the weighting functions. That is, the weighting functions may be too broad if the sensor has low spectral resolution. While the accuracy of low-resolution sensors is inherently limited, physical sounding schemes may be more severely affected than others.

Nonlinear sounding techniques share the same advantages as linear sounding techniques and also account for nonlinearities in the radiative transfer. Through the implementation of an iterative algorithm, it is possible to refine a solution at each step. The solutions are likely to be more accurate than linear solutions if convergence is achieved. However, convergence is never guaranteed and is not likely if the sensor is not optimally designed for sounding. A good initial estimate of the profiles is also required.

Model-matching techniques are also able to account for nonlinear effects in radiative transfer. These techniques can also account for nonlinearity effects due to coupling between the atmosphere and the surface. Like the statistical sounding approach, it is versatile because the number of possible parameter retrievals is limited only by the outputs of the forward model. Model-matching requires a parameterization of the model inputs. Convergence problems may arise depending on the complexity of the input parameter scheme.

Perhaps the biggest drawback of this technique is that it is suboptimal because the convergence criterion is based on how well the “at-sensor” radiance is matched by the model output and not on how accurate the atmospheric parameters are. That is, unless appropriately constrained an optimal match of the radiances does not imply an optimal match on the atmospheric parameters. This was demonstrated in section 2.1.5 with the comparison of least-squares and Twomey-Tikhonov regularized solutions (Figure 2.13).

The split-window technique provides a viable solution for a sensor with a limited number of bands. It is also simple and efficient. However, its accuracy is limited for land surface retrievals because the emissivity is unknown. Furthermore, it cannot be used to estimate spectral emissivity curves.

Alpha residuals are attractive because an estimate of the surface temperature is not required to derive a spectral curve that is related to the emissivity. This is particularly useful when absolute thermography is not needed (e.g., a classification or target detection application). However, a bias is introduced by Wien’s approximation of the Planck function. This bias is *temperature-dependent*, making it difficult to correct unless the temperature is known.

Finally, the TES algorithm has the unique ability to use an estimate of the downwelled radiance in its calculation of the surface temperature and emissivity. Also, both temperature and emissivity are obtained simultaneously. Unfortunately, the accuracy of the results depend on the validity of the algorithm’s main assumption: that the variability of the emissivity curve is related to a true minimum emissivity value for all targets of interest.

Clearly, no particular technique provides a suitable solution to the problem at hand. They do, however, provide enough theoretical and practical background to determine a suitable approach. Chapter 3 describes a unified and comprehensive approach that combines the advantages of some of these techniques, and introduces considerations that minimize the effect of their disadvantages. The goal is to determine the feasibility of this new algorithm for the exploitation of infrared hyperspectral data obtained from air and space.

Chapter 3

Approach

Simplicity, simplicity, simplicity! I say, let your affairs be as two or three, and not a hundred or a thousand; instead of a million count half a dozen, and keep your accounts on your thumbnail.

Henry David Thoreau, *Walden*, (1854)

The radiance reaching a remote sensor is the result of complex interactions between the inherent properties and thermodynamic state of the Earth's surface and atmosphere. Chapter 2 introduced the physics that govern radiative transfer and emission. From this analysis, a mathematical model describing a mapping from the atmospheric state space to the sensor measurement space was developed. From the perspective of classical physics, this mapping is deterministic. In reality, the laws of uncertainty cannot be avoided and the problem is stochastic. Regardless of the mechanism, the result is the same: each measurement is the combination of several—often indistinguishable—effects.

The remote sensing scientist really gets the raw end of the deal. While years of research have led to a good understanding of radiative processes, the same cannot be said about the inverse problem. This is mainly due to the unavoidable loss of information in the mapping from state space to measurement space. We attempt to circumvent this unfortunate state

of affairs by arming ourselves with as much *a priori* and ancillary information as possible. That is, we hope that other sources of information can replace the information lost in the radiative transfer. Finally, we turn to the old adage: “more is better”; and design instruments that make many more measurements than ever before. Hence, the recent proliferation of hyperspectral sensors.

However, it may be prudent to heed Thoreau’s advice and strive for “simplicity”. Here, “simplicity” does not mean *oversimplifying* the problem by making too many assumptions. Rather, we seek to cast the problem onto a framework where hundreds of variables can effectively be summarized by a few. There are several advantages gained with this approach:

- The ill-conditioning of the inverse problem may be avoided by working with a smaller set of variables.
- Nonlinear interactions can be avoided by working in a space where variables are independent.
- Resources are not wasted on redundant information.
- The mechanism used to cast the problem onto a more manageable framework may provide insight into the physics of the problem.
- An optimality criterion can be more easily implemented.

To understand how such a scheme may be developed, consider a set q parameters of interest that give rise to p observations. We wish to identify the *inherent* or *latent* relationships between the two sets. The relationships can be characterized by how a change in one set translates to a change in the other set. These related variations, or *correlations*, are a first order summary statistic that effectively describe how the parameters and observations are related. Rather than computing all the possible correlations, we are interested in the latent correlations that summarize *all* of the correlations. Furthermore, we wish to find

correlations that are orthogonal or *independent* so that redundancy is minimized. Depending on the nature of the mechanism relating the q parameters and the p observations, it is possible for a small number of latent correlations to summarize a large number of correlations between the data sets. Canonical Correlation Analysis (CCA) is a multivariate method that defines how these latent correlations can be obtained. By letting q represent atmospheric parameters of interest and p the number of spectral radiance values measured by p channels in an imaging system, we can begin to develop an analytical framework for the solution of the inverse problem in remote sensing.

This chapter outlines the theory of CCA and its implementation. Section 3.1 derives CCA and shows how it can be used to find optimal solutions of the inverse problem. It is also shown that CCA can be used to gain an understanding of what physical variables lead to the highest correlations between two data sets. In fact, CCA can be used to find the least number and optimal placement of spectral bands. A case study demonstrating this is presented in Chapter 4. Section 3.2 describes how CCA can be used to infer atmospheric and surface parameters directly from the observed radiance. This section also covers how CCA and TES can be used together to estimate surface temperature and emissivity. Finally, Section 3.3 describe the data and methodology used to test and validate the approach.

3.1 Theoretical Basis and Development

Consider the $p \times 1$ vector $\mathbf{x}$ which contains p variables (e.g., spectral radiance values at p wavelengths) and comes from a population with a multivariate probability $P(\mathbf{x})$. Now consider the $q \times 1$ vector $\mathbf{y}$ which contains q variables (e.g., atmospheric temperature at q altitude levels) and comes from a population $P(\mathbf{y})$. If there is any relationship between the populations, Bayesian statistics may be used such that

$$P(\mathbf{y}|\mathbf{x}) = P(\mathbf{y})P(\mathbf{x}|\mathbf{y})/P(\mathbf{x}) \quad (3.1)$$

yields the optimal solution for $\mathbf{y}$ given $\mathbf{x}$. Thus, the probabilities provide a *pathway* between the observation and the object or process that most likely caused the observation. The difficulty with this approach is in defining the multivariate probability distributions which describe the populations where $\mathbf{x}$ and $\mathbf{y}$ come from.

3.1.1 Ordinary Least Squares and Principal Components

In the absence of known multivariate probability distributions, the best estimate of $\mathbf{y}$ given $\mathbf{x}$ can be obtained by building an ensemble of $n \times p$ observations $\mathbf{X}$ and $n \times q$ dependent variables $\mathbf{Y}$ and finding the linear combination of $\mathbf{X}$ that results in predictions $\hat{\mathbf{Y}}$ such that the squared error

$$\epsilon^2 = \sum_{i=1}^n (\mathbf{y}_i - \hat{\mathbf{y}}_i)^2 \quad (3.2)$$

is minimized. This is the approach of Ordinary Least Squares (OLS) regression. The corresponding optimal linear combination of $\mathbf{X}$ is

$$\beta = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \quad (3.3)$$

where β is a $p \times q$ matrix representing a projection of $\mathbf{X}$ onto an *explanatory* space where the information in $\mathbf{X}$ about $\mathbf{Y}$ is maximally exploited. The regression coefficients are also the generalized inverse of $\mathbf{X}$. The arrows in Figure 3.1 denote the contribution of each variable in $\mathbf{X}$ to the prediction of a given variable in $\mathbf{Y}$. The contributions are weighted by the values of the regression coefficients. For this reason, the regression coefficients are often referred to as *weights*. The arrows are *unidirectional* because OLS implicitly assumes a *causal* relationship between $\mathbf{X}$ and $\mathbf{Y}$ (i.e., the model is not symmetric). The estimate of $\mathbf{Y}$ is given by $\hat{\mathbf{Y}} = \mathbf{X}\beta$.

Sometimes, it is appropriate to *mean-center* the data. Typically, this is done with respect to the variables represented by the columns of $\mathbf{X}$ and $\mathbf{Y}$. Thus, the mean of each column (i.e., variable) are computed over all n observations. The mean-centering is done by subtracting each observation in $\mathbf{X}$ and $\mathbf{Y}$ by its corresponding variable mean. The easiest

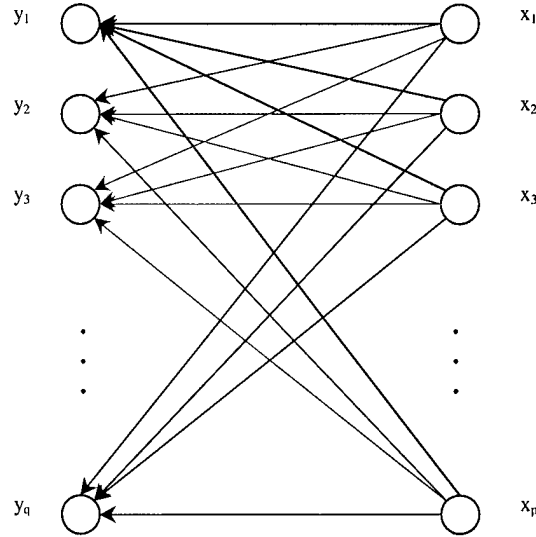


Figure 3.1: Path model for ordinary least-squares.

way to implement this is to compute the $(p \times 1)$ row mean vector $\bar{\mathbf{x}}$ and form a mean matrix $\bar{\mathbf{X}}$ by replicating the vector $\bar{\mathbf{x}}$ n times. The mean-centered data are then calculated by simple subtraction (e.g., $\mathbf{X} - \bar{\mathbf{X}}$). This is typically done for both $\mathbf{X}$ and $\mathbf{Y}$. Now, the covariance matrix of $\mathbf{X}$ is defined as

$$\frac{1}{n-1}(\mathbf{X} - \bar{\mathbf{X}})'(\mathbf{X} - \bar{\mathbf{X}}) \quad (3.4)$$

Therefore, when $\mathbf{X}$ and $\mathbf{Y}$ are mean-centered and scaled by the number of observations, $\mathbf{X}'\mathbf{X} = \Sigma_{\mathbf{xx}}$ and $\mathbf{Y}'\mathbf{Y} = \Sigma_{\mathbf{yy}}$ are the covariance matrices. This gives rise to an often overlooked interpretation of the least-squares regression coefficients of eq. (3.3). The coefficients are simply the projections of $\mathbf{X}$ onto $\mathbf{Y}$ scaled by the variances of $\mathbf{X}$. Thus, the variations in β are emphasize the variations in $\mathbf{Y}$.

The problem with OLS is that when p is large, the covariance matrix becomes ill-posed or rank-deficient and the inverse becomes impossible to calculate. Thus, the OLS solutions become unstable. A common approach is to decompose the covariance matrix

into *principal components*. The principal components are simply the eigenvectors of the $\mathbf{X}$ covariance matrix and result from

$$\Sigma_{\mathbf{xx}}\mathbf{A} = \mathbf{A}\mathbf{\Lambda} \quad (3.5)$$

where $\mathbf{\Lambda}$ is a $p \times p$ diagonal matrix containing the eigenvalues and $\mathbf{A}$ is a $p \times p$ matrix whose columns are the orthogonal eigenvectors. This analysis is referred to as Principal Components Analysis (PCA). The word “principal” is very appropriate, because the analysis finds the components (eigenvectors) that account for *most* of the variation in $\mathbf{X}$. That is, the eigenvectors point along the dimensions of maximum variance in $\mathbf{X}$.

If the original matrix is rank-deficient with rank $r < p$, then it is possible to condition the problem by retaining r eigenvectors and discarding the rest. The $\mathbf{X}$ data are then projected onto the *truncated* principal component space such that

$$\mathbf{U} = \mathbf{XA} \quad (3.6)$$

In this orthogonal space, the transformed matrix $\mathbf{U}$ (known as *scores*) is of full rank and suitable for OLS regression onto the dependent set $\mathbf{Y}$ as done in eq. (3.3) (Jackson 1991). This method is known as Principal Components Regression (PCR) and is shown via a path diagram on Figure 3.2. If $\mathbf{Y}$ is also rank-deficient, a *two-block* PCR may be implemented. That is, the $\mathbf{X}$ and $\mathbf{Y}$ data are transformed via independent principal components and the regression is done on the resulting scores.

3.1.2 Canonical Correlation Analysis

The underlying assumption in PCR is that the principal components of $\mathbf{X}$ will lead to a good prediction of $\mathbf{Y}$. However, the analysis is based on $\Sigma_{\mathbf{xx}}$ alone and there is no guarantee that the significant variance in $\mathbf{X}$ carries information about $\mathbf{Y}$. In Canonical Correlation Analysis (CCA) the joint structure between $\mathbf{X}$ and $\mathbf{Y}$ is considered and an optimal orthogonal space is created where the projections of $\mathbf{X}$ and $\mathbf{Y}$ are maximally correlated. This orthogonal space can be used for the same purpose as in PCR and provides a mechanism for dealing with

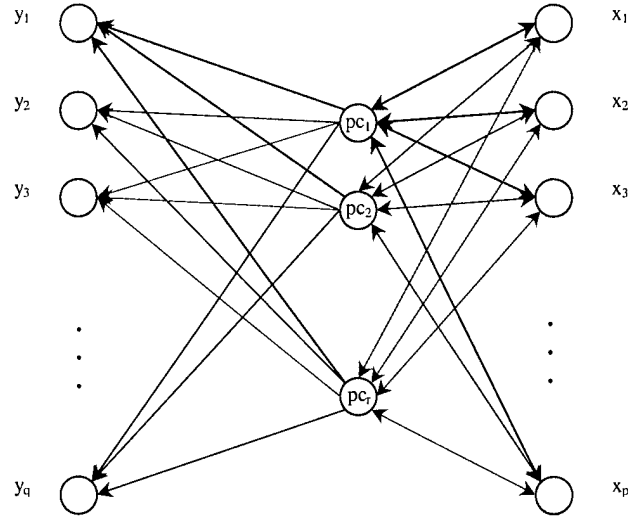


Figure 3.2: Path model for principal components regression.

rank-deficient matrices. In addition, the analysis can yield insight about which variables in $\mathbf{X}$ carry the most information about $\mathbf{Y}$. Finally, CCA is symmetric so that the distinction between predictor and dependent sets is not necessary.

The canonical correlations are the eigenvalues of

$$\begin{aligned}\Sigma_{xx}^{-1} \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx} \mathbf{A} &= \mathbf{A} \Psi \\ \Sigma_{yy}^{-1} \Sigma_{yx} \Sigma_{xx}^{-1} \Sigma_{xy} \mathbf{B} &= \mathbf{B} \Psi\end{aligned}\tag{3.7}$$

where Ψ is the $k \times k$ diagonal matrix of the *squared* canonical correlations and $k = \min(p, q)$. The eigenvectors defining the transformation of $\mathbf{X}$ are the columns of $\mathbf{A}$. Similarly, $\mathbf{B}$ contains the eigenvectors for the transformation of $\mathbf{Y}$ such that

$$\begin{aligned}\mathbf{U} &= \mathbf{XA} \\ \mathbf{V} &= \mathbf{YB}\end{aligned}\tag{3.8}$$

where $\mathbf{U}_{(n \times p)}$ and $\mathbf{V}_{(n \times q)}$ are the canonical variables (or scores) whose k squared correlations are defined by the diagonal entries of Ψ . Three properties are worth mentioning: (1) the

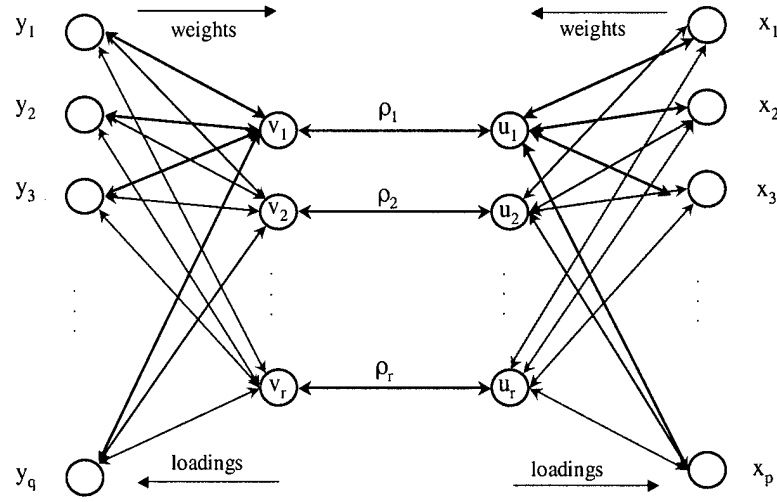


Figure 3.3: Path diagram for CCR

canonical variables are orthogonal (the variables $\mathbf{U}$ and $\mathbf{V}$ lie in an orthogonal canonical space spanned by the basis vectors in $\mathbf{A}$ and $\mathbf{B}$, respectively); (2) the canonical correlations are the maximum *linear* correlations that exist between the two data sets and are arranged in descending order of magnitude (the diagonal elements of Ψ are $\rho_1^2 \geq \rho_2^2 \geq \dots \geq \rho_r^2$); and (3) the canonical weights are normalized so that $\mathbf{A}'\Sigma_{xx}\mathbf{A} = \mathbf{B}'\Sigma_{yy}\mathbf{B} = \mathbf{I}_k$, where $\mathbf{I}_k$ is a $k \times k$ identity matrix (i.e., the canonical variables have unit variance). Other properties are given by Anderson (1984) and Johnson and Wichern (1992).

The flow between the original “observed” space $\mathbf{X}$ and the predictand space $\mathbf{Y}$ can be described via a path diagram as shown in Figure 3.3. The mapping from $\mathbf{X}$ to $\mathbf{U}$ is obtained from the canonical weights $\mathbf{A}$. Similarly, $\mathbf{V}$ is obtained from applying $\mathbf{B}$ to $\mathbf{Y}$. The inverse transformations going from the canonical to the original space are known as the *loadings*. The loadings are exact when the canonical dimensionality is the same as that of the original space. Otherwise, the loadings are the least-squares regression coefficients relating the canonical and original spaces (see proof in Section 3.1.3). In general, the loadings are smoother than the weights and are therefore more interpretable.

The regression is performed by applying eq. (3.3) to the canonical variables so that

$$\hat{\mathbf{V}} = \mathbf{U}\beta_{\mathbf{CC}} = \mathbf{U}(\mathbf{U}'\mathbf{U})^{-1}\mathbf{U}'\mathbf{V} \quad (3.9)$$

where $\beta_{\mathbf{CC}}$ is the matrix of regression coefficients for the canonical variables and is equal to Ψ . To find $\hat{\mathbf{Y}}$ from $\hat{\mathbf{V}}$, let $\hat{\mathbf{V}} \approx \mathbf{V} = \mathbf{Y}\mathbf{B}$ and use Property 3 such that

$$\hat{\mathbf{Y}} = \hat{\mathbf{V}}\mathbf{B}'\Sigma_{\mathbf{yy}} \approx \mathbf{Y}\mathbf{B}\mathbf{B}'\Sigma_{\mathbf{yy}} = \mathbf{Y}\mathbf{I}_q = \mathbf{Y} \quad (3.10)$$

A detailed proof is given in Section 3.1.3.

As in PCR, the orthogonal space may be reduced to dimensionality $r \leq k$ in order to stabilize the regression solution and prevent “overfitting” the data. This implementation of CCA, called Canonical Correlation Regression (CCR), is not widely used in the natural sciences with the exception of some implementation in climate modelling (Yaglom 1990; Yu et al. 1997). This is probably because its main use has been in econometrics and psychometrics where the emphasis is in the study of latent factors that are not physically measurable (e.g., intelligence, consumer preference, etc.). It has been typically shunned as a prediction tool because the optimality criterion is based on the latent variables and not on the observed variables. Therefore, it is possible that the canonical structure does not optimally explain the variance in $\mathbf{Y}$. However, the derivation in Section 3.1.3 demonstrates that the variability in $\mathbf{Y}$ is optimally predicted (in the least-squares sense) *subject to the constraint* that the canonical variables are maximally correlated. One of the goals of this research is to demonstrate that the CCR emphasis on the latent variables constrains the remote sensing problem adequately, leading to solutions that are physically interpretable. In other words, the CCR model is *physics-based*. This is because the model lies in a truncated lower-dimensional space made up of the highest correlations found between the two data sets. Assuming that the magnitude of physical correlations is larger than incidental “ensemble-dependent” correlations, this truncated space effectively summarizes the physics of radiative transfer. Thus, we can think of the CCR model as an “inverse model” of radiative transfer. This emphasis on physics ensures that the CCR inverse model is robust

and can be used for the estimation of parameters based on observations not used in the regression data.

Another hypothesis is that the CCR inverse model does not amplify sensor noise. This hypothesis stems from how the canonical variables are formed. When the data is projected onto the canonical space, the values leading to meaningful correlations are amplified. Conversely, uncorrelated variability across the data sets is suppressed. This is similar in concept to the Orthogonal Subspace Projection (OSP) background suppression method (Harsanyi and Chang 1994). In the worst case, noise in the observations will be linearly propagated through the model leading to an uncertainties in the estimated parameters that is proportional to the noise in the observations.

The CCR inverse model has the advantage of being symmetric. That is, there is no bias toward the prediction of $\mathbf{Y}$ or $\mathbf{X}$. Typical regression models are biased toward predicting $\mathbf{Y}$ based on $\mathbf{X}$. By doing so, it is implicitly assumed that the values in $\mathbf{X}$ are absolutely known. As will be shown in Section 3.2.1, neither $\mathbf{X}$ and $\mathbf{Y}$ are absolutely known in this application. Therefore, CCR is a suitable regression model for remote sensing.

3.1.3 Derivation of Canonical Correlation Analysis and Regression

The development presented in this section is divided in two parts. The first part derives the eigenvalue equation used for determining the canonical correlations and coefficients. The second part shows how canonical correlations can be cast into a predictive framework.

Eigenvalue Equation

The transformation for the first canonical variables is defined as $\mathbf{u} = \mathbf{X}\mathbf{a}$ and $\mathbf{v} = \mathbf{Y}\mathbf{b}$, where $\mathbf{u}$ and $\mathbf{v}$ are one-dimensional column vectors with n observations. CCA attempts to find

$$\max \{\text{corr}(\mathbf{u}, \mathbf{v})\} = \max \left\{ \frac{\mathbf{u}'\mathbf{v}}{\sqrt{\mathbf{u}'\mathbf{u}}\sqrt{\mathbf{v}'\mathbf{v}}} \right\} = \max \left\{ \frac{\mathbf{a}'\mathbf{X}'\mathbf{Y}\mathbf{b}}{\sqrt{\mathbf{a}'\mathbf{X}'\mathbf{X}\mathbf{a}}\sqrt{\mathbf{b}'\mathbf{Y}'\mathbf{Y}\mathbf{b}}} \right\} \quad (3.11)$$

where $\mathbf{a}$ and $\mathbf{b}$ are canonical coefficients or *weights*. The problem is constrained by requiring that the canonical variables have unit variance. Thus, $\mathbf{u}'\mathbf{u} = \mathbf{v}'\mathbf{v} = 1$. Subject to this constraint, the correlation to be maximized becomes

$$\text{corr}(\mathbf{u}, \mathbf{v}) = \mathbf{a}'\mathbf{X}'\mathbf{Y}\mathbf{b} \quad (3.12)$$

Following Phatak (1993), the method of Lagrangian multipliers may be used to construct the objective function

$$L = \mathbf{a}'\mathbf{X}'\mathbf{Y}\mathbf{b} + \frac{1}{2}\psi_x(\mathbf{a}'\mathbf{X}'\mathbf{X}\mathbf{a} - 1) + \frac{1}{2}\psi_y(\mathbf{b}'\mathbf{Y}'\mathbf{Y}\mathbf{b} - 1) \quad (3.13)$$

Where the parameters to be solved are $\mathbf{a}, \mathbf{b}$, and the Lagrangian multipliers ψ_x and ψ_y . Clearly, the function is at a maximum when $\mathbf{a}'\mathbf{X}'\mathbf{Y}\mathbf{b}$ is maximized and $\mathbf{a}'\mathbf{X}'\mathbf{X}\mathbf{a} = \mathbf{b}'\mathbf{Y}'\mathbf{Y}\mathbf{b} = 1$. The parameters are solved by setting the partial derivative of the objective function with respect to $\mathbf{a}'$ and $\mathbf{b}'$ equal to zero so that

$$\frac{\partial L}{\partial \mathbf{a}'} = \mathbf{X}'\mathbf{Y}\mathbf{b} + \psi_x\mathbf{X}'\mathbf{X}\mathbf{a} = 0 \quad (3.14)$$

$$\frac{\partial L}{\partial \mathbf{b}'} = \mathbf{Y}'\mathbf{X}\mathbf{a} + \psi_y\mathbf{Y}'\mathbf{Y}\mathbf{b} = 0 \quad (3.15)$$

Premultiplying 3.14 by $\mathbf{a}'$ and 3.15 by $\mathbf{b}'$ yields

$$\mathbf{a}'(\mathbf{X}'\mathbf{Y}\mathbf{b} + \psi_x\mathbf{X}'\mathbf{X}\mathbf{a}) = \mathbf{a}'\mathbf{X}'\mathbf{Y}\mathbf{b} + \psi_x = 0 \quad (3.16)$$

$$\mathbf{b}'(\mathbf{Y}'\mathbf{X}\mathbf{a} + \psi_y\mathbf{Y}'\mathbf{Y}\mathbf{b}) = \mathbf{b}'\mathbf{Y}'\mathbf{X}\mathbf{a} + \psi_y = 0 \quad (3.17)$$

where the unit variance constraint was applied. Now rearranging results in

$$\begin{aligned} \mathbf{a}'\mathbf{X}'\mathbf{Y}\mathbf{b} &= -\psi_x \\ \mathbf{b}'\mathbf{Y}'\mathbf{X}\mathbf{a} &= -\psi_y \end{aligned} \quad (3.18)$$

Since $\mathbf{a}'\mathbf{X}'\mathbf{Y}\mathbf{b}$ is a scalar then it is also equal to $\mathbf{b}'\mathbf{Y}'\mathbf{X}\mathbf{a}$ and

$$\mathbf{a}'\mathbf{X}'\mathbf{Y}\mathbf{b} = -\psi_x = -\psi_y = \psi_1 \quad (3.19)$$

Now substituting ψ_1 for $-\psi_y$ in 3.15 yields

$$\mathbf{Y}'\mathbf{X}\mathbf{a} = \psi_1 \mathbf{Y}'\mathbf{Y}\mathbf{b} \quad (3.20)$$

and solving for $\mathbf{b}$

$$\mathbf{b} = \frac{1}{\psi_1} (\mathbf{Y}'\mathbf{Y})^{-1} \mathbf{Y}'\mathbf{X}\mathbf{a} \quad (3.21)$$

Substituting $\mathbf{b}$ in 3.14 and letting $\psi_x = \psi_1$ results in

$$\mathbf{X}'\mathbf{Y}\mathbf{b} = \frac{1}{\psi_1} \mathbf{X}'\mathbf{Y}(\mathbf{Y}'\mathbf{Y})^{-1} \mathbf{Y}'\mathbf{X}\mathbf{a} = \psi_1 \mathbf{X}'\mathbf{X}\mathbf{a} \quad (3.22)$$

and rearranging

$$(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}(\mathbf{Y}'\mathbf{Y})^{-1} \mathbf{Y}'\mathbf{X}\mathbf{a} = \psi_1^2 \mathbf{a} \quad (3.23)$$

This is the eigenvalue equation used to solve for the canonical correlations and the linear combinations of $\mathbf{X}$. The same approach can be used to find the eigenvalue equation giving the linear combinations of $\mathbf{Y}$ by solving for $\mathbf{a}$ first using eq. (3.14) and then substituting into eq. (3.15) so that

$$\mathbf{a} = \frac{1}{\psi_1} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}\mathbf{b} \quad (3.24)$$

and

$$\frac{1}{\psi_1} \mathbf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}\mathbf{b} = \psi_1 \mathbf{Y}'\mathbf{Y}\mathbf{b} \quad (3.25)$$

Rearranging

$$(\mathbf{Y}'\mathbf{Y})^{-1} \mathbf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}\mathbf{b} = \psi_1^2 \mathbf{b} \quad (3.26)$$

If $\mathbf{X}$ and $\mathbf{Y}$ are mean-centered and scaled by the number of observations then $\mathbf{X}'\mathbf{X} = \Sigma_{xx}$, $\mathbf{Y}'\mathbf{Y} = \Sigma_{yy}$, $\mathbf{X}'\mathbf{Y} = \Sigma_{xy}$, and $\mathbf{Y}'\mathbf{X} = \Sigma_{yx}$. Thus,

$$\Sigma_{xx}^{-1} \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx} \mathbf{a} = \psi_1^2 \mathbf{a} \quad (3.27)$$

$$\Sigma_{yy}^{-1} \Sigma_{yx} \Sigma_{xx}^{-1} \Sigma_{xy} \mathbf{b} = \psi_1^2 \mathbf{b} \quad (3.28)$$

In general, the combinations of covariance matrices used in eq. (3.27) and eq. (3.28) are not symmetric. To simplify the implementation of this eigenvalue problem in a computer,

it is desirable to represent the problem in terms of symmetric matrices. This allows an SVD routine to compute “left” and “right” eigenvectors (i.e., the eigenvectors associated with the row and column spaces) that are identical. The matrices can be made symmetric by “factoring” out a square root matrix and redefining the eigenvectors. A square root matrix is defined as one where $\mathbf{A}^{1/2} \mathbf{A}^{1/2} = \mathbf{A}$. For the case of the $\mathbf{X}$ canonical coefficient solutions,

$$\Sigma_{xx}^{-1/2} \Sigma_{xx}^{-1/2} \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx} \mathbf{a} = \psi_1^2 \mathbf{a} \quad (3.29)$$

Multiplying both sides by $\Sigma_{xx}^{1/2}$ yields

$$\Sigma_{xx}^{-1/2} \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx} \mathbf{a} = \psi_1^2 \Sigma_{xx}^{1/2} \mathbf{a} \quad (3.30)$$

Now let $\mathbf{e} = \Sigma_{xx}^{1/2} \mathbf{a}$ and $\mathbf{a} = \Sigma_{xx}^{-1/2} \mathbf{e}$ so that the new eigenvalue equation becomes

$$\Sigma_{xx}^{-1/2} \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx} \Sigma_{xx}^{-1/2} \mathbf{e} = \psi_1^2 \mathbf{e} \quad (3.31)$$

Similarly for $\mathbf{b}$

$$\Sigma_{yy}^{-1/2} \Sigma_{yx} \Sigma_{xx}^{-1} \Sigma_{xy} \Sigma_{yy}^{-1/2} \mathbf{f} = \psi_1^2 \mathbf{f} \quad (3.32)$$

where $\mathbf{b} = \Sigma_{yy}^{-1/2} \mathbf{f}$.

So far, we have only discussed one canonical variable. In principle, there can be up to $k = \min(p, q)$ canonical variables and associated correlations. Fortunately, the eigenvalue analysis provides this readily:

$$\Sigma_{xx}^{-1/2} \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx} \Sigma_{xx}^{-1/2} \mathbf{E} = \mathbf{E} \Psi \quad (3.33)$$

$$\Sigma_{yy}^{-1/2} \Sigma_{yx} \Sigma_{xx}^{-1} \Sigma_{xy} \Sigma_{yy}^{-1/2} \mathbf{F} = \mathbf{F} \Psi \quad (3.34)$$

where Ψ contains the k canonical correlations along the diagonal and

$$\begin{aligned} \mathbf{A} &= \Sigma_{xx}^{-1/2} \mathbf{E} \\ \mathbf{B} &= \Sigma_{yy}^{-1/2} \mathbf{F} \end{aligned} \quad (3.35)$$

$\mathbf{A}$ and $\mathbf{B}$ are matrices comprised of the k canonical weights. The canonical variables are then computed by

$$\mathbf{U} = \mathbf{X} \mathbf{A} \quad \mathbf{V} = \mathbf{Y} \mathbf{B} \quad (3.36)$$

The eigenvector solutions $\mathbf{E}$ and $\mathbf{F}$ are orthonormal such that

$$\mathbf{E}'\mathbf{E} = \mathbf{F}'\mathbf{F} = \mathbf{I}_k \quad (3.37)$$

We can use the relationships in eq. (3.35) to express this property in terms of the canonical weights:

$$\mathbf{A}'\Sigma_{xx}\mathbf{A} = \mathbf{B}'\Sigma_{yy}\mathbf{B} = \mathbf{I}_k \quad (3.38)$$

By virtue of this property, the canonical variables are themselves orthonormal so that

$$\mathbf{U}'\mathbf{U} = (\mathbf{X}\mathbf{A})'\mathbf{X}\mathbf{A} = \mathbf{A}'\mathbf{X}'\mathbf{X}\mathbf{A} = \mathbf{A}'\Sigma_{xx}\mathbf{A} = \mathbf{I}_k \quad (3.39)$$

as long as $\mathbf{X}$ is appropriately centered and scaled. The same applies to $\mathbf{Y}$ and $\mathbf{B}$ so that $\mathbf{B}'\Sigma_{yy}\mathbf{B} = \mathbf{I}_k$. This property becomes very useful when using CCA for prediction because it ensures that the multivariate regression of the canonical variables will be well-conditioned.

Canonical Correlation Regression

In this section, CCA is put into a predictive framework. The advantage of using CCA for regression is that it works in a reduced and orthogonal space where correlations are maximized. The reduction of dimensionality conditions the inverse problem so that it is not ill-posed and provides a robust model that does not overfit the data. Thus, Canonical Correlation Regression (CCR) can be a powerful rank-reduced multivariate regression tool.

Before we can describe the regression approach, it is necessary to develop the *inverse* transformations from the canonical variables to the original variables. That is, we need to define the rotation that maps $\mathbf{U}$ to $\mathbf{X}$ and $\mathbf{V}$ to $\mathbf{Y}$. The simplest approach is to use the inverse of the *forward* rotations $\mathbf{A}$ and $\mathbf{B}$ such that

$$\mathbf{X} = \mathbf{U}\mathbf{A}^{-1} \quad \mathbf{Y} = \mathbf{V}\mathbf{B}^{-1} \quad (3.40)$$

Unfortunately, this is only applicable when the rank of $\mathbf{A}$ and $\mathbf{B}$ is $k = p = q$. If the dimensionality is reduced, as it is likely the case when p and q are large, then the canonical

weights are not directly invertible. To get around this, it is possible to use the orthogonality property

$$\mathbf{A}'\Sigma_{xx} = \mathbf{A}^{-1} \quad \mathbf{B}'\Sigma_{yy} = \mathbf{B}^{-1} \quad (3.41)$$

This gives a method for computing the inverse rotation without explicitly finding the inverse. However, the relationship is exact only when the rank is $k = p = q$. Otherwise, the transformation is the least-squares solution such that

$$\begin{aligned} \mathbf{A}'\Sigma_{xx} &= (\mathbf{U}'\mathbf{U})^{-1}\mathbf{U}'\mathbf{X} \\ \mathbf{B}'\Sigma_{yy} &= (\mathbf{V}'\mathbf{V})^{-1}\mathbf{V}'\mathbf{Y} \end{aligned} \quad (3.42)$$

Proof. First substitute $\mathbf{U} = \mathbf{XA}$ in equation 3.42 to get

$$\begin{aligned} (\mathbf{U}'\mathbf{U})^{-1}\mathbf{U}'\mathbf{X} &= [(\mathbf{XA})'(\mathbf{XA})]^{-1}(\mathbf{XA})'\mathbf{X} \\ &= [\mathbf{A}'\mathbf{X}'\mathbf{XA}]^{-1}(\mathbf{XA})'\mathbf{X} \\ &= [\mathbf{A}'\mathbf{X}'\mathbf{XA}]^{-1}\mathbf{A}'\mathbf{X}'\mathbf{X} \end{aligned} \quad (3.43)$$

Consider the case when $\mathbf{X}$ is mean-centered and scaled by the number of observations such that $\mathbf{X}'\mathbf{X} = \Sigma_{xx}$ so that 3.43 becomes

$$[\mathbf{A}'\Sigma_{xx}\mathbf{A}]^{-1}\mathbf{A}'\Sigma_{xx} \quad (3.44)$$

By applying the orthonormality property of 3.38 and $\mathbf{I}^{-1} = \mathbf{I}$, 3.43 becomes $\mathbf{IA}'\Sigma_{xx} = \mathbf{A}'\Sigma_{xx}$. The same approach can be used for the inverse transform of $\mathbf{V}$ to $\mathbf{Y}$. $\square$

The implication of the definition of the inverse CCA transform as defined in eq. (3.42) is that it provides a mapping from the canonical to the original space that can be performed even when the dimensionality of the canonical space is smaller than the original space. When this is the case, the mapping is no longer exact and becomes the coefficients of the least squares regression between the canonical and the original variables. The regression

coefficients are also the correlations between the canonical variables and the original variables that they were derived from. These coefficients are known as *loadings* because they describe how the canonical values load the original space with the inverse transformation. In cases where the dimensionality has been reduced, the loadings tend to be smoother than the canonical weights, thus making interpretation of the CCA results much easier.

Now that the inverse transformations have been developed, the regression framework may be built. The mapping from $\mathbf{X}$ to $\mathbf{U}$ is obtained from the canonical weights $\mathbf{A}$. Similarly, $\mathbf{V}$ is obtained from applying $\mathbf{B}$ to $\mathbf{Y}$. The mapping between the canonical variables $\mathbf{U}$ and $\mathbf{V}$ is obtained through multivariate least-squares regression. To predict $\mathbf{V}$ from $\mathbf{U}$, the least-squares solution to the regression coefficients is

$$\beta_{\mathbf{CC}} = (\mathbf{U}'\mathbf{U})^{-1}\mathbf{U}'\mathbf{V} \quad (3.45)$$

so that $\hat{\mathbf{V}} = \mathbf{U}\beta_{\mathbf{CC}}$. The regression coefficient matrix $\beta_{\mathbf{CC}}$ is a diagonal matrix with the canonical correlations along the diagonal (i.e., $\beta_{\mathbf{CC}} = \mathbf{\Psi}$).

Proof. The correlations between the canonical variables are given by eq. (3.11). However, because of the unit variance constraint, $\mathbf{U}'\mathbf{U} = \mathbf{V}'\mathbf{V} = \mathbf{I}_k$ and the correlation definition is simplified to

$$\text{corr}(\mathbf{U}, \mathbf{V}) = \mathbf{U}'\mathbf{V} = \mathbf{\Psi} \quad (3.46)$$

Applying the unit variance constraint to eq. (3.45) results in

$$\beta_{\mathbf{CC}} = \mathbf{I}_k \mathbf{U}'\mathbf{V} = \mathbf{U}'\mathbf{V} = \mathbf{\Psi} \quad (3.47)$$

□

Thus, the predicted canonical variables are $\hat{\mathbf{V}} = \mathbf{U}\mathbf{\Psi} = \mathbf{X}\mathbf{A}\mathbf{\Psi}$. The final step is to apply the inverse transform to the estimates of $\mathbf{V}$ such that

$$\hat{\mathbf{Y}} = \hat{\mathbf{V}}\mathbf{B}'\Sigma_{\mathbf{yy}} = \mathbf{X}\mathbf{A}\mathbf{\Psi}\mathbf{B}'\Sigma_{\mathbf{yy}} \quad (3.48)$$

Since the entire process is linear, the “cascaded” transformations may be represented by a single operation

$$\beta_{\text{CCR}} = \mathbf{A}\Psi\mathbf{B}'\Sigma_{\mathbf{Y}\mathbf{Y}} \quad (3.49)$$

and

$$\hat{\mathbf{Y}} = \mathbf{X}\beta_{\text{CCR}} \quad (3.50)$$

The major criticism about CCR is that it does not appear to satisfy a global optimization of the *prediction* of $\mathbf{Y}$ based on $\mathbf{X}$. However, since the process is a cascaded set of optimal linear transformations, then it follows that CCR optimally predicts $\mathbf{Y}$. The cascaded process can be divided in three optimal *linear* rotations:

1. An optimal rotation from $\mathbf{X}$ to $\mathbf{U}$ that ensures $\mathbf{U}$ is maximally correlated to $\mathbf{V}$.
2. An optimal least-squares mapping between $\mathbf{U}$ and $\mathbf{V}$ which is equal to the canonical correlations.
3. An optimal inverse transformation from $\mathbf{V}$ to $\mathbf{Y}$ that is exact when the dimensionality is not reduced and becomes a least-squares solution when the dimensionality is reduced.

Equation (3.50) shows that these linear transformations can be redefined as a single transformation from $\mathbf{X}$ to $\mathbf{Y}$. Since the individual transformations are optimal, the “overall” transformation β_{CCR} is also optimal.

3.1.4 Interpretation and Observations

Canonical Correlation Analysis may be interpreted in different ways. In its basic form, it is a method for data reduction and exploration of relationships between latent factors. Analysis of the canonical weights and loadings can provide insight into the nature of the factors and interpretation of the canonical variables.

The visual interpretation of CCA shown in Figure 3.3 is a path analysis that resembles a neural network diagram. Indeed, there is a close relationship between CCA and neural networks because they are both regression models based on indirect paths. However, there are also major differences. Neural networks are basically nonlinear regression models. The architecture of the neural network is very dependent on the application and is sometimes chosen to emulate some biological or physical process. In many cases, the architecture has no interpretation at all and is full of hidden layers and activation nodes. The activation nodes introduce the nonlinearity in the model. In contrast, CCR is a *linear* regression model. The canonical space can be thought of as a hidden layer but without any activation nodes. The architecture is simple and generally interpretable. There has been some research in the implementation of CCA as a neural network (Lai and Fyfe 1999). The predictive skills of neural networks and CCA have also been compared in the context of climate modelling (Tang et al. 2000). In the latter study, the neural network introduction of nonlinearity did not improve the parameter estimation significantly.

CCA can also be interpreted from an information theory perspective. In information theory, entropy is a measure of information and an optimal communication channel seeks to maximize the amount of information throughput (Shannon 1997). The maximization may be accomplished through Bayesian statistics if the probability distributions are known. The probabilities can then be used to build a channel that maximizes the mutual information between the source and the receiver. The CCA paths are analogous to communication channels. If the distribution of the variables is Gaussian, then the correlation is a measure of mutual information (Kullback 1997; Akaho et al. 1999; Becker 1996). Therefore, CCA is optimal from an information theory perspective if the variables are normally distributed.

3.2 Implementation

3.2.1 CCR Inverse Model

The CCR inverse models were built using *mean-centered* data. This is an extra *conditioning* step because it removes the mean value as an uncertain parameter. Therefore, the CCR inverse model is used to estimate *deviations* from the mean. That is, we assume that the climatological mean estimated in the ensemble does not change appreciably. This is a reasonable assumption if many samples are used or if the samples used are known to span the range of expected values. To implement the CCR inverse model, the mean of the ensemble $\mathbf{X}$ is subtracted from a new observation $\mathbf{x}$. The CCR coefficients are applied and the estimated $\mathbf{y}$ results. However, this estimate is also mean-centered so that the mean of the $\mathbf{Y}$ ensemble must be added to get the actual estimate of $\mathbf{y}$.

Inspection of eq. (3.7) reveals that the inverse of the covariance matrices must be calculated in CCR. As mentioned previously, the inverse may not exist depending on the rank r of the data. When $r < k$, it is necessary to perform a Singular Value Decomposition (SVD) of the covariance matrices and reconstruct these matrices using a truncated sequence of r singular values (Appendix B). The inverse matrix is a linear combination of the eigenvectors weighted by the reciprocal of the singular values. This truncation does not affect CCR because the number of significant correlations r_s is bounded by the “true” or “latent” rank (i.e., $r_s \leq r$). The practical difficulty is finding the appropriate truncation dimension. In this research, the sequence is truncated when the running sum of singular values totals 99.99% of the sum of all singular values. Finally, the significant correlations are determined by keeping the correlations (starting with the maximum and going in sequence towards the minimum) whose running sum is 85% of the total correlation in the data. These values were determined empirically by plotting the sum of squared errors against the kept number of dimensions and seeing where it began to level off. This is akin to analyzing a *scree* plot of PCA eigenvalues. An example plot is shown in Figure 3.4.

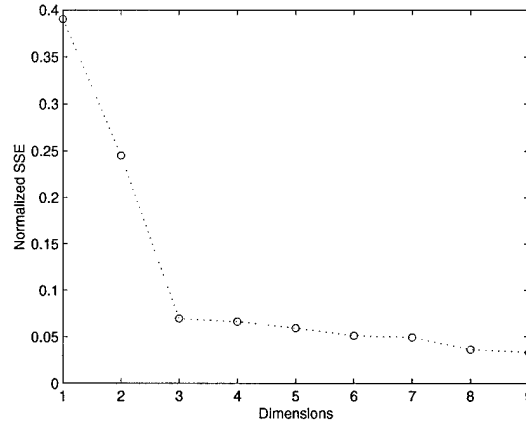


Figure 3.4: Plot of sum of squared error. In this example, the benefit of maintaining more than 3 dimensions is minimal.

The remaining issue is finding an appropriate ensemble of observations that can be used to build the CCR inverse model. This ensemble can be thought of as a *training* set. In this research, no attempt was made at finding an “optimal” ensemble. However, we can list a few ensemble traits that would be ideal:

1. *There are no errors in the observations.* Uncorrelated errors in the observations introduce variability within each data set that does not carry information. This leads to a classic signal-to-noise ratio (SNR) problem. This error introduces uncertainties in the estimated parameters and decreases precision. If there is too much noise, it may be impossible to predict a parameter of interest with any reasonable amount of certainty. Errors that are correlated (in the sense that the variables within a single data set are biased or modulated by structured noise) will skew the canonical correlations and decrease accuracy.
2. *There are an infinite number of observations.* The larger the ensemble, the more accurate the statistics. This is a consequence of the Law of Large Numbers. Also, the multivariate probabilities will tend to a Gaussian distribution because of the Central Limit Theorem. The normal distribution makes the mean and the covariance matrix

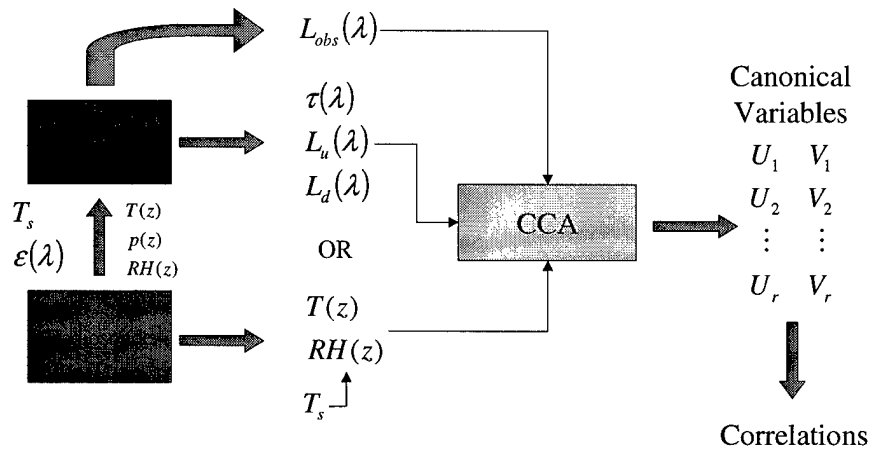


Figure 3.5: CCA implementation block diagram

sufficient statistics. Also, larger ensembles are more likely to have observations that span the range of expected physical conditions. If the database is too small, then the inverse model may not be able to extrapolate to an appropriate solution.

3. *All of the cross-set variabilities are correlated.* If all of the variability in the data sets are correlated, then the inverse mapping would be exact. Unfortunately, radiative physics ensures that some loss of information will occur in the mapping from the atmospheric state space to the measurement space. Therefore, some of the variability in the atmospheric and surface parameters of interest will not be correlated to the observations.

The first ideal situation could be achieved if a synthetic ensemble is used to build the CCR inverse model. That is, fictitious input parameters can be used in a forward model to generate simulated observations. A CCR inverse model can then be built relating the noiseless data sets. The other ideal situations are not realistic and the best we can do is approach the ideal condition as much as possible.

Figure 3.5 is a block diagram of how the algorithm is implemented. Radiosonde data are used as inputs to the MODTRAN forward model. MODTRAN was chosen because it

has been extensively validated over the course of the last 30 years. Furthermore, recent enhancements in MODTRAN allow the user to enter spectral surface emissivities, scale atmospheric profiles with a column factor, estimate hemispherical downwelled radiance from a single run more accurately, etc. Several input parameters to MODTRAN were varied to generate many observed radiance spectra. The goal of the CCA inverse model was to relate these observed radiance spectra to the MODTRAN-predicted atmospheric spectra (i.e., $\tau(\lambda)$, $L_u(\lambda)$, and $L_d(\lambda)$) or the atmospheric profiles used as inputs to MODTRAN. The estimate of the vertical profiles could then be used as a final product or as an input to MODTRAN to estimate the atmospheric spectra. In some cases, the CCA inverse model was used to estimate surface temperatures directly.

The CCR inverse models used in this research were built with three different atmospheric databases (with the exception of one of the experiments where synthetic profiles were used). These databases were generated with MODTRAN using radiosonde measurements obtained from the Forecast System Laboratory (FSL) of the National Climatic Data Center (NCDC), the CAMEX3 field campaign of the National Polar-orbiting Operational Environmental Satellite System (NPOESS) Aircraft Sounder Testbed-Interferometer (NAST-I) at Wallops Island, and the Space Science and Engineering Center (SSEC) at the University of Wisconsin-Madison. Figure 3.6 shows the geographic coverage of these measurements. These data were used because they provided different climates, thus allowing testing of the inverse model under various conditions. Choosing actual radiosonde profiles over synthetic profiles was an attempt at characterizing the “natural” variability in the atmosphere as measured by real data. Conversely, a synthetic database runs the risk of not being truly representative of the atmosphere.

The radiosonde databases included variations in temperature and water vapor profiles, surface elevation, time of day, date, and geographical coordinates. The only two parameters explicitly handled by the CCR inverse model (with respect to the atmospheric database) were the temperature and water vapor profiles. The rest of the variation acts as “noise”.

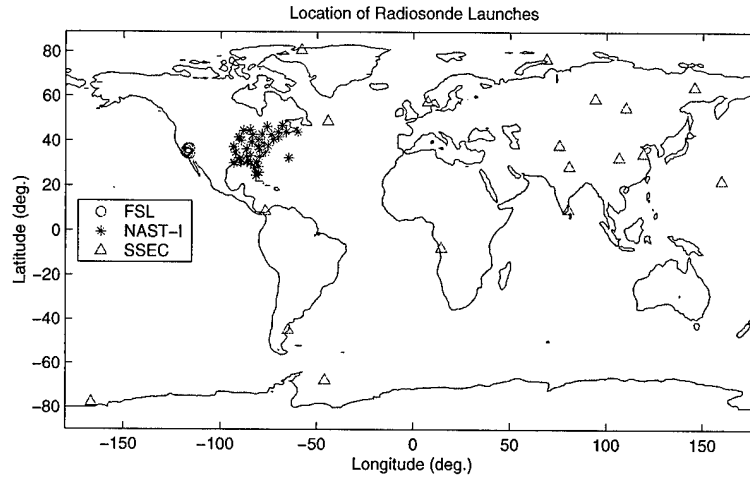


Figure 3.6: Geographic coverage of radiosonde measurements used to build the inverse model.

Thus, changes in the propagation path length due to changes in surface elevation introduce uncertainty in the retrieved parameters. Changes in solar geometry are not significant for LWIR and some MWIR observations. Finally, ozone measurements were not available with the radiosonde databases. Thus, cases implementing CCR inverse models built with radiosonde data do not account for changes in ozone concentration. This also introduces errors in the retrieved parameters.

3.2.2 Temperature and Emissivity Separation

The Temperature and Emissivity Separation (TES) algorithm was implemented as discussed in Section 2.2.3 with one difference: a new empirical relationship between the maximum-minimum difference (MMD) and the minimum emissivity ϵ_{min} was derived. The empirical relationship reported by Gillespie et al. (1999) was obtained from 86 laboratory reflectance measurements of rocks, soils, vegetation, snow, and water. These reflectance measurements were resampled to match the spectral response of the ASTER sensor. The empirical relationship is very much dependent on the data set and the sensor for which it was generated. Because the reported relationship is optimized for ASTER and for certain target classes,

Table 3.1: JHU spectral library reference table.

Group	Class	Filename	Number of Class Spectra	Number of Total Spectra
1	Igneous (coarse)	IGN_CRS.SLI	34	34
2	Igneous (fine)	IGN_FN.SLI	33	67
3	Lunar	LUNAR.SLI	17	84
4	Manmade1	MANMADE1.SLI	14	98
5	Manmade2	MANMADE2.SLI	19	117
6	Metals (coarse)	META_CRS.SLI	25	142
7	Metals (fine)	META_FN.SLI	29	171
8	Meteor	METEOR.SLI	59	230
9	Minerals1	MINERALS.SLI [†]	54	284
10	Minerals2	MINERALS.SLI	43	321
11	Minerals3	MINERALS.SLI	45	372
12	Minerals4	MINERALS.SLI	51	423
13	Minerals5	MINERALS.SLI	59	482
14	Minerals6	MINERALS.SLI	63	545
15	Minerals7	MINERALS.SLI	11	556
16	Sediments (coarse)	SED_CRS.SLI	15	571
17	Sediments (fine)	SED_FN.SLI	13	584
18	Snow	SNOW.SLI	4	588
19	Soils	SOILS.SLI	25	613
20	Vegetation	VEG.SLI	3	616
21	Water	WATER.SLI	3	619

[†]Class file split into seven groups for data handling purposes.

it was necessary to implement the TES algorithm with a new empirical relationship that had a broader scope. The new relationship was determined from a larger data set of laboratory spectra. The John Hopkins University spectral library included with ENVI contains 619 spectra of natural and man-made objects—including those used for ASTER (Table 3.1). These spectral measurements were made by J.W. Salisbury and are considered a standard in the geology and remote sensing communities. The larger data set was used to make the relationship more robust and not limited to a particular set of materials. No spectral resampling of the reflectance curves was performed in this analysis.

The relationship was built using a similar model to that implemented for ASTER. The model is

$$Y = \beta_0 + \beta_1 X + \beta_2 X^\gamma + \epsilon \quad (3.51)$$

where ϵ is some random error assumed to come from the unit normal distribution. This model retains a linear component in addition to the power law term. These parameters can be estimated by

$$\hat{Y} = b_0 + b_1 X + b_2 X^g \quad (3.52)$$

where b_0 , b_1 , b_2 , and g are estimates of β_0 , β_1 , β_2 , and γ , respectively. To perform a linear regression using this equation, the value for g must be determined. This was done iteratively, starting with the value 0.737. A linear regression was then performed for each intermediate estimate of g . Estimates were varied until the lack of fit was minimized.

The rest of this section describes the analysis and resulting empirical relationship that is implemented in the TES algorithm. It is assumed that the reader has some familiarity with regression analysis. For more background information, refer to Draper and Smith (1998). The initial results obtained from the regression using the value of $g = 0.737$ and the model used by Gillespie are shown in Figure 3.7. While there appears to be a relatively good fit, several points fall outside the prediction intervals. There are also points that appear to be extreme outliers (circled on the plot). Table 3.2 shows the analysis of variance (ANOVA), t-tests, standard error, and multiple squared correlation coefficients for the regression. It is clear that the regression and the regression parameters are significant. On the other hand, the lack of fit is also significant. This is probably due to the fact that the estimate of pure error is not an accurate one since there are only 10 degrees of freedom compared to 607 degrees of freedom associated with the lack of fit. The conclusions from the ANOVA were validated with an analysis of the residuals. The residuals are shown in Figure 3.8. Considering the results from the ANOVA, it is not surprising that the residuals do not appear to be normal, which indicates that the F-statistic calculations in the ANOVA are not accurate. There were also several large values for the residuals (all the way to 0.5!).

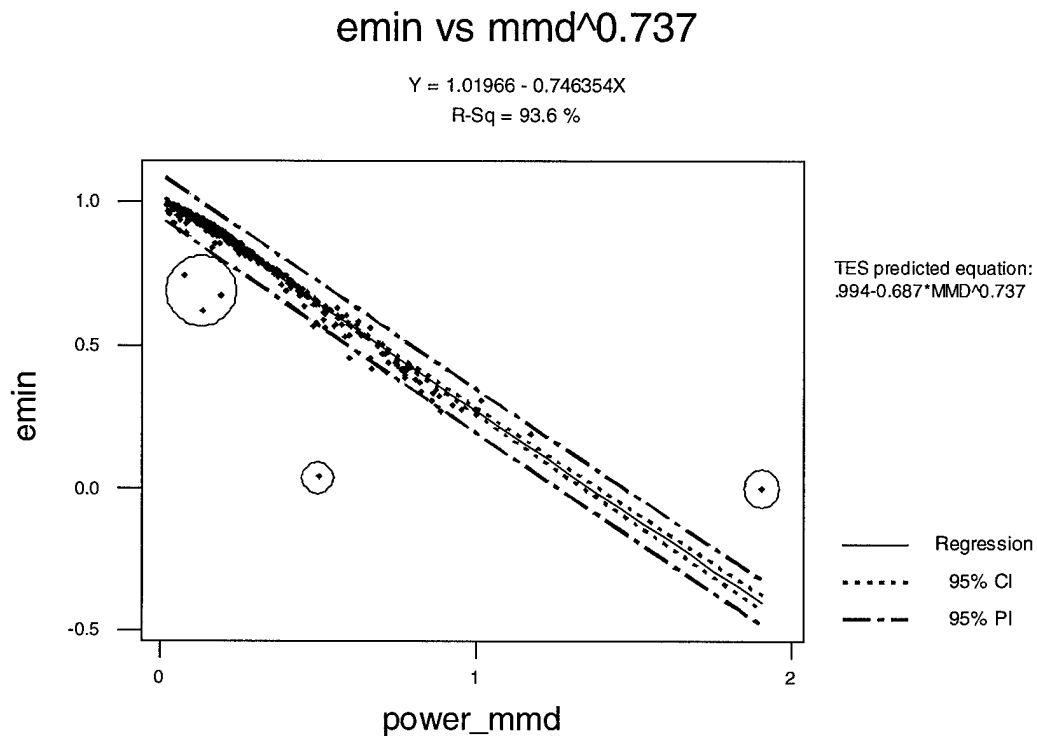


Figure 3.7: Plot of data and regression fit of the original empirical relationship.

Source	DF	SS	MS	F	P
Regression	1	13.404	13.404	9020.38	0.000
Residual Error	617	0.917	0.001		
Lack of Fit	607	0.917	0.002	3911.51	0.000
Pure Error	10	0.000	0.000		
Total	618	14.320			

Predictor	Coef	StDev	T	P
b_0	1.00487	0.00129	781.01	0.000
b_1	-0.09876	0.03071	-3.22	0.001
b_2	-0.68456	0.02925	-23.41	0.000

$$S = 0.03855 \quad R^2 = 93.6\% \quad R^2(\text{adj}) = 93.6\%$$

Table 3.2: Regression results using $g = 0.737$ and no repeats.

Residual Plots for Power-MMD

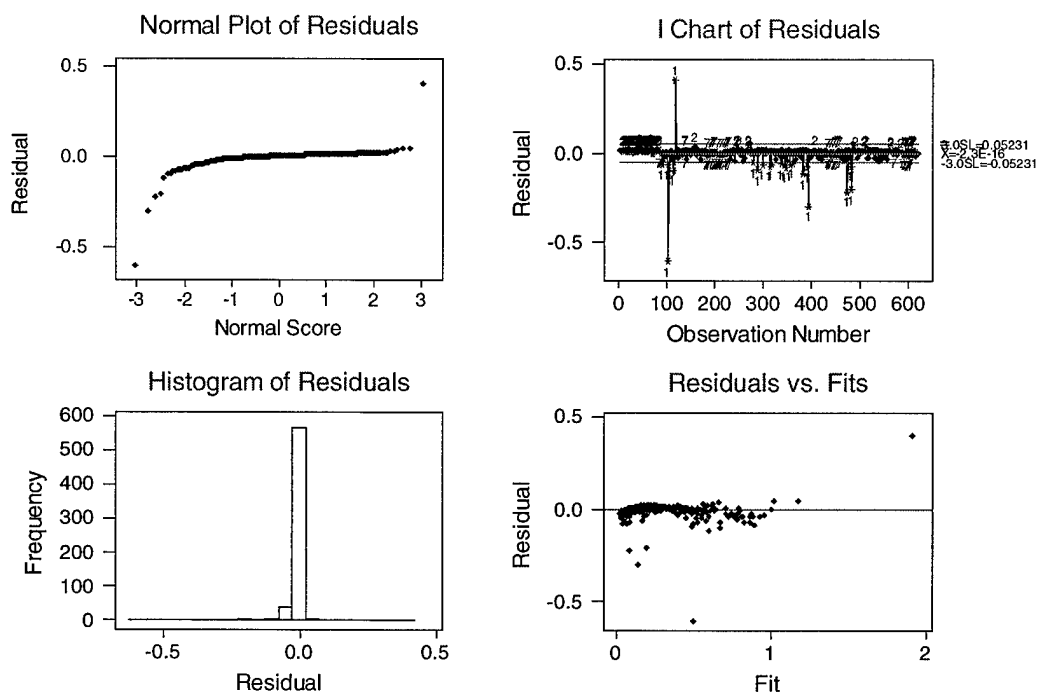


Figure 3.8: Residual plots for the original empirical relationship.

To obtain better estimates, the six observations that resulted in the largest errors were taken out of the data. The removed observations corresponded to the following materials in the spectral library:

- Point 101: Aluminum metal (Metal 0384UUUALM)
- Point 116: Copper metal (Metal 0682UWCOP)
- Point 382: Kyanite Al_2SiO_5 (Neosilicates (Isolated Tetrahedra)(AlSiO_5 Group)k.1)
- Point 393: Magnetite $\text{Fe} + 2\text{Fe}_2 + 3\text{O}_4$ (Spinel Group; magnet.1)
- Point 473: Pyrite FeS_2 (pyrite.1)
- Point 482: Pyrrhotite $\text{Fe}(1-x)\text{S}$ (Pyroph.1)

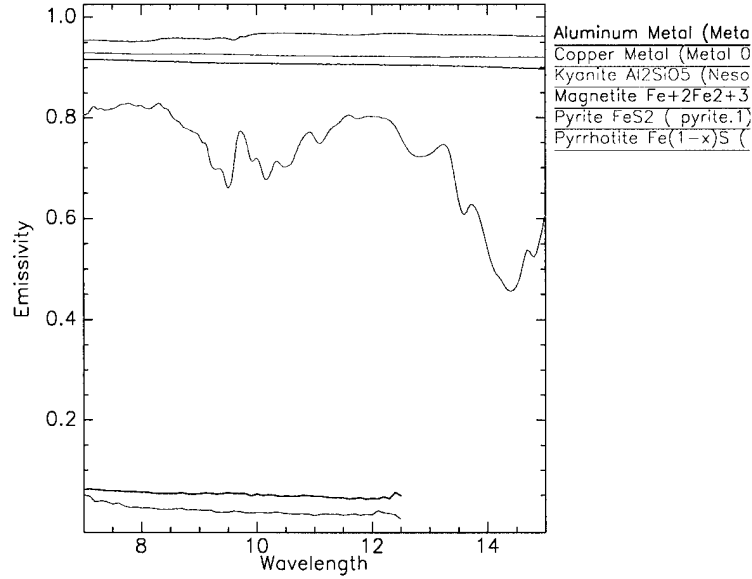


Figure 3.9: Spectral emissivity of extreme outliers.

The point number (e.g., Point 101) refers to the spectrum number assigned in Table 3.1. The emissivity curves of these materials are shown in Figure 3.9. Three of these materials exhibit very low emissivities. Also, the spectrally flat curves have relatively low emissivities (compared to a blackbody). Thus, the observations that do not fit the model correspond to materials that have a low minimum emissivity and a spectrally flat emissivity curve. These materials are “unusual” in the sense that they do not conform to the typical physical characteristics assumed when the model was devised. Deviations from these physical characteristics explain why the relationship between the minimum emissivity and the MMD do not fit the model. Further analysis indicated that up to 23 observations needed to be removed from the data set. (The observations removed from the data set were: 89, 100, 101, 111, 116, 273, 287, 297, 312, 315, 336, 342, 345, 354, 360, 382, 391, 393, 473, 480, 482, and 485 as referenced in Table 3.1). All of these observations came from manmade, minerals, and metal sources, which are the most likely to deviate from the model.

Source	DF	SS	MS	F	P
Regression	2	9.9256	4.9628	45461.86	0.000
Residual Error	593	0.0647	0.0001		
Lack of Fit	537	0.0604	0.0001	1.46	0.040
Pure Error	56	0.0043	0.0001		
Total	595	9.9903			

Predictor	Coef	StDev	T	P
b_0	1.00487	0.00129	781.01	0.000
b_1	-0.09876	0.03071	-3.22	0.001
b_2	-0.68456	0.02925	-23.41	0.000

$$S = 0.01045 \quad R^2 = 99.4\% \quad R^2(\text{adj}) = 99.3\%$$

Table 3.3: ANOVA results for the new fitted model.

To get a better estimate of the pure error, the independent variable (MMD) was rounded to 4 decimal places. This introduced some approximate repeats in the regression data. In addition, the linear term was added to the model as shown in eq. (3.51).

The next step was to obtain an appropriate estimate of g . The iteration was in the form of a binary search on the exponent. The search was constrained to values for g between 1.000 and 0.737 with the objective of minimizing mean-squared error (MSE) and F(lack of fit) and maximizing R^2 . This approach lead to an exponent of 0.818. Figure 2.18 on page 74 shows the fit with this model. The fitted equation was:

$$\varepsilon_{min} = 1.005 - 0.099MMD - 0.685MMD^{0.818} \quad (3.53)$$

Note that the fit is generally better and that there are no major outliers. However, there are still points that fall outside the 95% prediction intervals.

The ANOVA and other quantitative results for the new model are shown in Table 3.3. All of the parameters in the linear regression were found to be significant. The lack of fit is considerably lower, but it is still significant at a risk level of 0.04. The standard error is lower by more than a factor of 3, and the adjusted multiple squared correlation coefficient

is higher at 99.3%. Unfortunately, the validity of the ANOVA is slightly questionable. The residual plots shown on Figure 3.10 show that the residuals still deviate from a normal distribution. Attempts at fitting the data to both quadratic and exponential models did not reduce this pattern, and actually resulted in larger errors. Also, there is still a pattern in the residuals as shown in the residuals vs. fits plot. This pattern is due to the large concentration of materials that have high emissivities (and low MMD). Because the model fits well in this region, the residuals are generally lower than those corresponding to lower fitted values of emissivity. The “Run Chart” shows that there is a relatively large number of extreme values, which is partly due to the large amount of data used in this analysis. This makes extreme values much more likely. The serial correlation of the residuals was tested using the Durbin-Watson statistic. The value for these data was 1.538, which is lower than d_U at a risk level of 2.5% (this is a two-tailed test so 2.5% is used so that the total risk is 5%). This suggests that there is positive serial correlation in the data. Figure 3.11 shows a Lag-1 plot of the residuals. This plot, however, does not reveal any apparent correlation in the data. The disparity may be due to the fact that the value for d_U was obtained by extrapolating the value at 200 samples to 596. Because of this large extrapolation, the Durbin-Watson statistic may be biased and inferences about the serial correlation should probably be made based on the Lag-1 plot. In conclusion, there does not appear to be a significant correlation in the residuals that would alter the conclusions obtained from the other residual plots.

The 95% confidence intervals for the linear regression coefficients are:

$$\begin{aligned} 1.002 &\leq \beta_0 \leq 1.007 \\ -0.159 &\leq \beta_1 \leq -0.038 \\ -0.742 &\leq \beta_2 \leq -0.627 \end{aligned} \tag{3.54}$$

The β_1 parameter exhibits the largest variation. This is consistent with the regression analysis results showing that β_1 is the least significant of the parameters. These confidence

Residual Plot of Complex Model

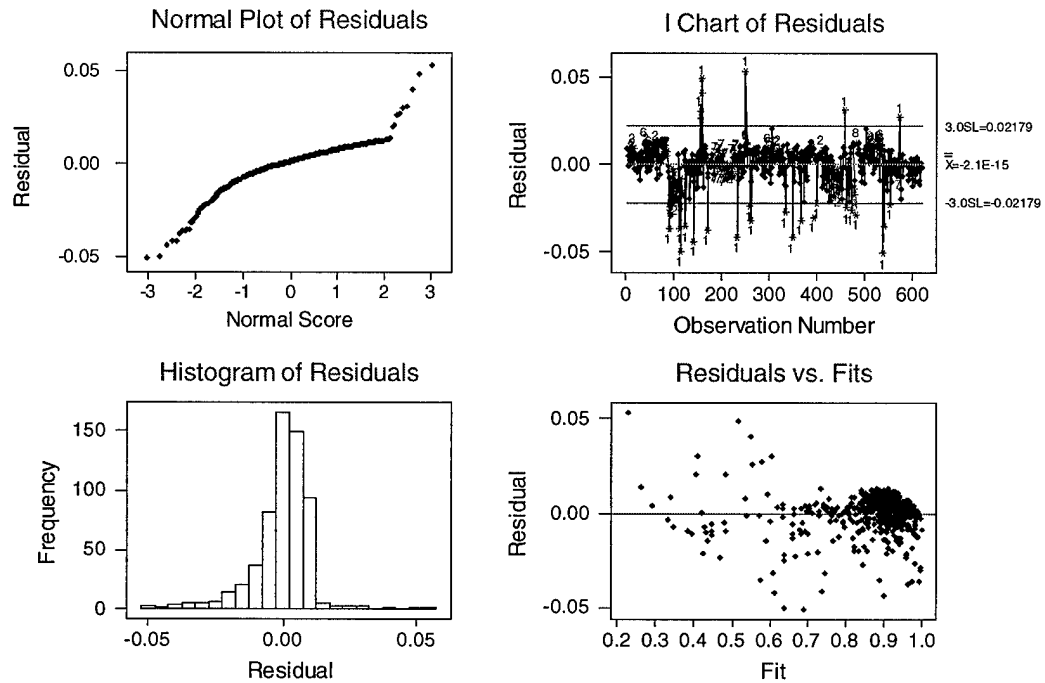


Figure 3.10: Residual plots for the new fitted model (with approximate repeats).

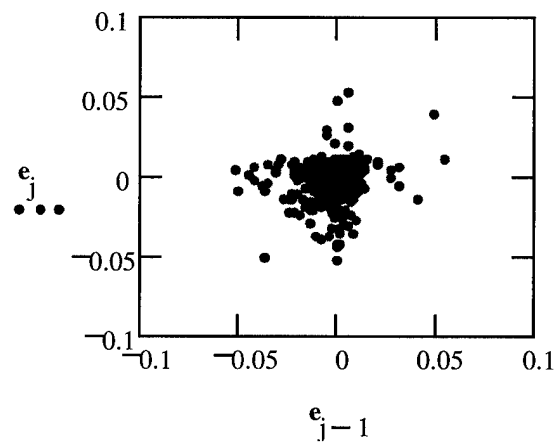


Figure 3.11: Lag-1 plot showing no correlation in the residuals.

intervals are based on independent estimates of the parameters and t-values. Thus, these intervals form a three-dimensional cube in the parameter space. The volume of this cube is the actual confidence region bounded by these intervals. In general, this is an overestimated figure. The ratio of the rectilinear region to the true (more ellipsoidal) region is given by the square root of the determinant of the variance-covariance matrix of the parameters. the variance-covariance matrix of the parameters is obtained from

$$\Sigma_{bb} = (\mathbf{X}^T \mathbf{X})^{-1} MSE \quad (3.55)$$

For this data, the ratio of the true confidence region to the rectilinear region is

$$\sqrt{|\Sigma_{bb}|} = 3.392 \cdot 10^{-8} \quad (3.56)$$

which indicates that the parameters are highly correlated, thus defining a very narrow region in the 3-D parameter space.

The quantification of confidence intervals about a nonlinear parameter can be performed in a way similar to that used for linear parameters. The main difference is in the resulting sum-of-squares function. In general, the sum-of-squares function for a nonlinear model is

$$S(\beta) = \sum_{i=1}^n [\mathbf{y}_i - f(\mathbf{x}_i; \beta)]^2 \quad (3.57)$$

where β is a vector of the parameters in the model, n is the number of observations, and $\mathbf{x}$ and $\mathbf{y}$ are the (possibly multivariate) predictor and response variables, respectively (Draper and Smith 1998). In the TES nonlinear model, the predictor and response variables are univariate (note that the way the regression was carried out, the *linear* model had multiple predictor variables; mainly *MMD* and *MMD $^\gamma$*). The confidence intervals may be constructed by finding all the values of the model parameters β that satisfy

$$S(\beta) = S(\mathbf{b}) \left\{ 1 + \frac{p}{n-p} F(p, n-p, 1-\alpha) \right\} = S_q \quad (3.58)$$

where $\mathbf{b}$ are the estimates of β , $F(\nu_1, \nu_2, 1-\alpha)$ is the single-sided F-statistic, p is the number of model parameters, and α is the risk factor. For a 95% confidence interval, we would use $\alpha = .05$.

There are $p = 4$ parameters in the TES nonlinear model. Thus, the confidence interval about β is really a four-dimensional *region*. To make the problem more tangible and easier to illustrate, it is possible to work with the *projection* of the multidimensional confidence region onto a two-dimensional space. The simplest way to do this is to hold two parameters constant and build a two-dimensional ellipsoid confidence region for the other two parameters. Since the parameter of interest is γ (hereafter referred to as β_3), this will be one of the parameters of the two-dimensional region. The other parameter that will be allowed to vary is β_1 . This parameter was chosen because it had the widest confidence interval in eq. (3.54). The sum-of-squares function then becomes

$$S(\mathbf{b}) = \sum_i n [y_i - 1.005 + .099x_i + .685x_i^{.818}]^2 = .065 \quad (3.59)$$

and from eq. (3.58)

$$S_q = .065(1 - .007 \cdot 2.387) = .064 \quad (3.60)$$

If we assume $\beta_0 = b_0$ and $\beta_2 = b_2$, then the confidence region is defined by all the points β_1 and β_3 that satisfy

$$\sum_{i=1}^n [y_i - 1.005 + \beta_1 x_i + .685x_i^{\beta_3}]^2 = .064 \quad (3.61)$$

The equation can be expressed as a quadratic function of β_1 :

$$A\beta_1^2 + B\beta_1 + C = .064 \quad (3.62)$$

where A , B , and C are functions of β_3 :

$$\begin{aligned} A &= \sum_i x_i^2 \\ B &= \sum_i (2x_i y_i - 2.010x_i + 1.370x_i^{\beta_3+1}) \\ C &= \sum_i (1.010 + 2.010y_i + y_i^2 + 1.370y_i x_i^{\beta_3} - 1.377x_i^{\beta_3} + .4692x_i^{2\beta_3}) \end{aligned} \quad (3.63)$$

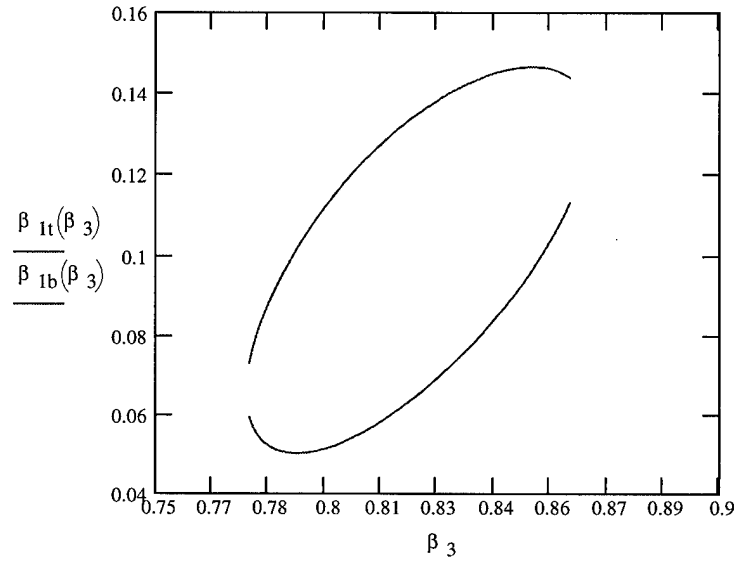


Figure 3.12: Two-dimensional confidence region of nonlinear parameters.

We can find the quadratic solution of β_1 and express β_1 as an explicit function of β_3 . By entering a range of values of β_3 , a confidence region is developed. Figure 3.12 illustrates this region. This analysis is consistent with the one used to derive eq. 3.54 because the projection of the ellipse unto the β_1 axis results in a similar confidence interval. The projection of the confidence region unto the β_3 axis shows that the 95% confidence interval for γ is

$$.776 \leq \gamma \leq .865 \quad (3.64)$$

These confidence intervals suggest that the difference between the published TES MMD regression line and the model given in eq. (3.53) is statistically significant, which is expected since the linear term is included in the new model. However, this conclusion is based on holding the β_0 and β_2 parameters constant at the center of their respective confidence intervals. Thus, the ellipsoid in Figure 3.12 is a potentially underestimated projection. Several values for β_0 were tested to determine the effect on the two-dimensional confidence region but no significant changes were observed.

The results from the ANOVA and the analysis of the residuals suggest that the TES model may not be adequate for the entire population of materials on the Earth. However, it does provide reasonable results for a large gamut of materials that *is* representative of the Earth's composition. The maximum errors in emissivities were about 5%. An error of this magnitude leads to a temperature error of about 2.5 °K. This represents a worst case scenario if proper atmospheric compensation and negligible sensor noise are assumed. In reality, the large spatial scale measurements made by high-altitude aircraft and satellite platforms result in hyperspectral pixels that consist of a mixture of spectral emissivities. These mixed pixels tend to "average-out" unusual emissivity features, resulting in an effective emissivity that adheres more closely to the phenomenology exploited by the TES model. Thus, it is expected that the retrieved emissivities and temperatures from this model would have an error lower than the 5% and 2.5°K.

Another option is to develop a model for each material class. Unfortunately, this requires *a priori* knowledge of the scene objects, which is often not available. Besides, the typical application of these hyperspectral sensors is to identify unknown targets to begin with. However, it may be possible to calculate a model for two or three broad classes that may be separable without having *a priori* knowledge of the materials. For example, vegetation may be identified by a ratio of two spectral channels (i.e., Normalized Difference Vegetation Index or NDVI). These "vegetation" pixels would be processed using one model while the rest of the image is processed by another model.

In summary, an empirical model relating the maximum-minimum difference of spectral emissivity curves measured by hyperspectral sensors and the true minimum emissivity value has been developed using standard regression analysis. The nonlinear aspect of the power law coefficient was resolved by performing a binary search which minimized the lack of fit and standard error from the linear regression ANOVA. The model yields reasonable results when applied to a large spectral library. Because of the broader range of materials for which this model applies, this new empirical relationship was implemented in this research.

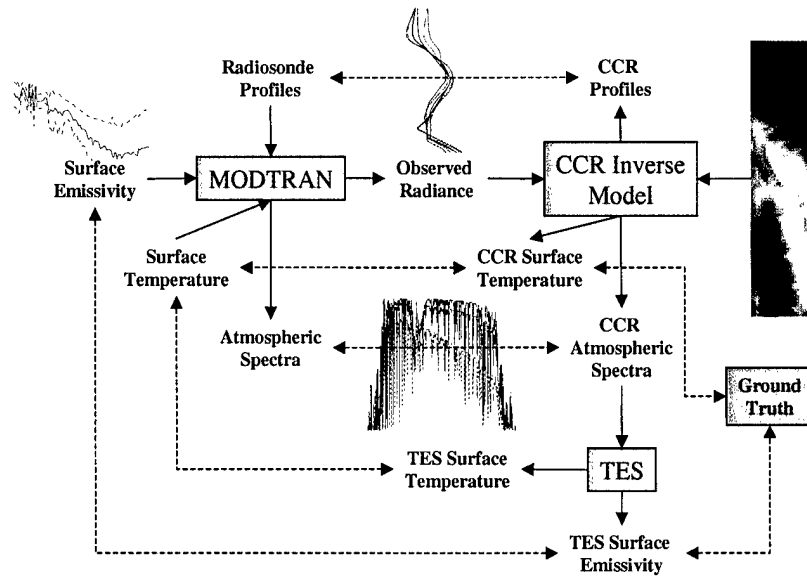


Figure 3.13: Schematic of test and validation approach.

3.3 Test and Validation

The ultimate goal of this research is to show that it is feasible to retrieve accurate estimates of land surface temperature and emissivity from remotely sensed infrared imagery. Section 3.1 described the approach used in this research to estimate these parameters. This section covers the methodology and data used to test and validate the approach. This methodology is summarized in Figure 3.13. Standard MODTRAN atmospheric profiles or radiosonde data were used as inputs into MODTRAN to generate simulated atmospheric and sensor spectra. These spectra were then used to build the CCR inverse models. Two kinds of models were built: (1) models inverting observed spectra to atmospheric optical parameters (i.e., τ , L_u , and L_d), and (2) models inverting observed spectra to physical parameters (i.e., surface temperature, temperature profiles, and water vapor profiles). The dashed double-headed arrows in Figure 3.13 indicate parameters that were compared to validate the inverse model.

When estimates of the atmospheric spectra are obtained from CCR, the sensor radiance can be solved using eq. (2.5):

$$\hat{L}_s(\lambda) = \frac{L(\lambda) - \hat{L}_u(\lambda)}{\hat{\tau}(\lambda)} \quad (3.65)$$

where $\hat{\cdot}$ denotes an estimated parameter and $\hat{L}_s(\lambda)$ is the estimated surface-leaving radiance. $L_s(\lambda)$ includes the reflected downwelled radiance component. That is,

$$L_s(\lambda) = \varepsilon(\lambda)L_{BB}(\lambda, T_s) + [1 - \varepsilon(\lambda)]L_d(\lambda) \quad (3.66)$$

This estimated surface-leaving radiance and downwelled radiance are then used as inputs to the Temperature and Emissivity Separation (TES) algorithm.

The estimated values were compared to the MODTRAN-generated spectra or to the vertical profiles used as inputs into MODTRAN. This process was done using different spectral configurations and bandpasses. Finally, the CCR inverse models were applied to multispectral thermal images from the MODIS Airborne Simulator (MAS) and MODIS and ASTER (MASTER) airborne sensors. The retrievals obtained from these data sets were compared to field measurements made coincident with the image acquisition.

3.3.1 Simulations

Simulated data are ideal for algorithm development because the experimental variables are easier to control. In addition, validation is less ambiguous than with real data because the algorithm retrievals can be compared to an *exact* value that was controlled in the experiment.

All of the CCR inverse models are built based on simulated MODTRAN spectra ($L(\lambda)$). These spectra are related to input parameters that are based on actual measurements (e.g., radiosonde profiles) or synthetic profiles (e.g., MODTRAN standard atmospheric models). Because there are no probability distributions governing the CCR inverse model, it is not possible to define confidence intervals on the retrieved parameters. Therefore, the performance of the inverse model was measured by calculating the RMS difference between

the model inputs and the parameters retrieved with the CCA inverse model. The RMS difference is the square-root of the mean-squared error (MSE) and is obtained from

$$\mathbf{y}_{RMS} = \sqrt{\sum_{i=1}^n (\mathbf{y} - \hat{\mathbf{y}})^2} \quad (3.67)$$

where n is the number of observations and $\mathbf{y}_{RMS}$ is a $q \times 1$ vector of RMS residuals. The RMS error is a biased estimate of the standard deviation. Thus, it describes how much error is expected 68.3% of the time.

Cross-validation of the CCR inverse models was also performed with simulated data. To do this, half of the MODTRAN runs were used to build the CCR inverse model. The model was then applied to the other half of the MODTRAN runs and RMS values were computed.

3.3.2 MODIS Airborne Simulator (MAS)

A useful source of data is the MODerate-resolution Imaging Spectrometer (MODIS) Airborne Simulator (MAS). This is an airborne sensor mounted on the NASA ER-2 high-altitude aircraft. The sensor is a breadboard of the Terra MODIS sensor. The MAS sensor is *not* a hyperspectral sensor because of the small number of bands. There are 9 longwave bands and 15 midwave bands. Nevertheless, this may be the only (relatively) high spectral resolution data from space available in the near-term (i.e., MODIS and ASTER onboard Terra). Since one of the goals of this research is to show the extendibility of the approach to spaceborne sensors, the MAS data were worthy of consideration in the development of the algorithms. That way, when MODIS data are available, it should be relatively easy to process the new data. Reported thermal noise values for the LWIR bands range between 0.09 to 2.00 °K (King et al. 1996). The noisy bands are channels 49 and 50, which correspond to 13.72 μm and 14.17 μm .

The MAS data is provided freely by NASA Goddard Distributed Active Archive Center (DAAC) in HDF format. The structure of the HDF data is specified in the “Level-1B Data

User's Guide" (Gumley 1994). ENVI has the capability to read this particular format. There is also a free IDL widget program called SHARP that can read the MAS HDF data. This was developed by Liam Gumley from the University of Madison-Wisconsin.

3.3.3 MASTER

The MODIS and ASTER (MASTER) sensor was developed as a breadboard sensor to validate the algorithms planned for the MODIS and ASTER sensors onboard the Terra satellite (Hook et al. 2000). It has been flown in a King Air Beachcraft B200 and DC-8 with plans for operations on the NASA ER-2. The system is a line scanner with a Gregorian telescope and uses diffraction grating spectrometers. There are four spectrometers covering the visible-near infrared, shortwave infrared (SWIR), midwave infrared (MWIR), and longwave infrared (LWIR). The LWIR focal plane array (FPA) is a Mercury-Cadmium-Telluride (HgCdTe) array with a cooled linear-variable filter. The LWIR FPA has 10 bands covering the region between 7.7 and 12.9 μm with a nominal $\text{NE}\Delta\text{T}$ of 0.3°K. The spectral resolution of these bands is about 0.5 μm . The FPA read-out is processed by a special set of 16-bit A/D converters. These high dynamic-range converters actively track the DC level detector signal, thus compensating for temporal thermal drifts (Hook et al. 2000).

3.3.4 SEBASS

SEBASS is an airborne infrared hyperspectral sensor that operates in the Mid-Wave Infrared (MWIR) and the Long-Wave Infrared (LWIR) atmospheric windows. Light from the telescope is imaged on the spectrograph entrance slit, then is split by a dichroic filter into wavelengths shorter and longer than 6.5 μm . The dispersed light is re-imaged by two prism spectrographs, one for MWIR and one for LWIR. The spectral range in the MWIR is between 2.9 μm and 5.2 μm . In the LWIR, the spectral range is between 7.5 μm and 13.6 μm . These regions are distributed over 128 spectral channels. The spectral resolution of the sensor in the LWIR is not constant, as shown in Figure 3.14. These plots show the

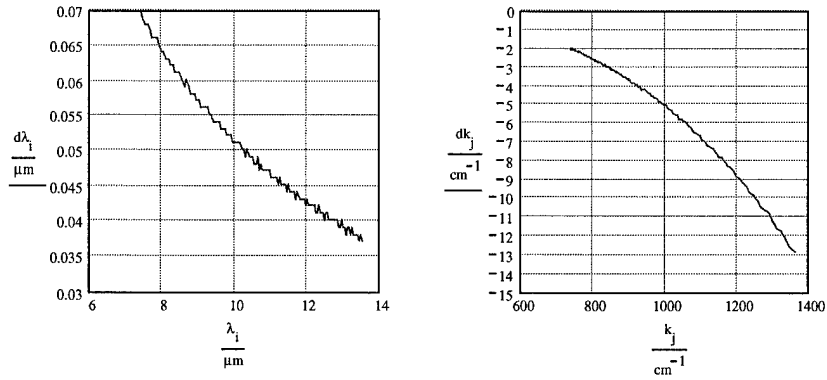


Figure 3.14: Spectral intervals for the SEBASS LWIR band.

intervals between the band centers as a function of wavelength or frequency. The sensor operates in a pushbroom mode with a swath defined by an array of 128 pixels with an Instantaneous Field of View (IFOV) of 1.1 milliradians per pixel. Thus, the FOV of the sensor is approximately 8.1° . The observed radiance is dispersed into 128 spectral bins. The 128 by 128 array is then scanned over the Earth's surface by the aircraft's motion. This generates a hyperspectral image cube that is band-interleaved by pixel (BIP) so that the depth is dictated by the number of frames collected over the flight path.

3.3.5 Experiments

A series of experiments using simulated and thermal imagery were carried out for the testing and development of the inversion technique. This section describes the experimental setup for each of these experiments, which are presented in chronological order. The results are outlined in Chapter 4.

Experiment #1

The initial test of the CCA inverse model was to determine whether the correlations between the observed spectra and atmospheric parameters were large enough for accurate

parameter estimation. To do this, 216 spectra were generated with MODTRAN 4.0. This was done with a 3-factor experimental design with no repeats. The factors were the vertical temperature profile, the vertical relative humidity profile, and the total amount of ozone. There were six different temperature and humidity profiles. There were also six different levels of ozone. This resulted in $6^3 = 216$ observations. No repeats were measured because the model is a physical model and the results are not random variables. The temperature and humidity profiles used were the default profiles for the six model atmospheres included in MODTRAN. These atmospheres are:

1. Tropical
2. Mid-latitude Summer
3. Mid-latitude Winter
4. Subartic Summer
5. Subartic Winter
6. 1976 U.S. Standard Atmosphere

The temperature, relative humidity, and ozone profiles for these models are shown in Figure 2.10. The profiles for temperature and relative humidity were used as radiosonde data so they could be “mixed” in the factorial design. Because the pressure and CO_2 profiles do not vary greatly, these were not a controlled factor. The radiosonde data contained the pressure and CO_2 profiles that corresponded to the model atmosphere where the temperature profiles were being extracted from. The total ozone concentration in the column of air was varied by adjusting the O3STR variable in CARD 1A of the Tape 5 file. The levels were 0.5, 1.0, 1.5, 2.0, 2.5, and 3.0 times the default value for the input model atmosphere. All of the runs were performed using a surface temperature of 300 °K. The output from the MODTRAN model is the simulated observed radiance for a sensor at an altitude of 100 km. The model also provides the spectral transmission and upwelled radiance based on the input parameters. The bandpass for the runs was between 7.34 μm and 13.57 μm

(737 cm^{-1} to 1362 cm^{-1} at 5 cm^{-1} steps). All of the observations were made assuming a nadir sensor-target geometry (i.e., the sensor is located at zenith).

Another goal of this experiment was to develop a scaling scheme of the observed spectra that would force CCA to use spectral features that are independent of the surface emission. This is because the observed radiance is typically dominated by the surface emission. Therefore, significant error may be introduced in the retrieval of *atmospheric* parameters because of uncompensated biases due to the surface temperature.

Experiment #2

There were several goals for this experiment:

1. To obtain temperature and water vapor retrievals with CCA inverse models built with radiosonde data.
2. To show that CCA is able to separate surface and atmospheric emission effects when these are allowed to vary in the model-building phase.
3. To couple the CCR inverse model with the TES algorithm.
4. To apply the CCR inverse model to real thermal imagery.

The CCA inverse models built in this experiment were based on the radiosonde data described in Section 3.2.1. Each radiosonde profile introduced variations in air-surface boundary layer temperatures, temperature profiles, water vapor profiles, altitude, surface elevation, surface latitude and longitude, time of day, and date. The surface temperature was set equal to the temperature of the lowest radiosonde level. Also, the radiosonde profiles were resampled to a common pressure altitude grid for each data set. Table 3.4 gives summary statistics of the radiosonde data. The variability in the global database is significantly higher than for the other two test cases. The surface temperatures for these data ranged between -42.3°C and 36.7°C with the SSEC data being the coldest and the NAST-I being the hottest.

Table 3.4: Description of radiosonde used for CCR

Dataset	N ^a	Geographic Coverage	Time Span	$\sigma_{T_{air}}$ ^b	σ_T ^c	σ_{wv} ^d
SSEC	117	Worldwide	1963-1972	19.00	11.23	16.06
FSL	192	34-38°N, 115-119°W	1995,1999	5.09	4.55	4.94
NAST-I	3,310	East Coast, U.S.A.	Jul-Sep 1998	5.35	3.71	12.33

^aNumber of profiles^bStandard deviation of surface-air boundary layer °K^cAverage standard deviation of temperature °K^dStandard deviation of column water vapor (mm)

The CCR inverse models built with the radiosonde data were used to retrieve parameters from MAS thermal imagery of Death Valley (Fig. 3.15) collected on 4 March 1997 (Flight 97-063/Track 2). This is a nighttime track so that self-emission and atmospheric radiation are the only relevant terms in the radiative transfer. An LST measurement was made by Wan (1999) coincident with the overflight. LST measurements of 18.7°C and 18.5°C were made via a thermal infrared (TIR) thermometer and a thermistor 1 mm beneath the surface, respectively. These measurements have an uncertainty of $\approx 0.5^\circ\text{C}$ due to errors in instrument calibration and emissivity estimates (Wan 1999).

To account for the altitude dependence of the observed spectral radiance, the MODTRAN sensor altitude (i.e., the parameter H2) was set to a nominal altitude of 21 km. This matched the nominal altitude of MAS for the imagery used in this experiment. The surface temperature T_s was set equal to the air-surface boundary layer temperature T_{air} . This introduced some variability in the surface temperature and was done to see if CCR would be able to separate the surface and atmospheric emission effects. In addition, the CCR inverse models were also used to retrieve the surface temperature directly. In this experiment, the surface albedo was set to 0.0. Therefore, the surface was assumed to be a blackbody. Default values for the standard mid-latitude summer model were used for all atmospheric parameters not described by the radiosonde data. Simulated observed spectral radiances as well as transmission, upwelled radiance, and downwelled radiance were recorded. Gaussian sensor response functions defined by specified FWHM and band centers

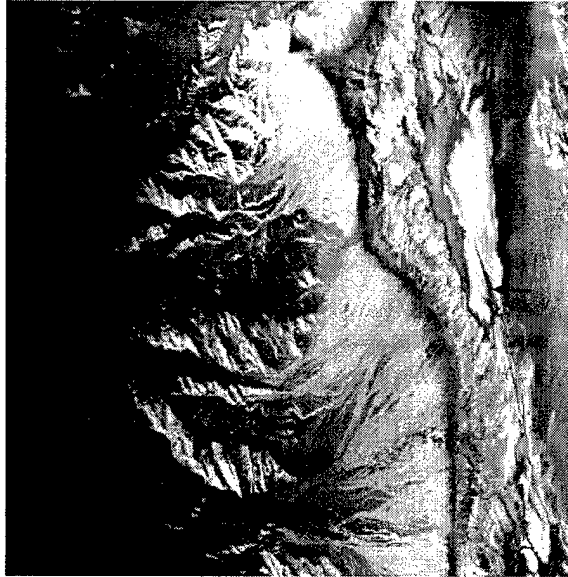


Figure 3.15: MAS infrared image of Death Valley (Band 44 equalized grayscale). The arrow indicates location of field temperature measurement.

were used to resample all spectral radiances and simulate MAS spectral observations. The observed radiances for each radiosonde set were then collected in an ensemble $\mathbf{X}$ with n observations for each of $p = 9$ spectral bands. The bands correspond to the longwave infrared (LWIR) MAS bands 42-50. The band configuration for the Death Valley collect are listed in Table 3.5. The observations were randomized prior to the analysis.

Experiment #3

The goal of this experiment was to demonstrate the ability of CCA to identify regions of the MWIR spectrum that are most useful for atmospheric sounding of temperature and water vapor. Another goal was to show that the CCR inverse model is “physical”. That is, the inverse mapping is based on physical properties of radiative transfer rather than on ensemble-dependent features that fortuitously lead to least-squares optimization.

The FSL and NAST-I data sets were used to generate simulated spaceborne MWIR observations. To build the $\mathbf{X}$ ensemble, radiosonde profiles and simulated observed radiances

Band	Center	Right 50%	Left 50%
42	8.53400	8.72800	8.31400
43	9.67500	9.98300	9.43900
44	10.5040	10.7120	10.2570
45	10.9930	11.2120	10.7280
46	11.9930	12.1560	11.7270
47	12.8670	13.0950	12.6960
48	13.3030	13.4950	13.0270
49	13.8330	14.0850	13.5240
50	14.2930	14.4720	14.0300

Table 3.5: MAS Thermal Bands for Death Valley Collect. The “Right” and “Left” columns denote the wavelengths to the right and left of the center wavelength where the response is 50% of the center response.

were processed at a nominal altitude of 100 km, which is the maximum for MODTRAN. Atmospheric optical properties were also generated and recorded. For each radiosonde observation, the surface temperature was varied about the air-surface boundary layer temperature by ± 6 °K at 2° increments. As in Experiment #2, the surface albedo was set to 0.0 and the parameters not specified by the radiosonde were set to mid-latitude summer profile values.

In this study, $n = 120$ observations were randomly chosen from both data sets to minimize computational time. The number of bands varied depending on the resolution at which MODTRAN was run. For this MWIR case study, the bandpass from 1950 to 3350 cm^{-1} (2.98 to 5.13 μm) was used. The highest resolution available with MODTRAN is 1 cm^{-1} . The results presented here are based on analysis done using high, medium, and low resolution cases defined by resolutions 1 cm^{-1} , 7 cm^{-1} , and 15 cm^{-1} , respectively. This corresponded to having 1401, 201, and 94 spectral bands in the $\mathbf{X}$ ensemble, depending on the test case. The number of correlations derived from CCA should provide insight into the number of *independent* channels of information. Also, the canonical weights should indicate the most influential regions of the spectrum that lead to the largest correlations with the atmospheric profiles.

Experiment #4

The goal of this experiment was to demonstrate the ability of CCA to define an inverse model for the prediction of atmospheric parameters and surface temperature under the influence of varying surface emissivities. The hypothesis is that CCA can be forced to find features that are independent of the surface emission and reflection by introducing variability in these parameters. The inverse model was also coupled with the TES algorithm to retrieve spectral emissivities and surface temperatures. Finally, the algorithm was applied to MASTER thermal imagery.

Ensembles were built with the FSL, NAST-I, and SSEC radiosonde profiles. In addition to the profiles, the surface temperature and emissivity were varied. The surface temperature was varied to $\pm 6^\circ\text{K}$ of the air-surface boundary layer temperature specified in the profiles. Ensembles were built using one of two surface cases: 1) using blackbody targets, and 2) using spectrally varying emissivities. The first case provides a baseline which is suitable for retrievals over near-blackbody targets such as water and certain types of vegetation. The second case is more general and applies to remote sensing over water and land. Three spectral emissivity classes were used: ocean, desert, and farmland. Thus, for each vertical profile, there were three different emissivity targets used as inputs in the MODTRAN model. These emissivities were chosen because they represent large generic classes and are conveniently referenced in the MODTRAN model. The emissivity spectra fluctuated from about 0.99 to 0.75.

In all cases, 60 observations were randomly selected from the radiosonde databases. For the blackbody cases, 7 temperature levels were used for each atmospheric observation, resulting in $n = 420$ observations. For the varying emissivity cases, 3 temperature and 3 emissivity levels were used resulting in $n = 540$ observations.

The MODTRAN runs were executed at the highest resolution (i.e., 1 cm^{-1}) and then resampled using MASTER sensor response functions. The MASTER sensor has 10 relatively wide bands (about $0.5\text{ }\mu\text{m}$ resolution) in the LWIR so it is considered multispectral. Three

Band	Lake Mead	RR/WR Valley
41	7.7574	7.7652
42	8.1599	8.1739
43	8.6120	8.6267
44	9.0487	9.0944
45	9.6855	9.7025
46	10.0966	10.1193
47	10.6186	10.6299
48	11.3079	11.3147
49	12.0984	12.1139
50	12.8712	12.8792

Table 3.6: MASTER band centers for Lake Mead and Railroad/White River Valley collects.

MASTER images were analyzed: 1) Flight 99-001-01 Track F over Lake Mead, NV on 02 December 1998; 2) Flight 99-006-14 Track F over White River Valley, NV on 29 September 1999; and 3) Flight 99-006-14 Track B over Railroad Valley, NV on 28 September 1999. The spectral band centers for these collects are listed in Table 3.6. Figure 3.16 shows the shape of the spectral response curves used to resample the MODTRAN runs for one of the collects (obtained from the ASTER web site: <http://asterweb.jpl.nasa.gov>). These collects were supported with ground truth measurements of surface temperature with self-calibrating radiometers and thermistors. In addition, emissivity measurements were made in the lab and the field in support of the Railroad Valley collect. The Railroad Valley and White River Valley images were collected at a nominal aircraft altitude of 10 km. The Lake Mead image was collected at about 6 km (Hook, Myers, Thome, Fitzgerald, and Kahle 2000; Palluconi 2000).

The images were obtained from the Earth Resources Observation Systems (EROS) Data Center (EDC) for the U.S. Geological Survey's (USGS), which also serves as the NASA Distributed Active Archive Center (DAAC) of MASTER data. The data are supplied georeferenced to latitude and longitude coordinates. These coordinates were used to select the pixels corresponding to ground truth measurements. The error in this procedure for the Railroad Valley and White River Valley images was less than $1.7 \mu\text{rad}$, which translates

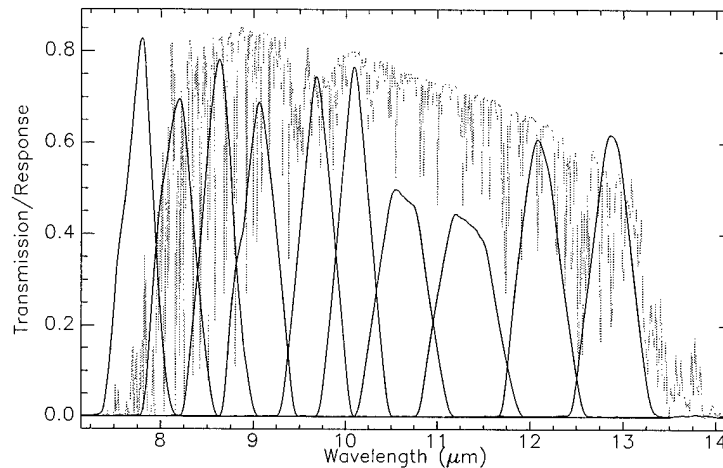


Figure 3.16: MASTER band spectral response for Railroad/White River Valley collect. An atmospheric transmission curve is superimposed for reference. Response functions were scaled for visualization.

to approximately 11 meters. For the Lake Mead image, the accuracy of the pixel location was about 5.6 meters. Figure 3.17 shows excerpts from these images using band 46 (at 10 μm). The location of the ground measurements are shown as asterisks. These pixels were processed through the inverse models built with the MODTRAN runs.

To gauge the performance of the algorithm with hyperspectral sensors, the MODTRAN output was resampled using spectral response functions for the Spatially Enhanced Broadband Array Spectrograph System (SEBASS) described in Section 3.3.4. MODTRAN spectra were resampled with the SEBASS band configuration for the 1997 Atmospheric Radiation Measurement (ARM) site collects over Lamont, Oklahoma.

3.3.6 Comparative Studies

Two comparative studies with other existing methods were made. One study compared the CCA inverse model results to those obtained with other multivariate regression models (see Appendix D for a description of these methods). Another study compared the CCR

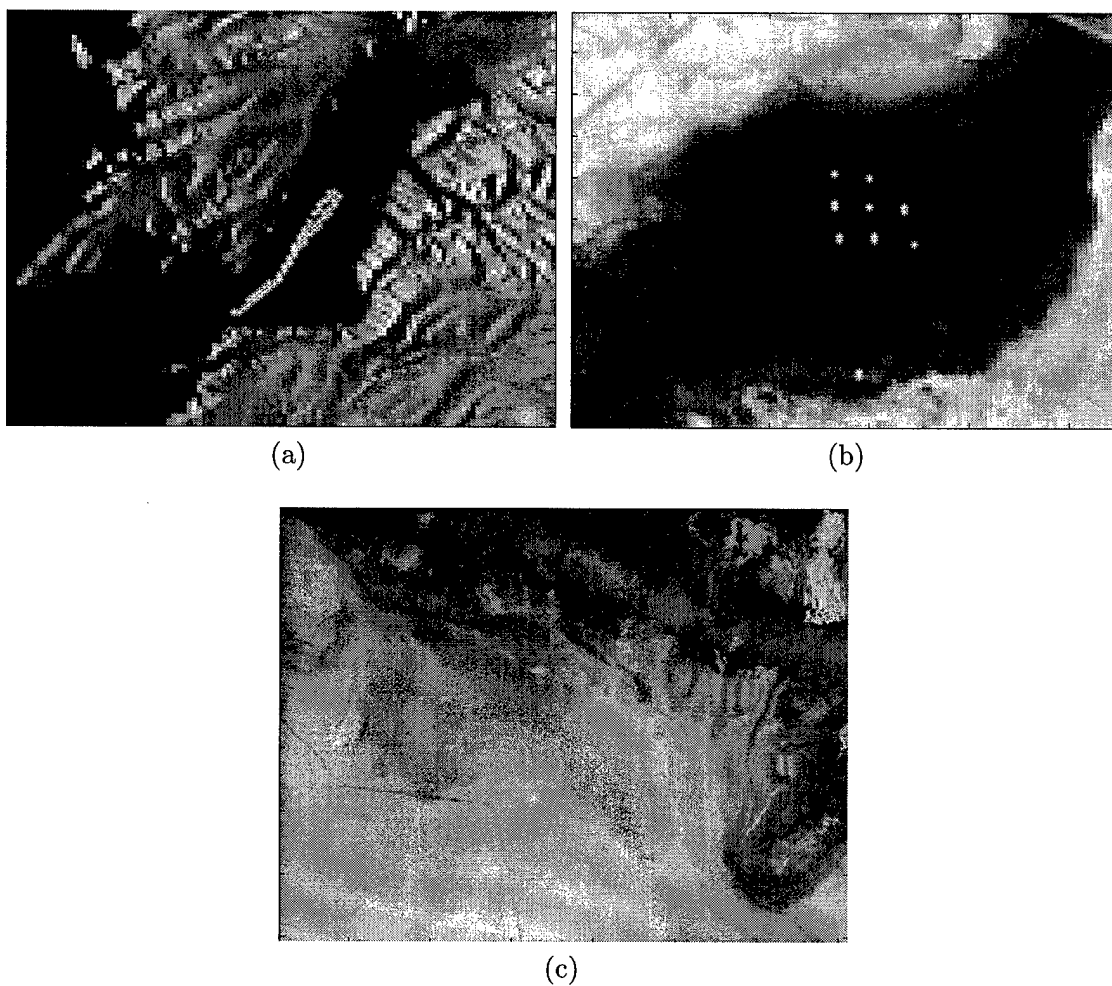


Figure 3.17: MASTER IR images with location of ground measurements: a) Gypsum Bay in Lake Mead (radiometer); b) Cold Springs reservoir in White River Valley (buoy-mounted thermistor); and c) Railroad Valley playa (FTIR).

inverse model estimates to results from the In-Scene Atmospheric Compensation (ISAC) method. The ISAC algorithm was implemented with the “maximum-hit” and Kolmogorov-Smirnov regression methods described in Section 2.1.6. The Kolmogorov-Smirnov ISAC algorithm used in this research was a modified version of the algorithm distributed by the Spectral Information Technical Application Center (SITAC), Central MASINT Office (CMO). Finally, a new “normalized” regression (NR) implementation of ISAC was also developed and tested. A detailed description of the different ISAC implementations is given in Appendix E. Although these studies were not comprehensive, they provided some indication of the performance of the CCA approach relative to other methods. ISAC is a good baseline for comparison because it is relatively popular in the community.

3.3.7 Validation of Linear Model

Finally, the appropriateness of using a linear inverse model was investigated. This was done through the analysis of canonical variable and residual vs. fitted value scatter plots. The canonical variable plots give insight into the “shape” of the data in the canonical space. If the canonical correlations are high, then the scatter plots should follow a linear pattern and the linear model is appropriate. On the other hand, low correlations in the canonical space result from: (1) all of the correlation being explained by the first few canonical variables, in which case the linear model is appropriate and the plot of the of the low-correlation canonical variables has no pattern, and (2) outliers in the data due to errors or nonlinear relationships, in which case the canonical variable plots should exhibit some pattern. The residual vs. fitted value plots give an indication to how the errors change relative to the values being estimated. A pattern in the error plot gives an indication of the appropriateness of the model. For example, the residual vs. fits plot for the TES algorithm shown in Figure 3.10 indicate that emissivity spectra with low minimum emissivity values and high variability are not modelled well by the $\text{MMD-}\epsilon_{\min}$ linear model.

Chapter 4

Results

*No amount of experimentation can ever prove me right;
a single experiment can prove me wrong.*

Albert Einstein

The detailed description of the experimental design for the results presented in this chapter are found in Section 3.3. However, schematics showing the design of each experiment are presented throughout the material as reference.

4.1 Experiment #1

Figure 4.1 shows a schematic of the experimental design. The goal of this experiment was to demonstrate that the canonical correlations relating the observed radiance and atmospheric parameters were large enough to build an accurate inverse model. In addition, a scaling scheme for surface temperature biases was explored.

The analysis relating the observed radiance and the atmospheric transmission and upwelled radiance resulted in 6 significant correlations with squared values of 0.9999, 0.9969, 0.9873, 0.97164, 0.8996, and 0.8118. A complete sum-of-squares of error (SSE) matrix would be too large to show here and difficult to interpret. However, the RMS errors for trans-

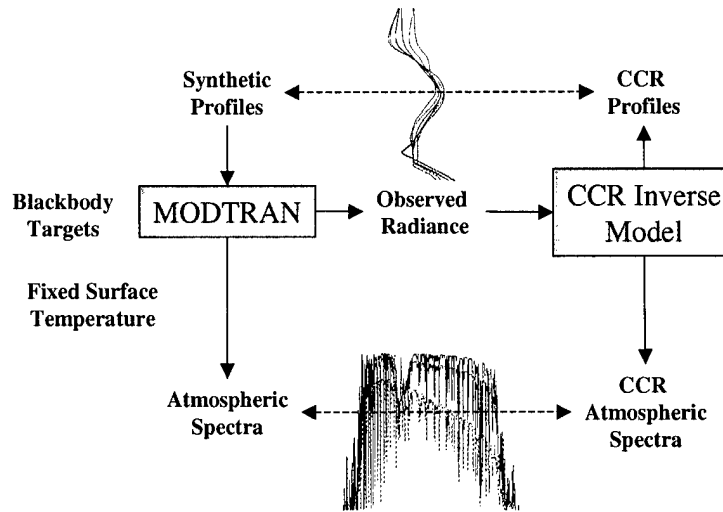


Figure 4.1: Schematic of design of Experiment # 1

mission and upwelled radiance over all wavelengths were 0.011 and 15.49 μf , respectively. Clearly, there is a strong linear correlation between the observed radiance and the atmospheric spectra. These results should not be surprising since eq. (2.5) define the observed radiance as a linear combination of transmission and upwelled radiance effects.

Figure 4.2 shows an example of the results for the atmospheric transmission and upwelled radiance predictions. The black curves are the “true” values obtained from the MODTRAN model. The red curves are the CCR predictions. The data shown are representative of observation #2, which is a Tropical model case. Both the temperature and relative humidities are from the tropical model. The O_3 content was the standard default concentration for this observation.

One of the aspects that makes CCR very appealing is the analysis of the linear combinations used in the transformation. These “canonical modes” are the *orthogonal* basis of the space the data are transformed to. These “modes” or “weights” are often interpretable and provide insight into the nature of the problem. Figure 4.3 shows the first two modes for the transmission and upwelled radiance predictions. The black curves are the modes

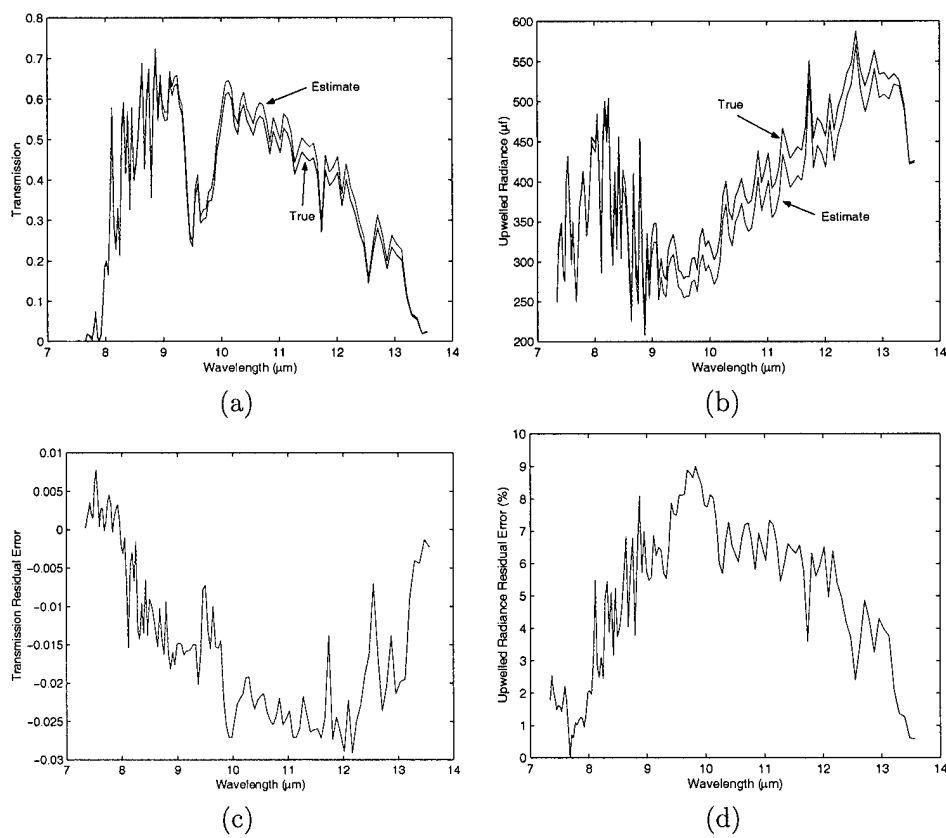


Figure 4.2: Comparison of “true” and CCR-predicted spectra: (a) transmission, (b) upwelled radiance, (c) transmission residual, and (d) upwelled residual (% error).

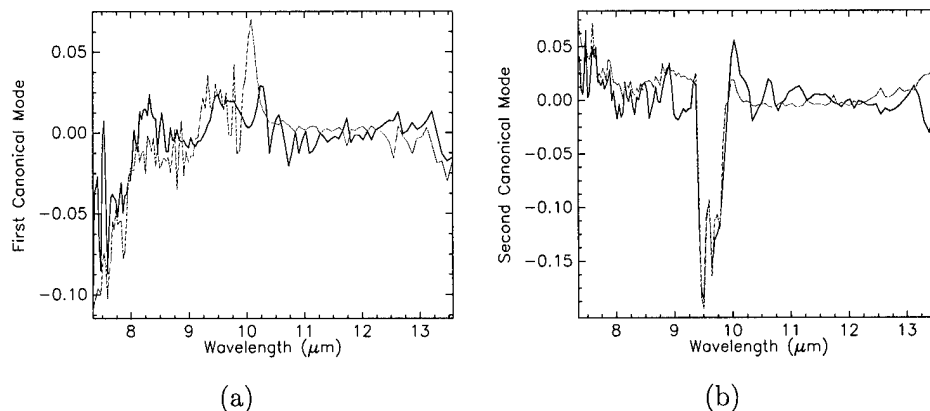


Figure 4.3: First pair of canonical modes obtained from the radiance and transmission spectra: (a) First canonical modes, and (b) Second canonical modes.

used to transform the observed radiance spectra. The green curves are the modes used to transform the transmission spectra. Note that the modes for both sets of data are very similar, which indicates that there is a tight relationship between the observable features in each spectra. The first mode appears to be picking up on the wing of the water absorption band centered at $6.3 \mu\text{m}$. The second mode is clearly based on the shape of the major ozone absorption band at $9.6 \mu\text{m}$. The modes for the upwelled radiance were consistent with these results. These interpretations of the modes indicate that the CCA inverse model is physical.

CCR was also used to relate the observed spectra to the vertical temperature profiles used as inputs to MODTRAN. Figure 4.4 shows a comparison of the true input and the predicted temperature profile for observation # 79. This is a mid-latitude winter temperature profile, which was run with a mid-latitude summer relative humidity profile and the default O_3 concentration. The curve in red is the predicted profile. All of the residuals were within 1°K of the true profile. There were 4 significant canonical correlations: 0.9995, 0.9988, 0.9466, 0.8531. The RMS error over all altitude levels was 0.93°K .

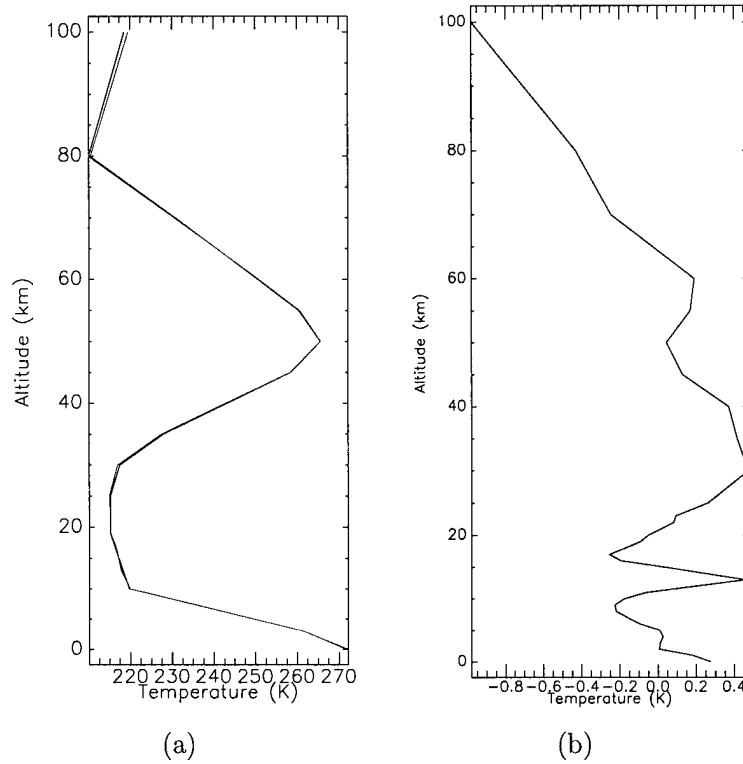


Figure 4.4: Comparison of “true” and CCR-predicted temperature profiles (observation #79): (a) temperature profile, and (b) Residual.

The second part of the experiment addresses a practical issue that may potentially introduce large errors in the estimated spectra. Inspection of eq. (2.5) will show that the surface-leaving radiance can significantly dominate the observed radiance. This is particularly the case when the surface temperature is larger than the apparent temperature of the atmosphere. Although this particular scenario would normally be welcomed in infrared remote sensing of the surface, it is not ideal when the goal is to remotely sense the *atmosphere*. Therefore, it is necessary to scale the data so as to *minimize* the effect of the surface-leaving radiance on the prediction of the atmospheric parameters. The emphasis is placed on minimization of the effect because it is physically impossible to get around the effects of a poor signal-to-noise ratio (SNR). In the atmospheric parameter retrieval problem, the signal of

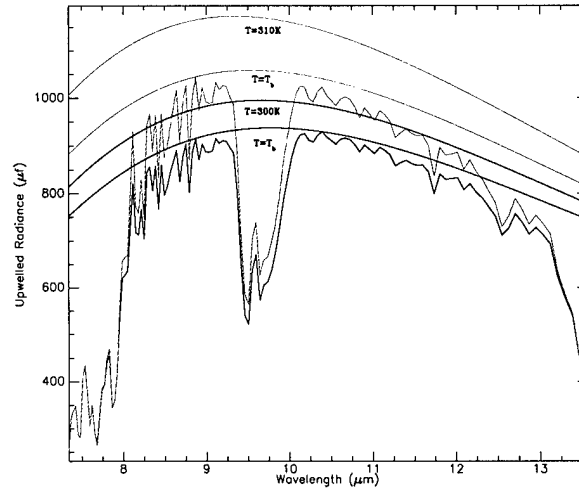


Figure 4.5: Observed radiances for two different surface temperatures. The black curves correspond to a surface temperature of 300 °K. The green curves are for a surface temperature of 310 °K.

interest is the upwelled radiance while the “noise” is the surface-leaving radiance. The best that we can do is to *force* the canonical correlation regression algorithm to use features in the observed radiation that are *independent* of the surface-leaving radiance. A heuristic approach was developed that seemed to work well with the data discussed in section 3.3.5. This procedure is an intermediate step necessary to characterize the atmosphere. Later on, the atmosphere will be the “noise” and the estimated atmospheric parameters will be used to calculate the surface radiance.

The approach is based on the careful inspection of the observed radiances for two cases. Figure 4.5 shows the curves for two cases. The first case corresponds to observation #2 of the regression analysis. This is obtained from the standard MODTRAN tropical model with a surface temperature of 300 °K. The second case is an observation *outside* of the regression data. The surface temperature used in the new observation is 310 °K. This 10°K difference should introduce a large enough bias to test the algorithm against. The structured curves are the actual MODTRAN observed radiances. The smooth curves are

Planck functions derived with either the maximum brightness temperatures of the observed radiances or with the actual surface temperature. For both cases, the Planck curves obtained with the brightness temperatures are lower than the Planck curves obtained with the true surface temperatures. This is expected, since the maximum brightness temperature is an underestimation of the true temperature because of attenuation by the atmosphere. Note however, that the Planck curves derived from the maximum brightness temperatures follow the shape of the observed radiance much more closely than the Planck curves obtained from the surface temperatures. The difference between the Planck curves derived from the maximum brightness temperature is proportional to the difference between the observed radiances except for at the edges of the band and at the center of the ozone feature. This is because these are areas where the atmosphere is highly absorbent (low transmission). In those regions, the contribution from the surface component to the observed radiance is low. Thus, for the case of equal atmospheres, the observed radiances are the same in those regions. A suitable scaling scheme would be one that would shift the spectra *only* in spectral regions where the atmosphere is transmissive.

Unfortunately, there is no *a priori* knowledge of the actual atmospheric transmission. In fact, this is the quantity that we wish to predict with the analysis! However, note that the observed radiance is *proportional* to the transmission of the atmosphere. This will generally be true for cases where the surface-leaving radiance component dominates over the atmospheric upwelled radiance. Luckily, this is the scenario that we wish to “compensate” for. Thus, a suitable scaling of the observed radiance may be obtained by multiplying the difference in the Planck curves obtained with the brightness temperatures by a fraction that is proportional to the observed radiance. This defines the bias that the observed radiance is shifted by. Mathematically, the fraction is obtained from

$$\gamma(\lambda) = \frac{L(\lambda) - \min [L(\lambda)]}{\max \{L(\lambda) - \min [L(\lambda)]\}} \quad (4.1)$$

The numerator is the radiance subtracted by its minimum. This ensures that the minimum value of the correction factor is zero. The resulting curve is then divided by its maximum

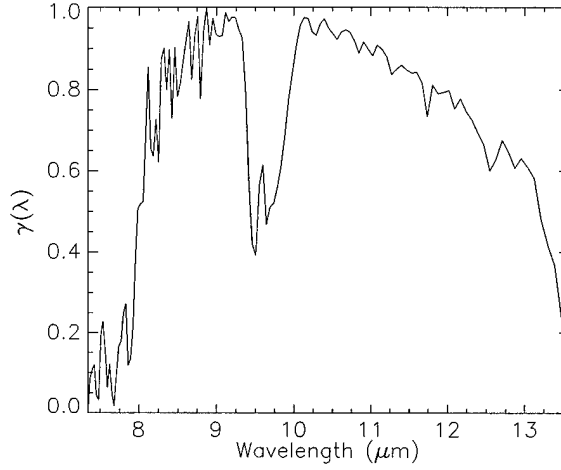


Figure 4.6: Scaling factor applied to the difference in Planck curves.

value so that the peak value is 1.0. Thus $0 \leq \gamma \leq 1.0$. This fraction is a scaling factor that is proportional to the atmospheric transmission. This curve is shown in Figure 4.6. The next step is to multiply the scaling factor by the difference in the Planck curves obtained from the maximum brightness temperatures. The bias then becomes

$$b(\lambda) = \gamma [L_{BB}(\lambda, T'_{b1}) - L_{BB}(\lambda, T'_{b2})] \quad (4.2)$$

where T'_{b1} and T'_{b2} are the maximum brightness temperatures for the two observations.

Unfortunately, for any particular observation, we do not know what T'_{b2} the radiance should be compared to. We do know, however, that all of the observations to be used in the regression analysis were generated at a single surface temperature $T_s = 300^\circ\text{K}$. Thus, we would scale *all* of the data to the Planck curve derived with this temperature. The bias to be used is then

$$b_i(\lambda) = \gamma [L_{BB}(\lambda, T'_{bi}) - L_{BB}(\lambda, T'_s)] \quad (4.3)$$

where $b_i(\lambda)$ is the bias subtracted from the i^{th} observation. Figure 4.7 shows the effect of the scaling factor on the observation biased by the high surface temperature. The scaling

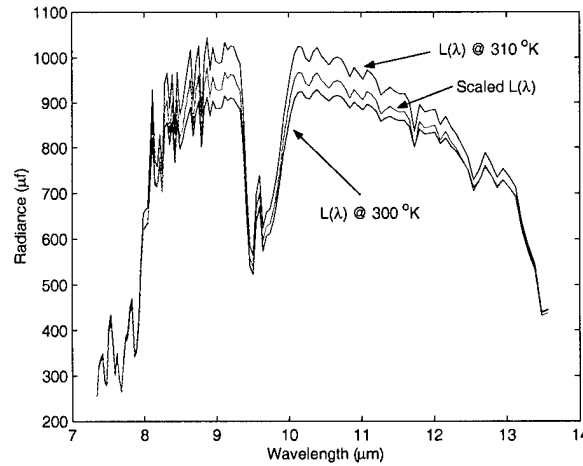


Figure 4.7: Comparison of scaled and unscaled observed radiances for observation #2 of the regression analysis.

brings the observed radiance closer to the curve obtained with 300 °K but it is not perfect. This will translate to an increased error in the atmospheric spectra estimates. Nevertheless, the results should be an improvement over those obtained with unscaled data.

Figure 4.8 shows the results that would be obtained if the data were not scaled. The errors are very large and the retrieved transmission and upwelled radiance contain values that are not physically possible.

Figure 4.9 shows the results obtained with the scaled data. CCR was done with all of the observations in the regression data scaled with the bias correction. The scaling had no significant effects on the regression analysis since all of the observations had a common surface temperature. However, the scaling dramatically improved the retrievals obtained with the biased observation outside of the regression data. These results are much better than those obtained with the unscaled analysis. However, the residuals are larger than those obtained with just the regression data. There is also a distinct pattern to the residuals, suggesting that there may be a better way to minimize the effect of changes in the surface radiation.

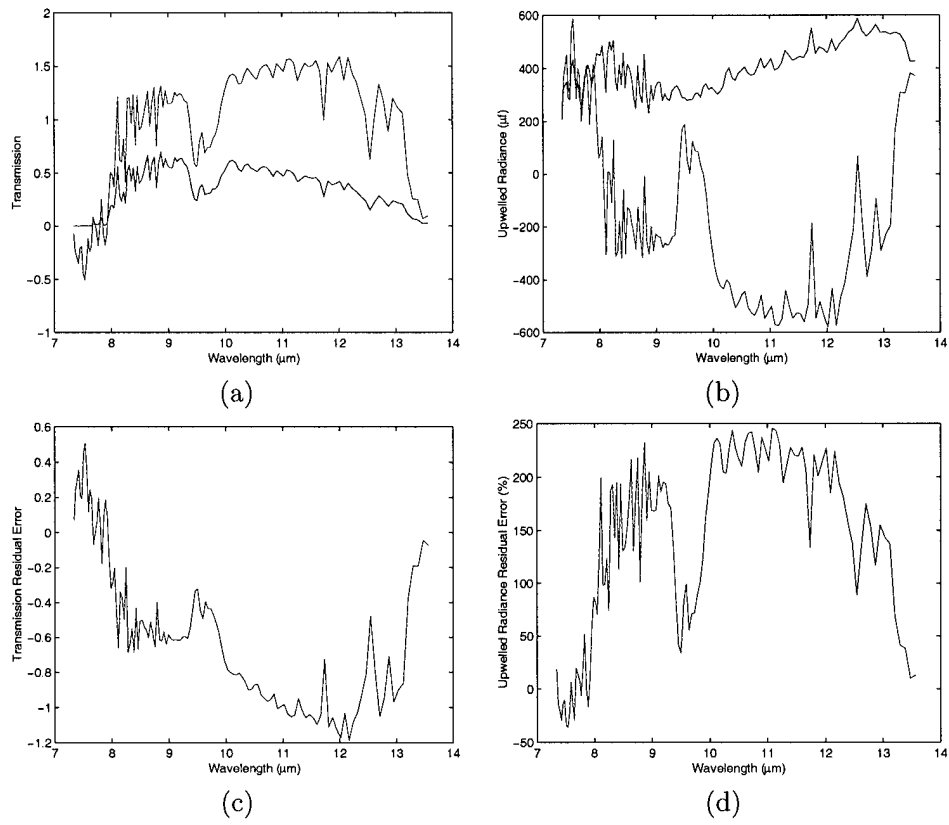


Figure 4.8: Comparison of “true” and CCR-predicted spectra (before scaling the data): (a) transmission, (b) upwelled radiance, (c) transmission residual, and (d) upwelled residual (% error). Blue curves are the true values; red curves are the estimated spectra

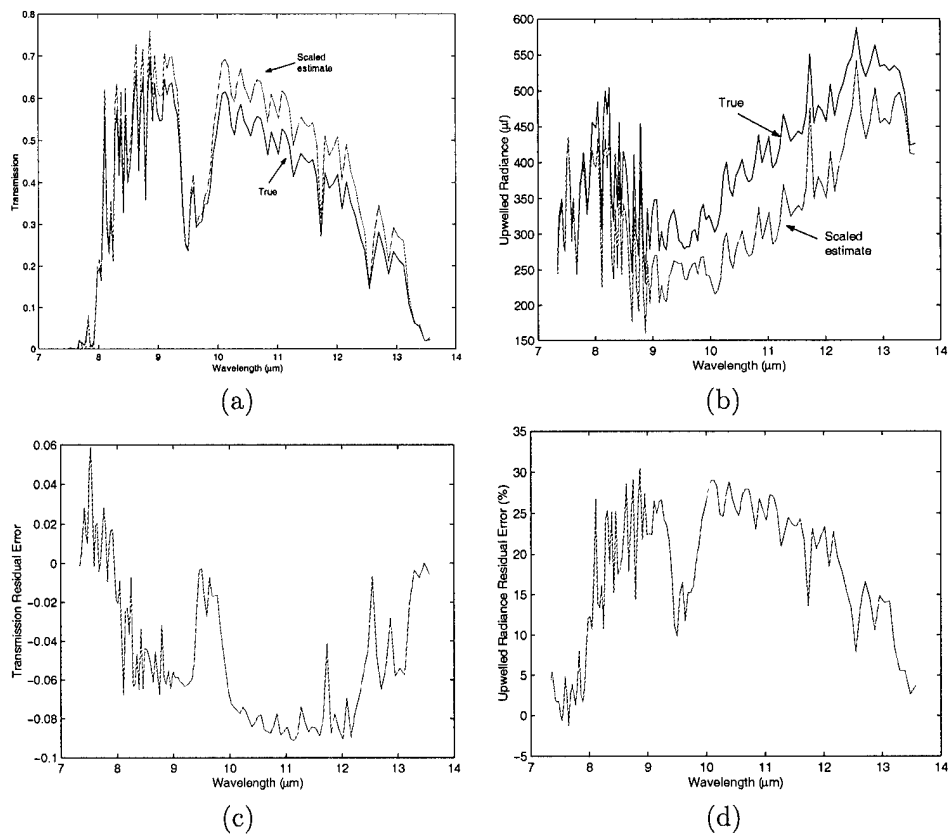


Figure 4.9: Comparison of “true” and CCR-predicted spectra (after scaling the data): (a) transmission, (b) upwelled radiance, (c) transmission residual, and (d) upwelled residual (% error).

This *heuristic* approach minimizes the effect of differences in the surface-leaving radiance due to temperature changes because all of the observed radiance spectra are scaled to one common (or *reference*) surface temperature. The reference temperature does not necessarily have to be the surface temperature used in the regression data. However, the reference temperature *should* be in the domain of surface temperatures used in the regression analysis. This adds flexibility in the experimental design by allowing observations that were obtained with different surface temperatures to be used in the regression analysis. Finally, the scaling takes into consideration the fact that we do not wish to apply any correction to spectral regions where the surface contribution is minimal.

The scaling approach developed for this experiment was *not* used throughout the rest of this research. This is because subsequent experiments included the the surface temperature as a parameter to be estimated by the CCR inverse model. Nevertheless, the scaling results are presented here since they may a suitable implementation of CCR when only the atmospheric parameters are of interest.

In summary, the results from this experiment demonstrate that the canonical correlations between atmospheric physical and optical parameters and the observed spectra are large enough to build an accurate inverse model. In addition, analysis of the canonical modes showed that the CCA inverse model is physical. A scaling scheme for mitigating biases due to changes in surface temperatures was developed.

4.2 Experiment #2

CCR inverse models were built for the prediction of atmospheric profiles and spectra. In both cases, the parameters were retrieved *simultaneously*. That is, the ensemble $\mathbf{Y}$ contained all of the parameters of interest. For the case of atmospheric profile retrievals defined at q altitude levels, $\mathbf{Y}$ was a $n \times 2q$ *partitioned* matrix where the first q columns were the temperature profiles and the second q columns were the water vapor profiles. Similarly, an ensemble $\mathbf{Y}$ of atmospheric transmission, upwelled radiance, and downwelled radiance was a

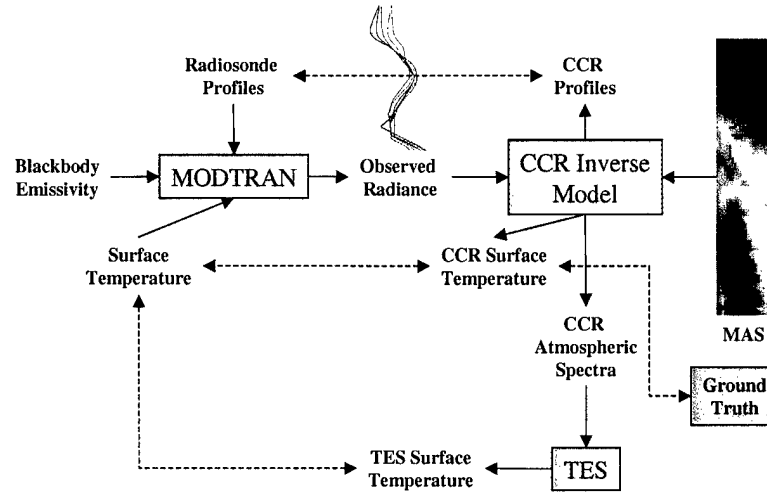


Figure 4.10: Schematic of design of Experiment # 2

$n \times 3p$ matrix. This was done to constrain the algorithm and prevent CCR from attributing the same observed radiance effect to different parameter estimates. The experimental design is shown in Figure 4.10. In this experiment, the surface temperature as allowed to vary (as defined by the lowest layer in the radiosonde profiles) and was estimated by the CCR inverse model. In addition, TES was coupled with CCR to refine the surface temperature estimates.

The results are summarized in Table 4.1. Since no field emissivities were measured, the TES emissivity estimates cannot be verified. However, the estimated surface temperatures agree to within 0.4 °C of the field temperature measurements. The accuracy of the retrieved TES temperatures is improved by as much as 1 °C compared to the direct CCR retrievals. This suggests that the TES algorithm is adequately compensating for reflected downwelled radiance and emissivity. In addition, the direct retrieval is based on all of the bands while the TES retrieval uses only the bands in regions of high transmission. These bands also had lower sensor noise than those at the edge of the LWIR bandpass (Wan 1999). The estimated uncertainty in these retrievals is within 1 °C for the FSL and NAST-I data and about 4 °C for the SSEC data. These results are consistent with the analysis on the databases used to

Parameter	SSEC	FSL	NAST-I
MAS			
RMS Error T_s ($^{\circ}\text{C}$)	1.15	0.21	0.23
RMS Error Profile ($^{\circ}\text{C}$)	2.13	2.20	4.65
RMS Error CWV (mm)	4.66	1.92	4.65
CCR Retrieval MAS pixel T_s ($^{\circ}\text{C}$)	19.82	17.50	19.32
TES Retrieval MAS pixel ^a T_s ($^{\circ}\text{C}$)	18.9	18.7	18.3
MODIS			
RMS Error T_s ($^{\circ}\text{C}$)	1.31	0.26	0.26
RMS Error Profile ($^{\circ}\text{C}$)	2.00 ^b	2.10	1.53
RMS Error CWV (mm)	4.96	2.10	4.65

^aOnly MAS bands 42,44-48 used.

^bFor pressure levels greater than 100 mbar.

Table 4.1: Summary results for experiment # 2

build the correlation coefficients. Recall that the SSEC data had a lot of variability in the profiles as a result of the sparse geographic coverage over the entire globe. This results in estimates that are less precise than for the FSL and NAST-I retrievals.

Finally, MODIS data simulated with MODTRAN were processed. These data were generated using the maximum altitude in MODTRAN (100 km). The residual errors in the atmospheric parameters are shown to be of the same order as those obtained with the MAS data. This shows the extendibility of the algorithm to spaceborne remote sensing platforms.

The results obtained in this experiment are *not* optimized for the observations in the training data. That is, it is possible to obtain more accurate results for the specific ensembles used to build the regression coefficients by keeping more canonical dimensions in the inverse model. However, this may lead to “overfitting” of the data. Therefore, the model is more robust and applicable to observations *outside* of the training set because only 3-4 significant correlations were retained. Figure 4.11 shows how the cross-validation results are nearly identical to those obtained with the ensembles used to build the model. To demonstrate the robustness of the model, 5 observations from the NAST-I data were used to predict all of the 3,310 NAST-I temperature profiles and corresponding air-surface temperatures. The RMS errors in the estimates were within 3 $^{\circ}\text{K}$ and 1 $^{\circ}\text{K}$, respectively.

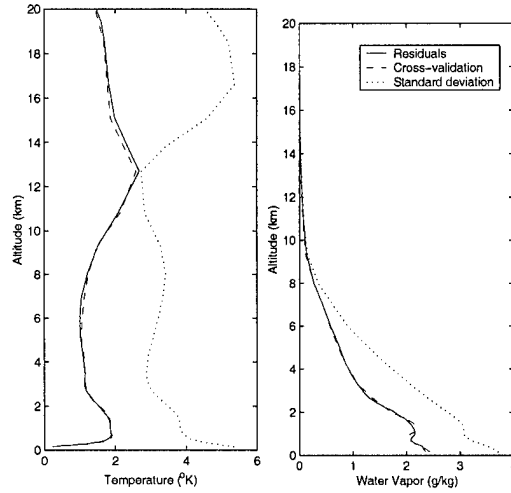


Figure 4.11: Cross-validation results for MAS profile retrievals for the NAST-I data. The standard deviation of the ensembles are shown to show the amount of variability accounted by the model.

4.3 Experiment #3

4.3.1 Number of Bands Determination

CCR provides a mechanism for finding the minimum number of bands required for accurate estimation of atmospheric parameters from the observed radiance. As described in Section 3.2.1, the algorithm truncates the singular value spectrum so that a lower dimensional space is used to build the canonical correlations. This ensures that the model does not overfit the data and the solutions are stable. The retained dimensionality is also an indication of the number of independent bands (i.e., the inherent number of independent variables) needed to achieve a certain level of accuracy.

The experimental design is shown in Figure 4.12. CCR was performed on the FSL and NAST-I data sets for the high, medium, and low resolutions. The RMS errors in the estimated temperature and water vapor profiles obtained with the FSL and NAST-I data are shown in Figure 4.13 and Figure 4.14, respectively. The lowest layer in the plotted profiles is actually the surface temperature, which was estimated with CCR as well. In

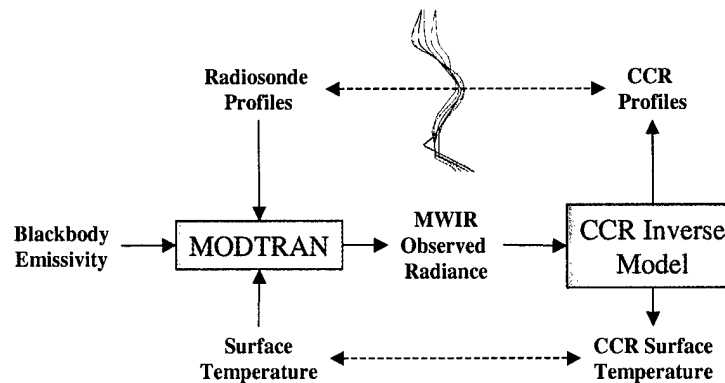


Figure 4.12: Schematic of design of Experiment # 3

general, CCR found strong correlations with the surface temperature and predicted it with high accuracy. The errors in the temperature profiles are larger in the stratosphere. This is because the contribution from stratospheric temperatures to the observed radiance is minimal due to lower atmospheric density and constituent population. For the water vapor profiles, the error is largest close to the Earth's surface where the water vapor content and variability is largest. The standard deviation of the profiles is plotted as reference to show what the residual error would be if the algorithm simply predicted the mean. This provides a baseline to help determine how much of the variability the algorithm is truly predicting. The error in the temperature retrievals approaches the standard deviation curve just below the tropopause (at about $z = 12$ km or $p = 200$ mb). This implies that there is little information in the MWIR spectrum about this atmospheric level. There could be several reasons for this: (1) there is no significant emission from this atmospheric level, (2) emission from this level does not reach the sensor because of path absorption, (3) the simulated spectral resolution is too low for the emission to be detected, (4) there are errors due to discretization and regularization, or (5) there are errors in the forward model.

The latent dimensionality of the problem is driven by the number of features available in the profile data that can be related to features in the observed radiation spectra. Clearly,

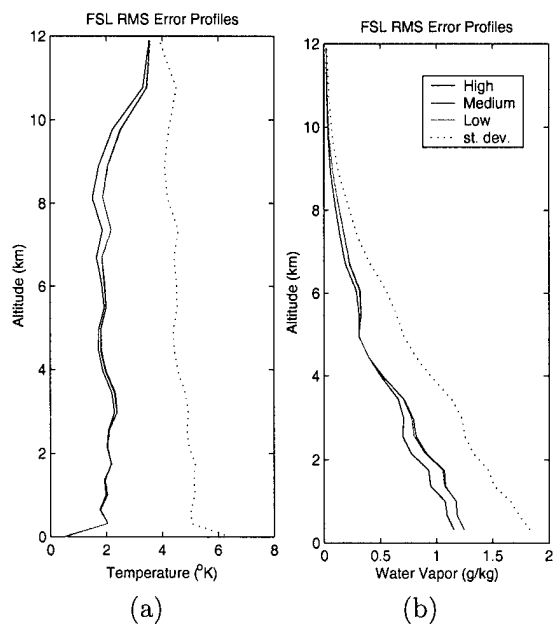


Figure 4.13: RMS error profiles for FSL dataset

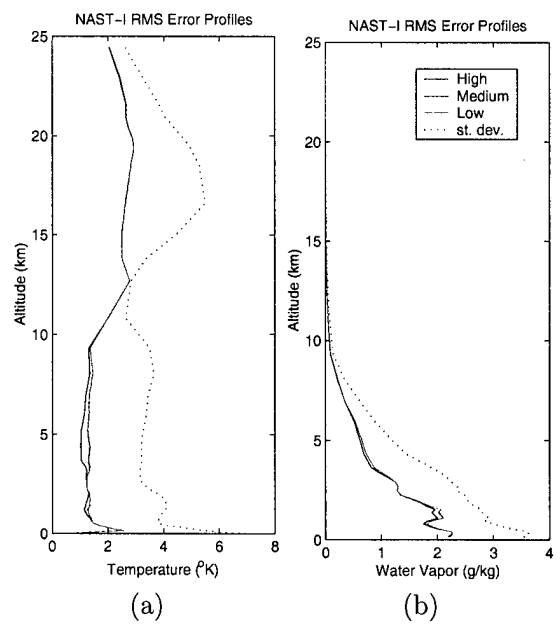


Figure 4.14: RMS error profiles for NAST-I dataset

atmospheric spectra are much more feature-rich than temperature and water vapor profiles. In addition, the amount of information about the profiles available in the observed radiance is limited by the discretization of the atmospheric profiles. By virtue of the true profiles being continuous functions, working with discrete measurements introduces a certain level of ambiguity and ill-conditioning. To avoid problems arising from this situation, the CCR model necessarily has to use a lower dimensional space. This regularization introduces a certain level of error.

The residual errors are a manifestation of variability not accounted by the CCR model. This “left-over” variability may be attributed to errors in the radiosonde measurements, errors in the radiative transfer code, and nonlinear relationships. Errors in the radiosonde measurements lead to inconsistencies and nonphysical behavior. Thus, the resulting predicted observed radiance from an erroneous profile will not be representative of the rest of the ensemble. These inconsistencies may reduce some of the “weaker” relationships that would otherwise be detected by the algorithm. How well the CCR model can represent the physics of radiative transfer also depends on the quality of the ensemble generated with the forward model. If the predicted observations associated with the input profiles do not accurately reflect what would truly be observed, then this has the effect of lowering the relationships in the data. Thus, errors in the band and radiative transfer models used in MODTRAN would contribute to the residual error. Finally, recall that the CCR model finds the maximum *linear* correlations in the data. If there are nonlinear relationships, these could go unaccounted for and would therefore contribute to the residual errors.

These errors are inherent to the problem and are present regardless of the method used for atmospheric compensation. In this experiment, the interest is in the *relative* performance of the algorithm as the number of bands are varied. In light of these inherent errors, CCR provides insight into the number of bands required so that more error is not introduced. The residual RMS errors did not change appreciably as the resolution of the observed radiation was varied. This indicates that the basis (i.e., modes; weights) spanning the canonical space

have broader spectral features than the modelled resolution. For the FSL data, the latent dimensionality was 4. For NAST-I, the dimensionality was 5. Presumably, the NAST-I set required an extra dimension because of the higher water vapor content and variability in the East Coast atmospheres. Thus, the analysis suggests that 5 bands may be enough to achieve close to the same level of accuracy as obtained with 94, 201, or even 1401 bands. The next section describes the optimal placement of these bands.

4.3.2 Band Selection

CCR provides regression coefficients between $\mathbf{Y}$ and $\mathbf{X}$ which are useful for prediction. The regression coefficients are formed from the canonical weights $\mathbf{A}$ and $\mathbf{B}$ and the canonical correlations. The columns of these matrices are the CCR weights associated with the significant correlations. In addition, the *loadings* are the correlations between the canonical and original variables. The loadings also represent the inverse transformation from the canonical to the original space and are given by $\mathbf{A}'\Sigma_{xx}$ and $\mathbf{B}'\Sigma_{yy}$ (see Section 3.1.3). These weights and loadings provide insight into how the spectral bands in $\mathbf{X}$ are used to infer information about $\mathbf{Y}$.

Figure 4.15 shows the significant CCR weights and loadings for the FSL medium resolution case. For this case, four dimensions were found to be sufficient. The four curves shown are associated with the four dimensions retained. All of the weights have been normalized so that they can be shown on the same plot. This is because the weights associated with the lower correlations have a smaller contribution to the predictions than those associated with the higher correlations. The plot of the weights for the observed radiance are overlaid with a normalized transmission curve. This curve is the mean transmission for all of the profiles in the FSL data and helps in the interpretation of where in the spectrum the canonical weights and loadings are finding the most information. In general, the loadings offer a better interpretation because irrelevant variance in the weights is removed. However, regions emphasized by both canonical weights and loadings should contain much information.

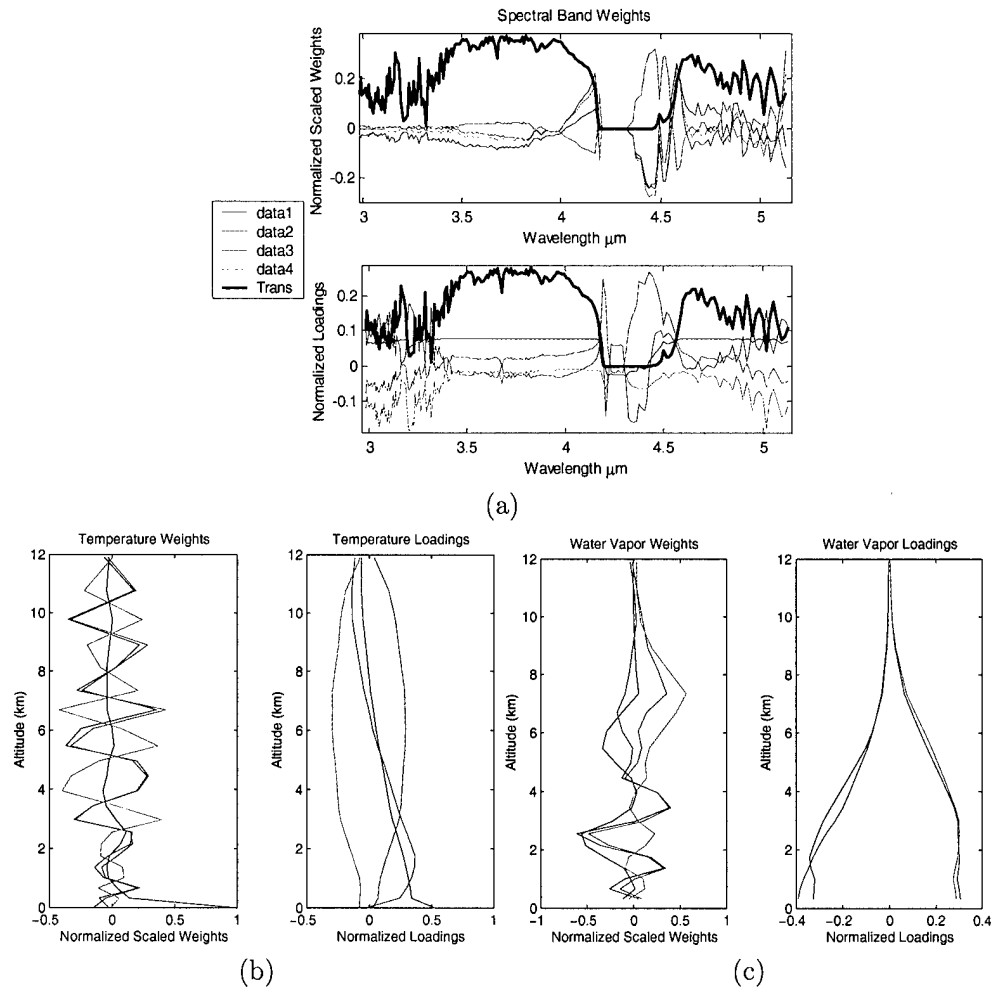


Figure 4.15: Results with FSL dataset (medium resolution): (a) Weights and loadings for the observed radiance, (b) weights and loadings for temperature profiles, and (c) weights and loadings for water vapor profiles

By relating the observed radiance to the profiles, the analysis is forced to find bands that would be useful for sounding. To the extent that the physics of radiative transfer are manifested statistically, the CCR method should provide results similar to the weighting functions. In the results shown in Figure 4.15, the sounding interpretation is evident in the way CCR relates the radiance to the profiles. For instance, consider the relationships built between the observed radiance and the temperature profiles. The blue curve denotes the first dimension in the analysis and is clearly a weighted average of the observed radiance in regions of relatively high transmission. Clearly, these regions offer the most information about the surface temperature and this is manifested in the weights and loadings for the temperature profiles where the emphasis is in the surface layer. The other 3 profile loadings are associated with information in upper layers of the atmosphere. The curves are akin to the weighting functions used for atmospheric sounding but they tend to be broader. This is because we are looking at the entire spectrum and not at just a single absorption feature. The general sounding approach is evident since information about the upper atmosphere is associated with regions of high absorption. The water vapor profiles tend to be wider and smoother suggesting that the CCR approach could not find significant correlations between higher order features in the profiles and the observed radiation spectra.

The canonical weights and loadings for the NAST-I data are shown in Figure 4.16. In general, the significant absorption bands in the observed radiance are similar to those obtained with the FSL data. The main difference between the two sets is that the variability in water vapor plays a larger role in changes observed in the observed radiance. This is because the average content of water vapor for the NAST-I data is higher than for the FSL data. Because of this extra factor, the number of canonical correlations is larger by one dimension. Also, the water vapor weights have more structure than in the FSL case and the loadings are better defined. There are two factors contributing to this. One has already been mentioned, and it is the fact that there's more water vapor in the NAST-I atmospheres and therefore a larger contribution to the observed radiance. The other is that the NAST-I

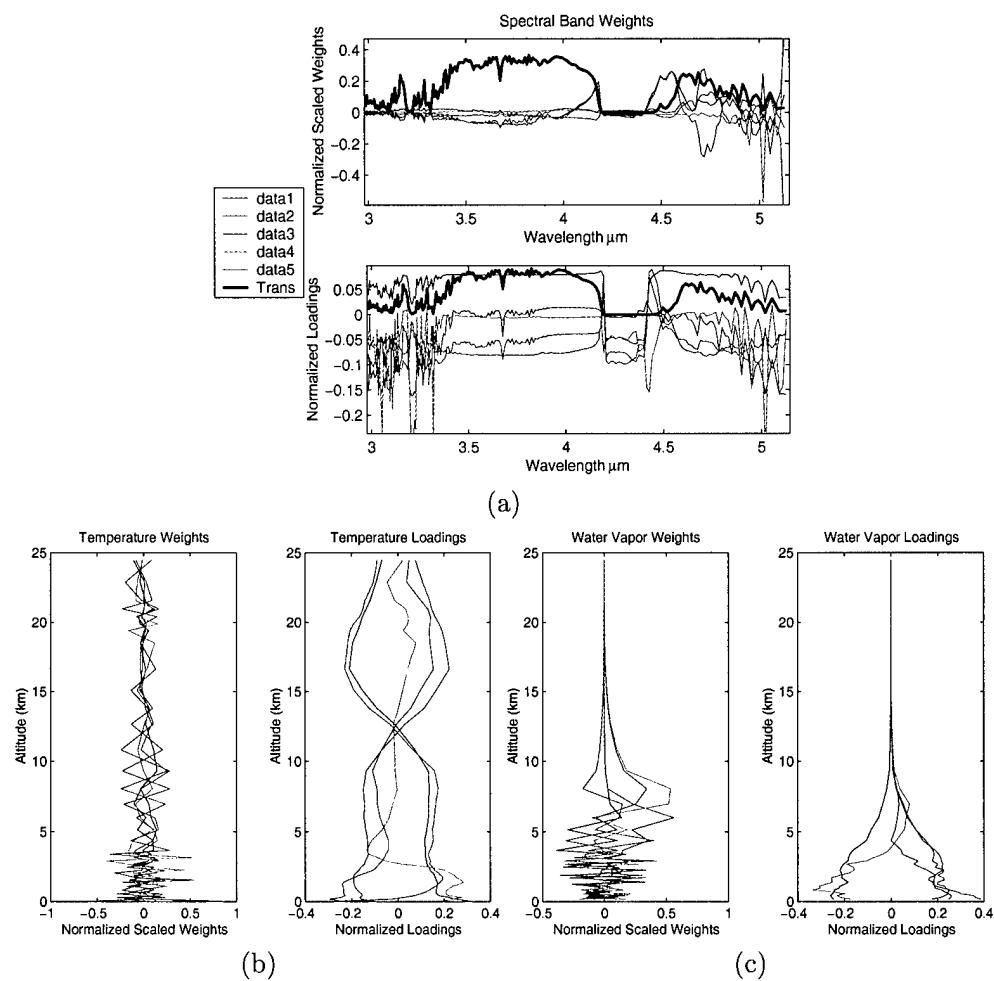


Figure 4.16: Results with NAST-I dataset (medium resolution): (a) Weights and loadings for the observed radiance, (b) weights and loadings for temperature profiles, and (c) weights and loadings for water vapor profiles

profiles are measured at more discrete levels than the FSL profiles. Thus, less errors are introduced by discretization.

Figure 4.17 emphasizes the *emissive* portion of the MWIR region. There are two reasons for doing this: (1) CCR finds more information about the profiles in this region, and (2) the effects of reflected direct solar radiation are minimized. Indeed, CCR may be focusing in this region to avoid variation introduced by solar radiation, which was certainly present in the data (i.e., the data contained day and night cases). The plots shown in Figure 4.17 are for the FSL and NAST-I medium and high resolution cases. The medium resolution case provides a smoother version of the high resolution case and is more applicable to a sensor with relatively wide FWHM. The high resolution case gives more insight into the fine structure of the atmospheric spectra and where wide bands and narrow bands would be more applicable. For example, to appropriately sample the large CO₂ band “wings” originating at 4.3 μm , two relatively wide bands are sufficient (namely, two centered at 4.44 and 4.52 μm). However, to sample some of the water vapor features, narrow bands are necessary (e.g., bands centered at 4.9 μm and 5.01 μm).

Based on the results in Section 4.3.1, five bands were selected. The location of these was based on where in the spectrum the canonical weights and loadings were placing an emphasis. The criterion for selection was that the band had to appear to be significant in the weights *and* the loadings for *both* data sets. The selected bands were 4.44, 4.52, 4.6, 4.9, and 5.01 μm . All of the bands except band 3 are used to characterize the atmosphere. Band 3 was chosen for imaging and surface temperature determination because it corresponds to a region of relatively high transmission and blackbody emission, thus taking into consideration practical issues dealing with signal-to-noise ratio. It is worth noting that bands 1 and 2 are the same bands used by MODIS for temperature sounding. For this study, a nominal FWHM of 0.05 μm was chosen to build a Gaussian-shape sensor response function for each band.

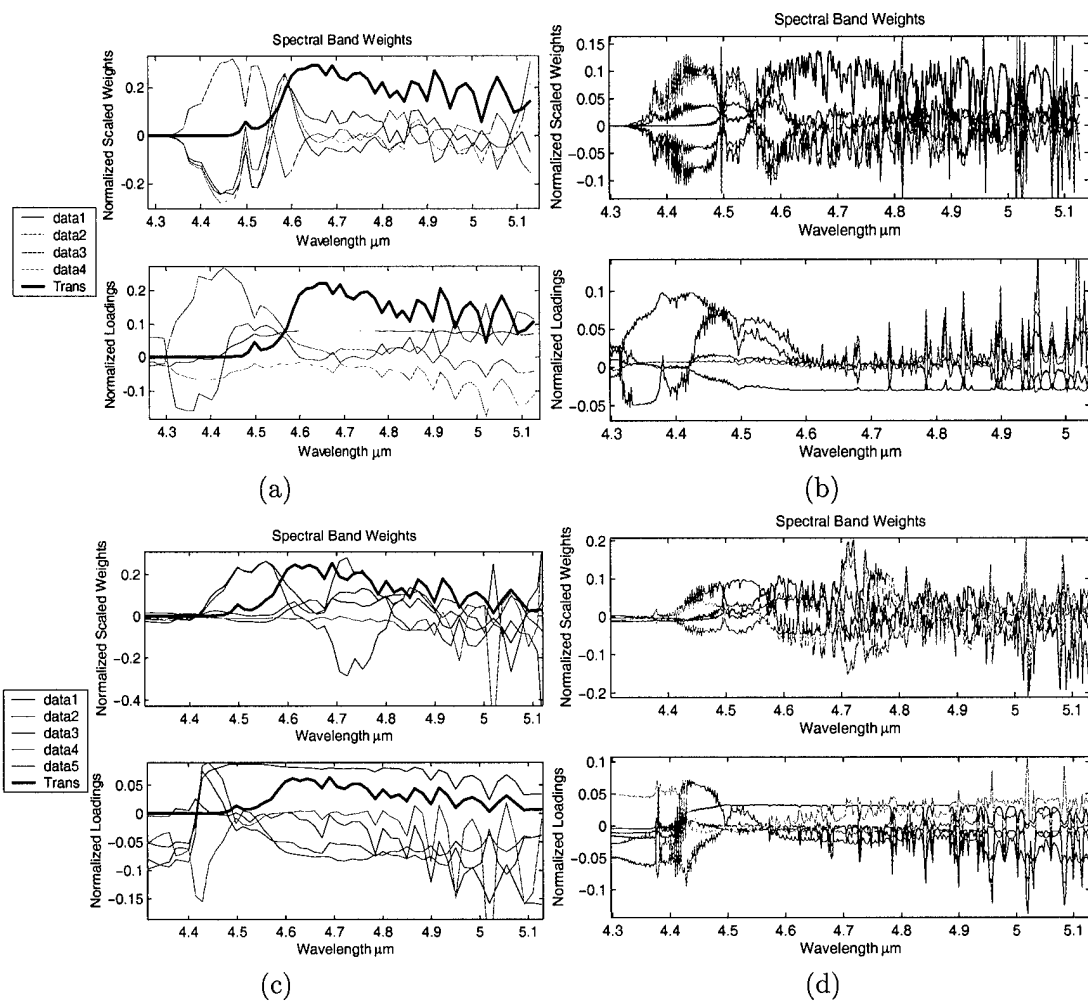


Figure 4.17: Canonical weights and loadings for MWIR emissive region: (a) FSL medium resolution, (b) FSL high resolution, (c) NAST-I medium resolution, (d) NAST-I high resolution.

Table 4.2: Summary Band-Selection Results

CONFIGURATION	NAST-I RMS ERRORS		FSL RMS ERRORS	
	Direct (°K)	CWV (mm)	Direct (°K)	CWV (mm)
5 bands	0.78	4.97	0.80	2.52
Bands 1,2	1.57	9.33	2.10	3.39
Bands 1,2,5	1.39	5.16	2.10	2.55
Bands 1,2,4,5	1.01	4.94	0.81	2.69

4.3.3 Testing and Validation of Band Configurations

Surface temperature and column water vapor estimates were obtained using the bands selected in Section 4.3.2. The RMS errors in these parameters serve as a metric that may be used to gauge the performance of the selected band configuration and are shown on Table 4.2. The results show that bands 1 and 2 provide information about the temperature profile but not about the surface temperature. For both data sets, using just bands 1 and 2 resulted in the highest error in estimated surface temperature. The same was true for column water vapor. This is not surprising since there are no resolvable water vapor features at these band locations. The addition of band 5 reduces the error in column water vapor significantly while not affecting the surface temperature retrieval. The 3-band configuration produces a better estimate of column water vapor at the expense of increased error in surface temperature, indicating that the temperature and water vapor errors are not mutually exclusive. This is because the temperature and water vapor profiles are estimated simultaneously. Figure 4.18 shows how the retrieved error profiles are affected by the different band configurations. The black profiles are the retrievals obtained with the medium resolution case and are shown as a reference. In general, the profiles were not affected significantly with the exception of the 2-band configuration (red profiles).

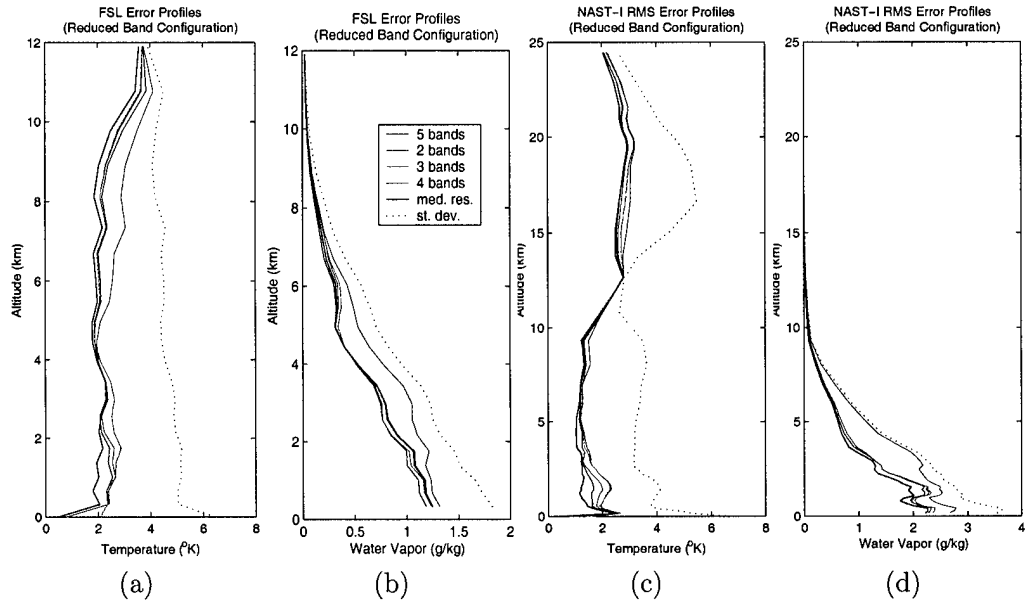


Figure 4.18: RMS error profiles for different band configurations.

4.4 Experiment #4

The design for this experiment is shown in Figure 4.19. The CCR inverse models built for this experiment were the most extensively tested because surface emissivity variations were included. The models were tested with MASTER thermal imagery as well as SEBASS and high resolution simulations.

Table 4.3 summarizes the data used to build the inverse models. The italicized runs correspond to the White River Valley and Railroad Valley collects. The other runs are for the Lake Mead collect. Separate ensembles were needed because of the altitude differences. The summary statistics give an indication of the nominal climatological conditions. These are the average surface temperature (μ_{T_s}), standard deviation of the surface temperature (σ_{T_s}), the average standard deviation of the temperature profile (σ_T), the average column water vapor (μ_{cww}), and the standard deviation of column water vapor (σ_{cww}). Of the three major climatologic datasets, the FSL most closely resembles the weather conditions found

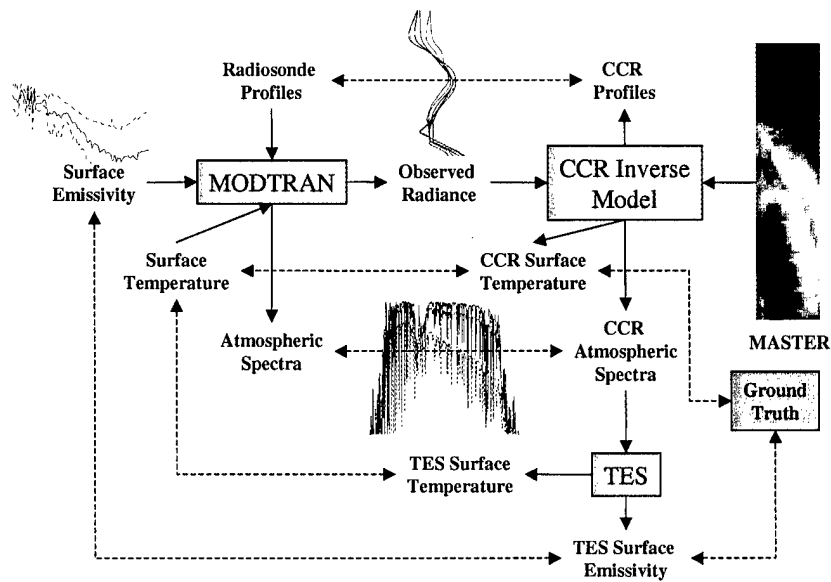


Figure 4.19: Schematic of design of Experiment # 4

Dataset	Geographic Coverage	Time Span	Run Seed	μ_{Ts} (°C)	σ_{Ts} (°C)	σ_T (°C)	$\mu_{c_{wv}}$ (mm)	$\sigma_{c_{wv}}$ (mm)
FSL	34-38 °N, 115-119 °W	1995 1999	13406*	9.01	6.31	4.73	11.08	5.15
			1460	9.58	7.05	4.41	11.52	5.41
			4768*	8.64	6.04	4.17	14.11	7.44
			17836	9.40	7.47	5.12	14.14	7.31
NAST-I	East Coast, U.S.A.	Jul-Sep 1998	24603*	22.77	6.17	3.63	38.97	12.83
			23072	23.32	6.94	3.05	38.12	12.36
			6085*	23.32	6.34	3.08	42.39	15.00
			1280	22.71	7.02	3.63	38.45	14.83
SSEC	Worldwide	1963- 1972	30165*	0.97	20.09	13.77	17.28	17.92
			5129	-2.61	19.26	12.86	13.60	15.40
			4984*	-11.05	15.95	10.97	8.31	6.57
			8729	-1.71	19.84	14.57	17.04	18.63

Table 4.3: Atmospheric database description and statistics for Experiment #4. Italic runs correspond to Lake Mead. Other runs are for Railroad/White River Valley. *Generated using blackbody targets.

Database	Avg. Profile RMS Error (°K)		CWV RMS Error (mm / %)	
	Blackbody	Variable Emissivity	Blackbody	Variable Emissivity
FSL	2.14	2.77	2.78 / 33.2	6.02 / 52.8
NAST-I	1.87	2.39	6.73 / 16.8	9.81 / 26.6
SSEC	2.30	2.99	2.84 / 49.4	8.34 / 130

Table 4.4: Temperature and column water vapor errors for Lake Mead.

Database	Avg. Profile RMS Error (°K)		CWV RMS Error (mm / %)	
	Blackbody	Variable Emissivity	Blackbody	Variable Emissivity
FSL	2.14	2.77	2.78 / 33.2	6.02 / 52.8
NAST-I	1.87	2.39	6.73 / 16.8	9.81 / 26.6
SSEC	2.30	2.99	2.84 / 49.4	8.34 / 130

Table 4.5: Temperature and column water vapor errors for Railroad/White River Valley.

in the MASTER images since it is based on observations made in the same region. The NAST-I databases tend to be more humid than the FSL and SSEC data. Finally, the SSEC has the highest amount of variability. In general, there's a high degree of variation that must be accounted by the model.

4.4.1 Atmospheric Parameter Retrievals

Atmospheric physical parameters (i.e., temperature and water vapor profiles) were estimated from simulated MASTER "at-sensor" spectral radiance observations using the CCR inverse model. Tables 4.4 and 4.5 show the average RMS error in the temperature profiles and column water vapor for the Lake Mead and Railroad/White River Valley configurations, respectively. The results are shown for both blackbody and variable emissivity cases. The errors do not appear to be dependent on the type of surface used in the retrievals. On the other hand, the errors differ quite significantly depending on which atmospheric ensemble is analyzed. More specifically, the errors are the lowest for the NAST-I set and the largest for the SSEC data.

There are two major reasons why the errors for the SSEC data are so large. The most important factor is the large variability in the data with respect to the average temperature

and water vapor concentrations. The lowest error for the SSEC retrievals was obtained with the blackbody run for Lake Mead. Table 4.3 shows that this particular run had the least amount of variation of all of the SSEC runs. The other source of error is the discretization of the vertical profiles. This error is common to both the SSEC and FSL data. For example, the Railroad/White River Valley collects were done at an altitude of about 10 km. At this altitude, only 18 discrete atmospheric layers (between 300 and 900 mb) are defined for the FSL data and only 17 (between 300 and 1050 mb) are defined for the SSEC data. In contrast, 26 layers (between 350 and 1000 mb) are defined for NAST-I.

Another source of error, independent of the atmospheric database used, is the spectral resolution of the MASTER sensor. Figure 4.20 on page 161 shows the canonical weights and loadings for the MASTER sensor obtained with the NAST-I (Run 23072) data. Three significant correlations were identified. The canonical weights for the sensor observations emphasize the overall shape of the continuum as well as water vapor and ozone band absorption features. The canonical weights for the temperature and water vapor profiles exhibit a high degree of variability, indicating that information is being extracted from the profiles and that the profiles are highly correlated. The loadings for the temperature profiles are broad—a direct result of the low spectral resolution. Water vapor loadings are also broad but have a more definable shape. Figure 4.21 on page 162 shows the canonical weights and loadings for a high-resolution configuration (maximum MODTRAN resolution resulting in 751 bands between 650 and 1400 cm^{-1}) implemented with Run 23072. There are now 8 significant correlations and there is quite a bit more structure in the temperature and water vapor loadings. The canonical weights for the sensor observations show that an appropriate emphasis is being placed on the continuum as well as the major water vapor band absorption and narrow line absorptions. The average RMS error in the temperature profiles estimated with the high-resolution observations was 1.21 °K. The RMS error in column water vapor was 3.76 mm (10.3%). These are significant improvements over the MASTER (NAST-I variable emissivity) retrievals shown in Table 4.5. The errors for the SSEC data (variable

emissivity) were reduced to 2.02 °K and 6.23 mm (113%) for temperature and column water vapor, respectively. However, the errors are still too large which indicates that the increased spectral resolution could not completely account for all of the variability in the data.

In general, the atmospheric optical parameter retrievals were very accurate for the FSL and NAST-I data. Figure 4.22 on page 163 shows the RMS errors in transmission, upwelled radiance, and downwelled radiance for SEBASS simulations using Run 23072 (NAST-I data). The simulations were done with MODTRAN and with the same sensor altitude as for the MASTER airborne images. An example of the actual retrieved spectra is shown for observation 412 (chosen randomly). The highest RMS error for transmission is about 0.07 transmission units while the highest RMS error for upwelled and downwelled radiance was about 18%. As with the physical parameter retrievals, the errors for the SSEC data were much larger. This was particularly so for the retrievals of upwelled and downwelled radiance which had RMS errors of up to 75%! The next section demonstrates the effect of this error on the retrieval of surface temperature and emissivity. It should be noted that all of these retrievals were done under the influence of varying surface temperature and emissivity.

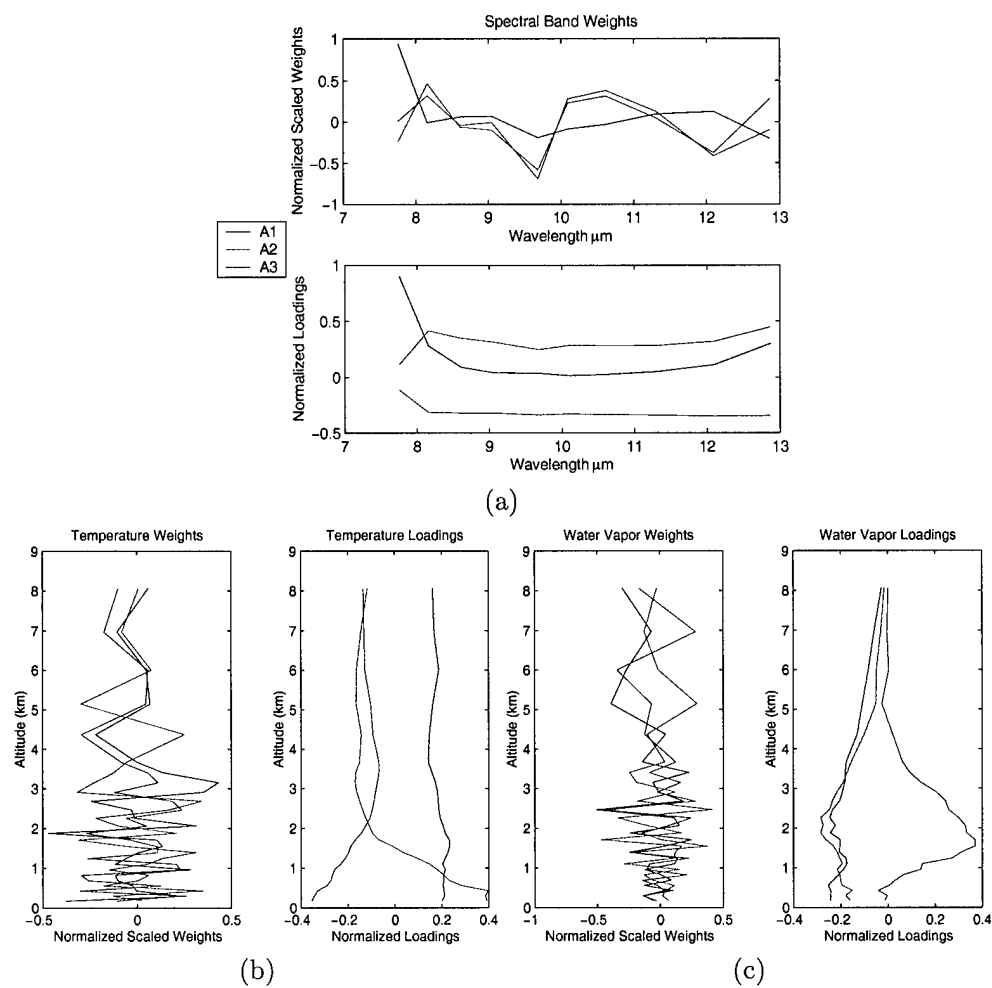


Figure 4.20: NAST-I (Run 23072) canonical weights and loadings for MASTER: (a) sensor observed spectra, (b) temperature profiles, and (c) water vapor profiles.

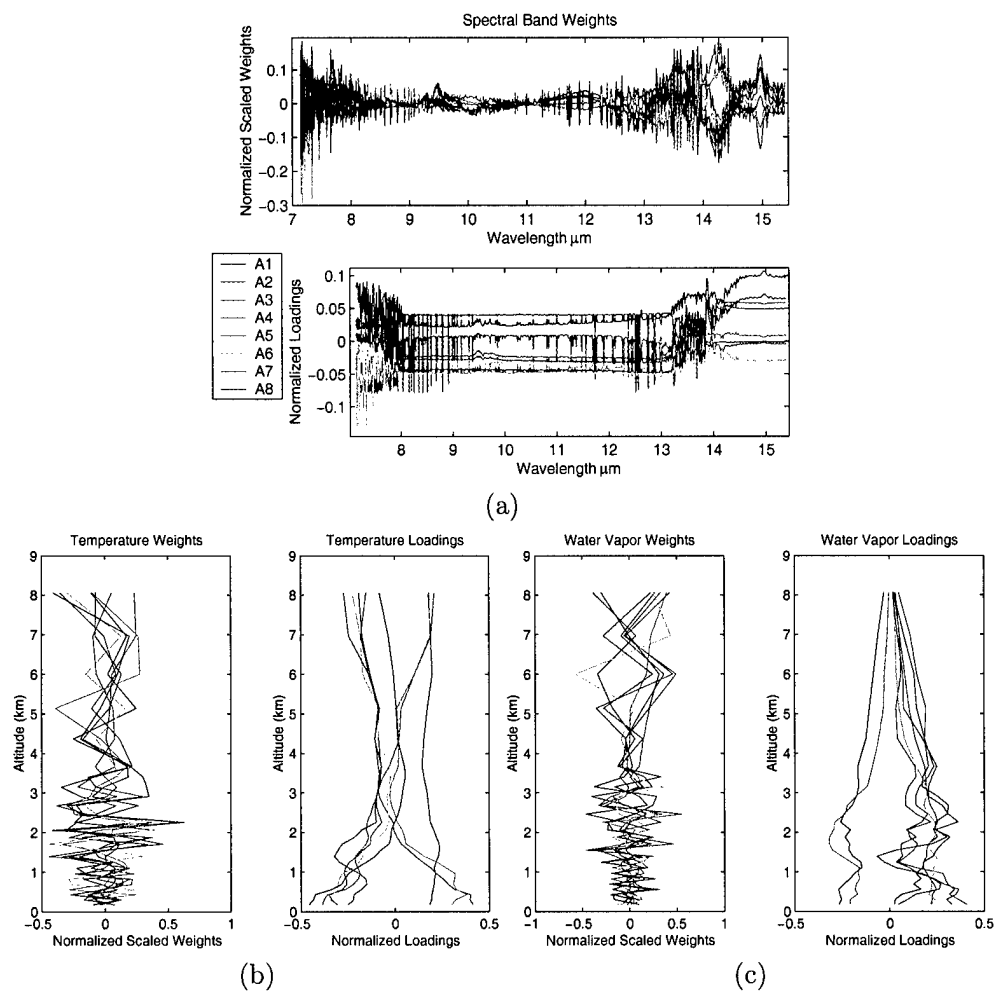


Figure 4.21: NAST-I (Run 23072) canonical weights and loadings for high-resolution case: (a) sensor observed spectra, (b) temperature profiles, and (c) water vapor profiles.

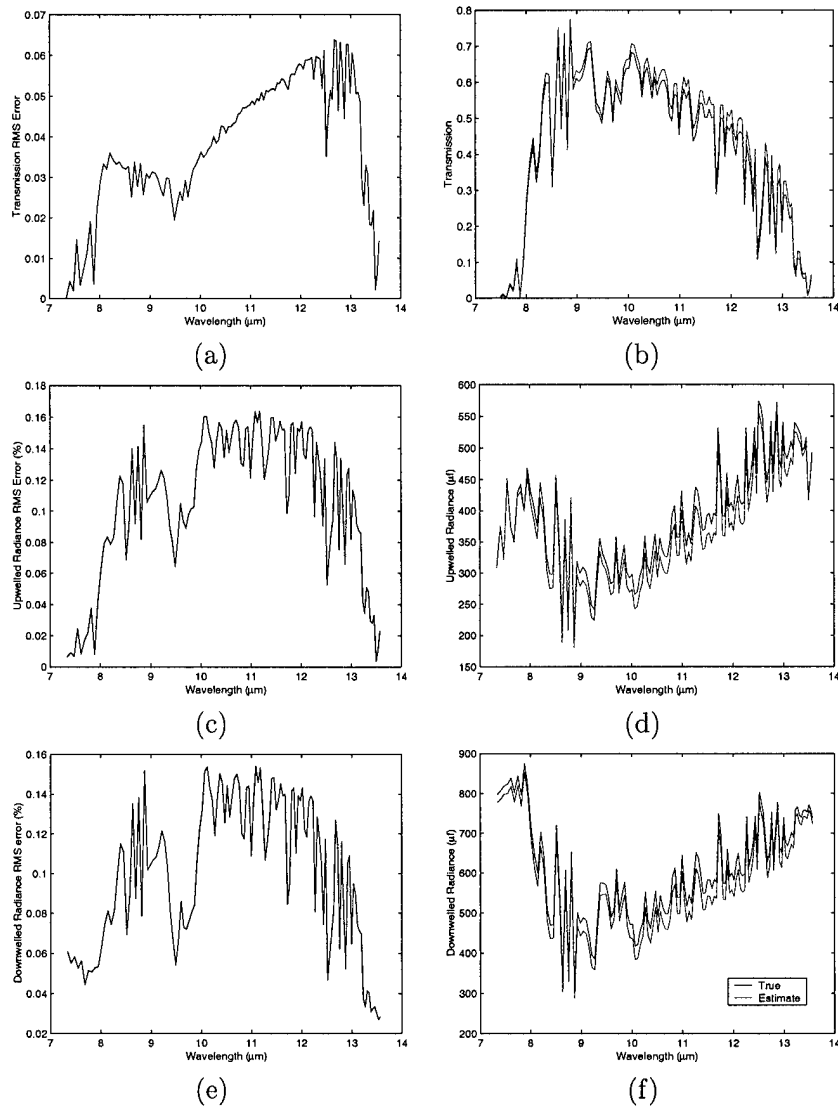


Figure 4.22: RMS Errors and retrievals of atmospheric spectra from SEBASS-resolution NAST-I data (Run 23072).

Test Case	Simulated MASTER T_s (RMS °K)	MASTER (L. Mead & C. Springs) T_s (RMS °K)
<i>Lake Mead</i>		
FSL	0.11	0.37
NAST-I	0.15	0.71
SSEC	0.08	0.33
<i>Cold Springs</i>		
FSL	0.07	0.71
NAST-I	0.16	1.16
SSEC	0.27	0.88

Table 4.6: Errors in retrieved surface temperature for blackbody targets.

4.4.2 Surface Temperature and Emissivity Retrievals

Blackbody Target Results

The estimation of surface temperature when the target is a blackbody is unique because the problem is greatly simplified. For these cases, two unknowns are removed: the surface emissivity and downwelled radiance. This allows the use of a simpler approach. For these cases, the observed brightness temperatures were related directly to the surface temperature T_s with the CCA inverse model. The rationale for the use of brightness temperatures is discussed in Section 4.4.3.

Table 4.6 shows the results obtained with the Lake Mead and Cold Springs (located in White River Valley) configurations, respectively. The first column is the RMS error in the retrieved surface temperature for the observations used to build the model. The second column is the RMS error in the retrieved surface temperature for the actual MASTER observations at Cold Springs reservoir and Lake Mead. The results are very accurate for the simulated blackbody target case. The model validation shows that the highest error introduced by the procedure is 0.27 °K. This occurs when the SSEC database is used, which is not surprising considering this database had the highest variability in surface temperature. Thus, this residual error is due to variance unexplained by the CCA inverse model. The RMS errors increase for the actual MASTER observations. This is expected since these errors

include not only the model errors, but sensor noise, calibration error, ground measurement error, and pixel registration error as well. Except for the NAST-I results, the Lake Mead errors were within the uncertainty due to sensor noise and the White River errors were slightly higher than that. This is because the NAST-I atmospheres are much more humid than the dry Nevada atmospheres. The larger errors for the White River Valley MASTER retrievals may be due to uncertainty in the ground surface temperature measurements. The measured temperature was actually bulk water temperature, which is not necessarily the same as the *skin* (or *kinetic*) surface temperature (Palluconi 2000). The skin temperature for Cold Springs was assumed to be 0.5 °K lower than the bulk water temperature. This assumption was based on comparisons made with other similar measurements (Palluconi 2000). Finally, the Cold Springs image was taken at a higher altitude.

Varying Emissivity Results

The surface temperature for these cases was estimated using the direct approach used for blackbody targets and the atmospheric compensation-TES approach. Table 4.7 shows the RMS errors in the retrieved surface temperatures for the various test cases. The “Simulated MASTER” column contains the results from comparing the input surface temperatures and those derived with the CCA models. These results were obtained using the MASTER spectral resolution. The “MASTER” (middle) column contains the RMS errors in the retrieved surface temperatures based on the Lake Mead or Cold Springs MASTER observations. Finally, the “Simulated SEBASS” column contains the errors in the retrieved surface temperatures using the the same observations as in the “Simulated MASTER” column but resampled to SEBASS resolution. All available spectral bands were used to estimate the surface temperature directly. This was also true for the estimation of the atmospheric parameters. However, only bands where $\tau(\lambda) > 0.4$ were used when applying eq. (3.65) for compensation of the SEBASS observations. This was done to minimize numerical instability in the solution. The sensor altitude for the White River collect was high enough to

Test Case	Simulated MASTER		MASTER (L. Mead & C. Springs)		Simulated SEBASS	
	TES RMS (°K)	Direct RMS (°K)	TES RMS (°K)	Direct RMS (°K)	TES RMS (°K)	Direct RMS (°K)
<i>Lake Mead</i>						
FSL	2.81	1.13	0.81	1.87	2.50	0.60
NAST-I	2.51	1.19	0.65	1.75	2.33	0.53
SSEC	2.68	1.99	0.99	2.70	-	1.24
<i>Cold Springs</i>						
FSL	2.83	1.45	0.67	3.50	2.28	0.47
NAST-I	2.30	1.91	0.61	1.95	2.11	0.55
SSEC	3.60	2.59	1.40	2.05	-	1.23

Table 4.7: Errors in retrieved surface temperature for variable emissivity targets.

make ozone absorption significant. However, ozone effects were held constant in the atmospheric databases so they were not accounted for. Thus, band 45 (centered at the 9.6 μm ozone absorption feature) was removed from the analysis of the MASTER White River observations.

The RMS error in the TES surface temperatures obtained from the simulated MASTER observations were about 2.5 to 3.5 °K. This high residual error was due to errors in the retrieved emissivities. Figure 4.23 shows the retrieved emissivity for the farm and desert classes using the MASTER resolution and the FSL/White River database. The bias error for the farm emissivity retrievals was 0.09 with a standard deviation of 0.03. The bias is due to the prediction of ε_{min} from the MMD. The MMD for the farm emissivity is about 0.15, which leads to $\varepsilon_{min} \approx 0.85$. This value is shown on the graphs as a dotted line. Note that the average minimum value of the retrieved emissivity is 0.85. For the desert emissivity, the true MMD is about 0.06 corresponding to $\varepsilon_{min} \approx 0.92$. The desert retrievals had a large error in band 41 due to atmospheric compensation. This error offset the overall bias by making the MMD larger—thus reducing ε_{min} . This offset was about 0.02. The overall bias between the predicted and the true desert emissivity was about 0.05. Again, this is mostly due to the MMD regression line. The errors in the temperatures retrieved directly from the observed brightness temperatures were about 1 °K lower than the TES-derived

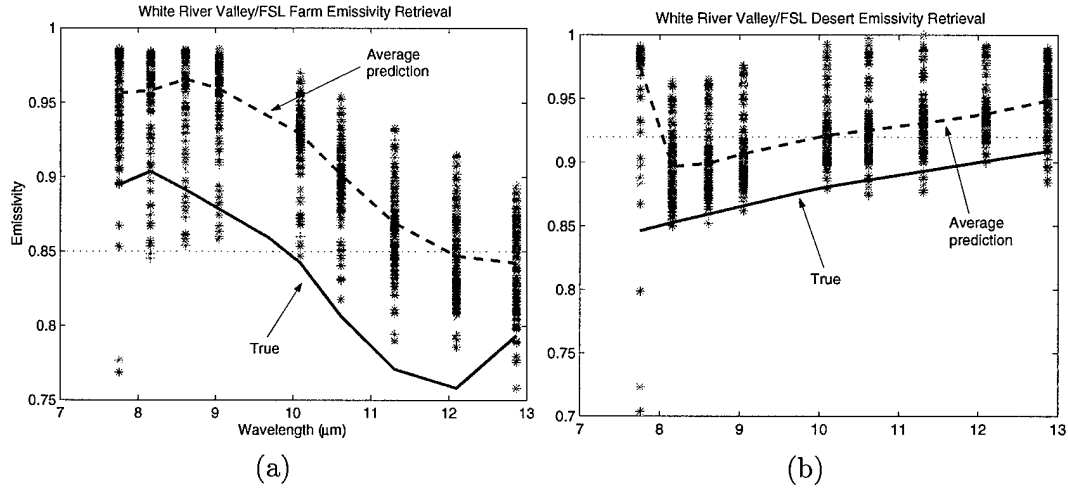


Figure 4.23: Emissivity retrieval for MASTER simulations: a) farm and b) desert classes.

temperatures. This indicates that the error introduced by the compensation computation is about 1 °K. Also, these errors are relatively low, indicating that CCA is able to find a model for the prediction of temperature that is relatively insensitive to emissivity and atmospheric variations.

The TES-derived temperatures from the MASTER images were all within 1 °K of the ground measurements (with the exception of the SSEC dataset used for Cold Springs). These results are much better than for the entire dataset because the target was water, which has high emissivity. Figure 4.24a shows the emissivity estimated from MASTER Cold Springs observations and the ocean emissivity class used in MODTRAN. The TES emissivity is an unbiased estimate except at the edges of the spectrum. Despite the sensor noise, the average deviation in the emissivity retrievals was less than 0.01. The errors in the temperatures derived directly from the MASTER brightness temperatures were higher by about 1-2 °K. This may be due to the lower resolution of the MASTER spectra, which limits the model's ability to separate the atmospheric and emissivity variation from the surface temperature. The reason why this does not manifest itself in the atmospheric compensation and TES process is because the information about the atmospheric parameters

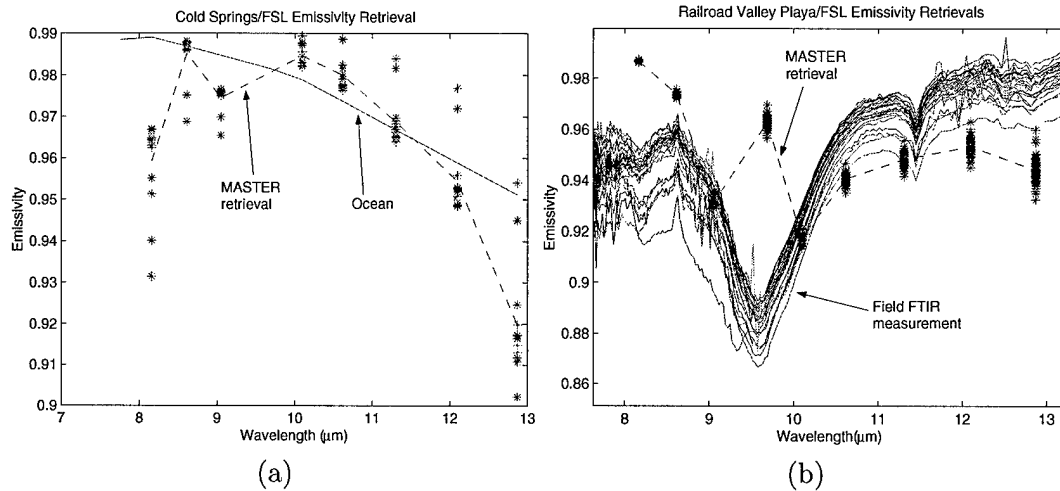


Figure 4.24: MASTER emissivity retrievals for: a) Cold Springs reservoir and b) Railroad Valley playa.

is in the regions of high absorption where the contribution from the surface emission is low. Figure 4.24b shows the emissivity estimated from MASTER Railroad Valley observations. These spectra are compared to measurements made with a field FTIR (Palluconi 2000). Except for band 45, the spectra agree quite well. The feature in the playa spectra matches that of ozone absorption. When the CCA model is built including this band, the model interprets the feature as due to atmospheric absorption and compensates for it. Thus, the feature is eliminated from the resulting emissivity. This is an inherent problem with the retrieval of emissivity spectra that have reflectance features at the same location and of the same width as atmospheric absorption spectra. As the emissivity values for these features decrease, the amount of reflected downwelled radiation increases. Depending on the surface and sky temperature, this reflected radiance can “mask” the emissivity feature completely.

The SEBASS results show that increasing the number of bands and spectral resolution does not lead to better estimates of surface temperature when using the atmospheric compensation and TES approach. The errors in the TES-derived temperatures are about the same as those obtained with MASTER resolution observations. This is because the

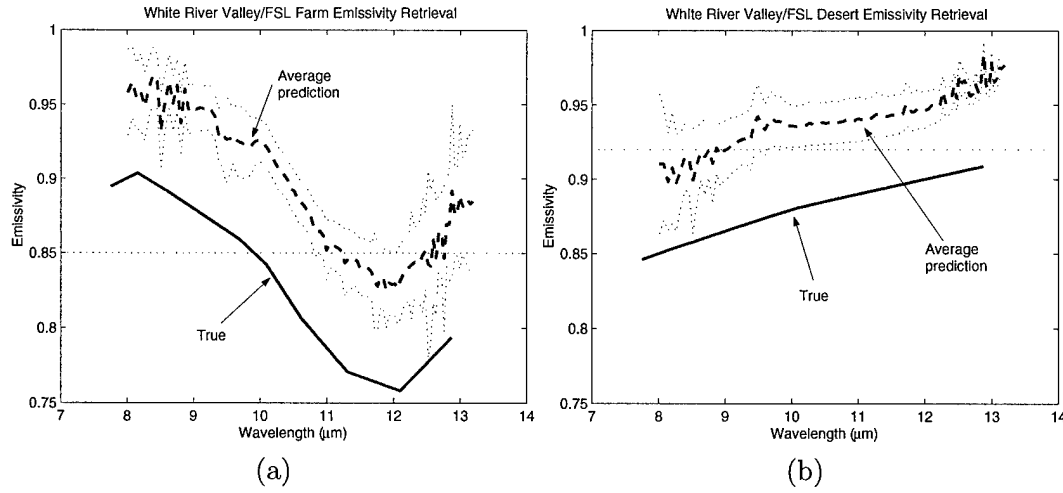


Figure 4.25: Emissivity retrieval for: a) farm and b) desert classes using SEBASS resolution.

error is still being driven by the bias from the MMD regression. Figure 4.25 illustrates this by comparing the retrieved and true emissivity for the farm and desert classes. The dotted lines about the average prediction curves are one standard deviation away from the mean. The results were nearly the same as those obtained with the MASTER resolution. The spectral error (i.e., the relative band emissivity error) was generally low but larger at the edges of the LWIR bandpass. These regions are characterized by strong and narrow water vapor absorption features. For the SSEC database, the errors were actually larger. Because of this, TES was not able to converge to a solution (see Section 4.4.3 for a detailed explanation). The direct temperature retrievals, however, are significantly improved. This is because there is more spectral information that can be used without introducing more unknowns. That is, as the number of bands p grows larger, $p+1 \approx p$. Also, as the resolution increases, the difference in the emissivity between adjacent bands decreases. This results in several observations with approximately the same temperature and emissivity, thus improving the accuracy of the temperature estimate. With the exception of the SSEC data, all the retrievals had an accuracy of about 0.5-0.6 °K. Again, the larger errors for SSEC are due to the large variability in the global database.

4.4.3 Other Findings

Observed Radiance vs. Brightness Temperature

In this experiment, the brightness temperatures were used instead of the observed radiance for the estimation of surface temperature and atmospheric profiles. Recall that the brightness temperature is the apparent temperature at *each* wavelength and is obtained by inverting the Planck function (see eq. (2.84)). This was done because the CCR model depends on the strength of *linear* correlations. However, the observed radiance is proportional to the Planck function, which is nonlinear with respect to temperature. This is particularly the case when the temperature has a large range. This nonlinearity is demonstrated in Figure C.1 in Appendix C. On average, the error in surface temperature retrievals from the brightness temperature was about 0.25 °K less than for estimates obtained from the observed radiance. The estimation of water vapor from brightness temperatures was tested as well since the temperature profiles are solved simultaneously with the water vapor profiles. Analysis showed that errors in water vapor estimates obtained from the brightness temperatures were not significantly different than those obtained from radiance. Figure 4.26 shows two examples of RMS errors in water vapor profiles obtained from CCR using brightness temperature $T_b(\lambda)$ and observed radiance $L(\lambda)$.

Issues with Unconstrained Solutions

In all previous experiments, the solutions obtained with the CCR inverse model were all physical. That is, the solutions did not take on values that were not physically possible (e.g., $\tau > 1$ or $L_u < 0$). This fortunate state of affairs was not attained by design since the solutions from the CCR inverse model were not constrained in any way.

In this experiment, some of the solutions for transmission spectra had values greater than 1.0 or less than 0.0. This occurred only with the CCR inverse model built with varying emissivity and SEBASS spectral resolution (Figure 4.27). The error was the worst

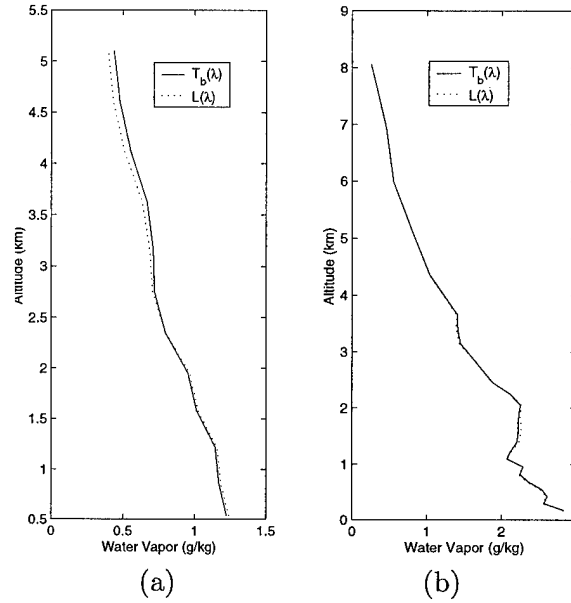


Figure 4.26: Comparison of water vapor profile retrievals obtained with the sensor brightness temperature and observed radiance. The results were obtained from CCR inverse models built with (a) FSL Run 4768 and (b) NAST-I Run 23072.

for the SSEC training set. The same problem occurred with the estimation of upwelled and downwelled radiance from the CCR inverse model built with the FSL and SSEC data. Again, the SSEC estimates exhibited the worst errors. One plausible explanation for this problem is that the errors are larger when the atmospheres in the ensemble are thin. Thus, the emissivity variability is more likely to affect the observed radiance at the sensor. On average, the FSL and SSEC atmospheres are thinner than the NAST-I atmospheres; thus explaining the increased error in these data.

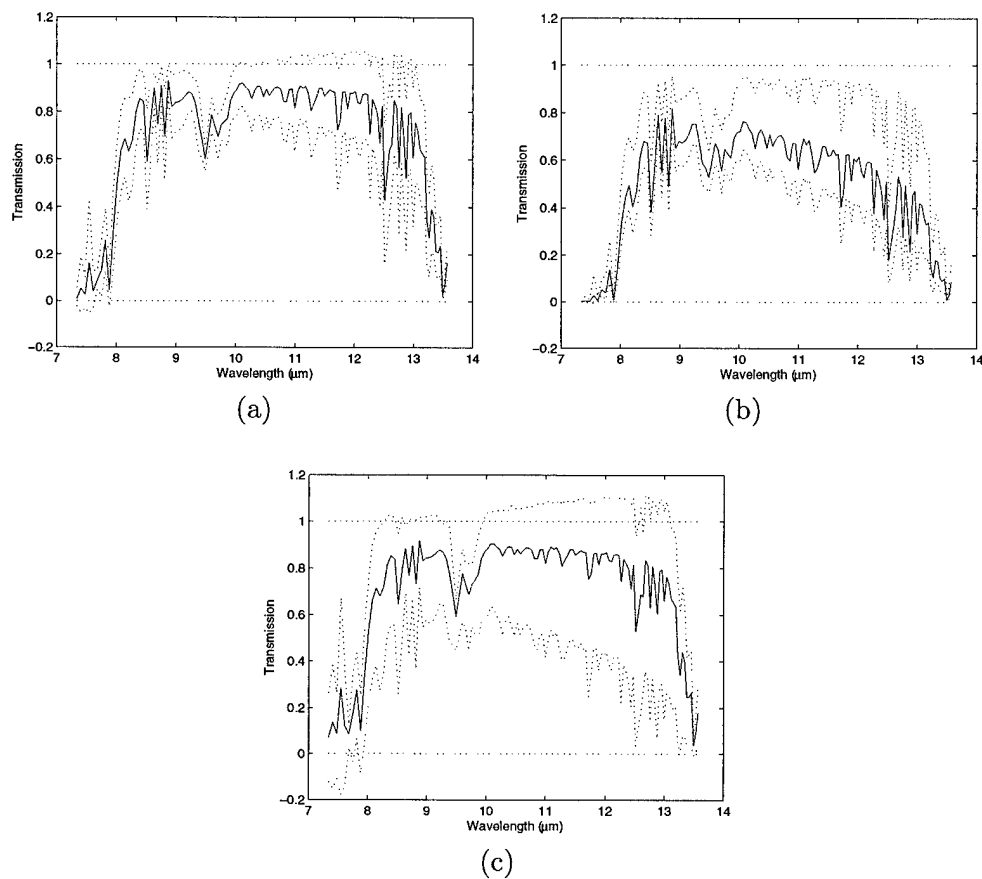


Figure 4.27: Transmission spectra retrievals for SEBASS simulations with variable emissivity: a) FSL; b) NAST-I; and c) SSEC. The solid curves are the mean values in the training set. The dotted curves are maximum and minimum estimates obtained with CCR.

4.5 Comparison of Multivariate Regression Methods

This section compares the CCR inverse model estimates with those obtained from other multivariate regression methods. Three other methods are tested: 1) Principal Components Regression (PCR), (2) Maximum Redundancy (MR), and (3) Partial Least Squares (PLS). A description of these methods is provided in Appendix D.

Tables 4.8, 4.9, and 4.10 show the RMS errors obtained with the multivariate regression methods. The test case was Run 23072 from Experiment #4 at different spectral resolutions. The dimensionality of the multivariate models was the same within each test case. In general, CCR performed better than MR, PCR, and PLS. This was particularly the case for surface temperature estimates. The largest difference between CCR and the other methods occurred with the SEBASS and high resolution test cases. The results obtained with MR, PCR, and PLS were about the same except for the estimate of column water vapor where PLS yielded better estimates than MR and PCR (and slightly better—although probably not significantly better—than CCR for the high resolution test case).

Parameter	PCR	CCR	MR	PLS
T_s RMS ($^{\circ}\text{C}$)	2.05	1.91	2.05	2.01
Temp. profile RMS ($^{\circ}\text{C}$)	2.99	2.97	2.98	2.98
CWV RMS (mm)	12.10	12.12	12.10	12.10

Table 4.8: Comparison of multivariate methods for the MASTER resolution case (10 bands). Three dimensions were retained for all methods.

Parameter	PCR	CCR	MR	PLS
T_s RMS ($^{\circ}\text{C}$)	1.91	0.56	1.91	1.90
Temp. profile RMS ($^{\circ}\text{C}$)	1.56	1.47	1.54	1.54
CWV RMS (mm)	5.49	4.49	5.33	5.23

Table 4.9: Comparison of multivariate methods for the SEBASS resolution case (128 bands). Five dimensions were retained for all methods.

Parameter	PCR	CCR	MR	PLS
T_s RMS ($^{\circ}\text{C}$)	1.85	0.51	0.75	0.75
Temp. profile RMS ($^{\circ}\text{C}$)	1.84	1.80	1.81	1.80
CWV RMS (mm)	4.38	4.22	4.22	4.21

Table 4.11: Comparison of multivariate methods for the MWIR medium-resolution case (200 bands). The multivariate model dimensionality was 5.

Parameter	PCR	CCR	MR	PLS
T_s RMS ($^{\circ}\text{C}$)	3.11	0.80	0.80	0.80
Temp. profile RMS ($^{\circ}\text{C}$)	2.06	1.99	1.99	1.96
CWV RMS (mm)	5.11	4.96	4.95	4.94

Table 4.12: Comparison of multivariate methods for the MWIR 5-band resolution case. The multivariate model dimensionality was 3.

Parameter	PCR	CCR	MR	PLS
T_s RMS ($^{\circ}\text{C}$)	1.67	0.54	1.67	1.68
Temp. profile RMS ($^{\circ}\text{C}$)	1.42	1.31	1.41	1.38
CWV RMS (mm)	4.51	4.16	4.36	4.12

Table 4.10: Comparison of multivariate methods for the high resolution case (751 bands). Eight dimensions were retained for all methods.

The MWIR NAST-I ensemble generated in Experiment #3 was also tested. Tables 4.11 and 4.12 show the results. Again, CCR outperformed MR, PCR, and PLS in the estimation of surface temperature. Otherwise, the methods yielded almost identical results for the medium resolution test case. The results obtained with the CCR-selected bands show that MR and PLS are able to adequately exploit the information contained in these bands. However, PCR residuals are considerably higher. This may be due to the methodology used for band-selection, which was to maximize the information about the parameters of interest based on the CCR weights and loadings. The resulting configuration appears to have principal components that are poor estimators of the parameters of interest.

The methods were also applied to Run 5129 at SEBASS resolution (128 bands). Sections 4.4.3 and 4.4 described the difficulties associated with this test case. The goal of

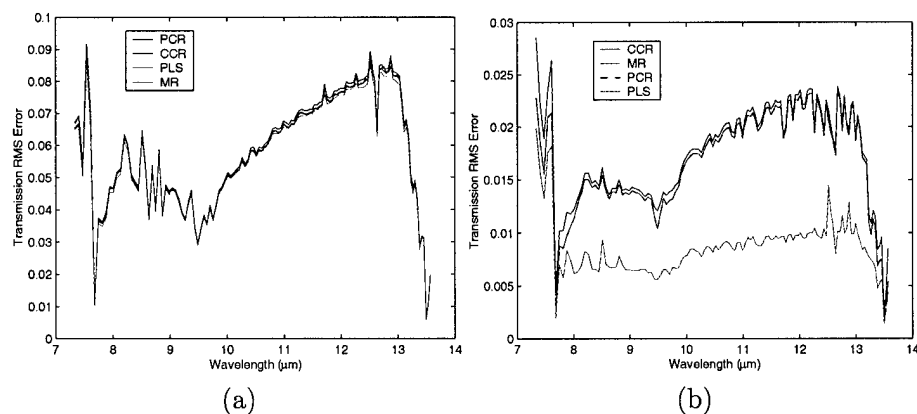


Figure 4.28: Comparison of transmission spectra RMS errors with (a) 4 dimensions and (b) 40 dimensions.

this exercise was to see how well the different methods could estimate atmospheric spectra within the respective physical boundaries. Figure 4.28 shows the RMS errors in transmission spectra estimates obtained with 4 and 40 dimensions. All of the multivariate methods give very similar results for the low-dimensional case. If 40 dimensions are kept, the overall RMS error decreases for all of the methods. However, PLS errors are considerably lower than for CCR, MR, and PCR. As the number of dimensions approaches the number of original variables, MR and PCR converge to the same solution. This is because all of the relevant information about $\mathbf{X}$ and $\mathbf{Y}$ is contained in the principal components as the number of dimensions increases. Therefore, a regression of the principal components in PCR is almost identical to the regression of the principal components and the OLS estimate of $\mathbf{Y}$ in MR. It is interesting to note that CCR performs about the same as PCR and MR. This suggests that the canonical correlations and weights of $\mathbf{Y}$ explain as much of the variance in $\mathbf{Y}$ as the principal components do. Figure 4.29 shows the maximum transmission and minimum upwelled radiance spectra estimated with the multivariate methods when 40 dimensions were retained. The $\tau = 1.0$ boundary (red line) is shown in Figure 4.29(a) as a boundary reference. Likewise, the reference boundary $L_u = 0.0$ is shown in Figure 4.29(b). The MR results (not shown) were identical to the PCR estimates. Again, CCR and PCR

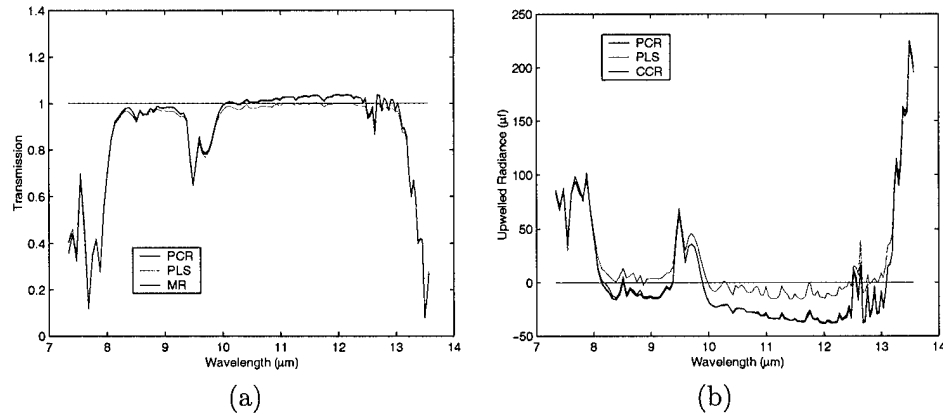


Figure 4.29: Comparison of (a) transmission and (b) upwelled radiance residuals with 40 dimensions retained.

give similar results with the PLS estimates being the closest to the physical boundaries for transmission and upwelled radiance.

4.6 Comparisons to ISAC

The CCR inverse model approach was compared to the generally accepted In-Scene Atmospheric Compensation (ISAC) approach. Three implementation of the ISAC algorithm (as described in Appendix E) were tested: (1) Kolmogorov-Smirnov (KS); (2) Normalized Regression (NR); and (3) SITAC Kolmogorov-Smirnov. The comparison was done in terms of errors in the surface temperature and emissivity retrievals. To this end, CCR and ISAC were coupled with TES. However, TES requires an estimate of downwelled radiance, which ISAC does not compute. To make the comparison fair, the estimated downwelled radiance for ISAC had to be obtained. A comparison of MODTRAN upwelled and downwelled radiance calculations showed that the downwelled radiance can be estimated from the upwelled radiance through a scalar factor:

$$L_d(\lambda) \approx 1.6L_u(\lambda) \quad (4.4)$$

Method	Lake Mead RMS (bias) °K	C. Springs RMS (bias) °K
CCR	0.81 (-0.75)	0.67 (0.14)
ISAC-KS	0.17 (-0.07)	0.67 (-0.33)
ISAC-NR	0.16 (-0.02)	0.69 (0.22)
ISAC-SITAC	0.30 (0.26)	0.63 (-0.23)

Table 4.13: Comparison of CCR and ISAC surface temperature retrievals with TES for Lake Mead and Cold Springs reservoir.

Using this approximation, the ISAC downwelled radiance could then be calculated from the ISAC upwelled radiance. Clearly, this is a rough approximation since there is no consideration for the altitude of the sensor or for the heterogeneity of the sky composition (e.g., presence of clouds). Another option was to use the CCR estimates of downwelled radiance with the ISAC transmission and upwelled radiance estimates. Both methods yielded almost identical results, indicating that the sky inhomogeneity was negligible for the cases under consideration. Finally, the ISAC retrievals were based on the unscaled parameters.

Table 4.13 shows the results for Lake Mead and Cold Springs temperature retrievals. The CCR inverse model results were obtained using the correlations computed with the FSL (runs 17836 and 1460) data. RMS and bias temperature errors are listed for all methods. In general, the majority of the error in temperature is due to a bias in the estimate. ISAC performed significantly better than CCR for the Lake Mead case. For this case, the ISAC-TES temperature estimates were practically unbiased except for the SITAC implementation. In contrast, the difference in the retrievals for Cold Springs were negligible. The bias in the temperature estimates for CCR and ISAC were also about the same.

The results shown in Table 4.13 indicate that the temperature error is largely driven by the bias in the estimates. The question then becomes: are the differences in the biases due to TES or to the atmospheric compensation? The answer is both. However, our interest is in comparing the differences in *atmospheric compensation* between CCR and ISAC. The use of TES as the final temperature and emissivity retrieval step confounds the comparison.

Figure 4.30 shows the CCR and ISAC surface-leaving radiance retrievals for Lake Mead. Both retrievals resemble blackbody radiation curves. The Planck function for the average Lake Mead surface temperature is overlayed on top of the surface radiance retrievals. Although similar, there are some small differences in the shape of the curves. TES identifies this variation and translates it to variation in emissivity, thus introducing a bias in the surface temperature retrieval. CCR and ISAC were coupled with the Normalized Emissivity Method (NEM) to isolate the differences in biases due to the atmospheric compensation. The NEM is simply the first module of the TES algorithm. The results are not scaled by an empirical formula and are based only on the retrieved surface-leaving radiance and downwelled radiance. Table 4.14 shows the surface temperature errors obtained with CCR and ISAC coupled with NEM. The errors are generally higher, indicating that TES MMD-regression line is at least scaling the emissivities in the right direction (the exception is the CCR retrievals for Lake Mead where the NEM results are unbiased). The differences in the biases (and therefore the temperature errors) are due only to the atmospheric compensation step. In general, CCR results in lower errors than all of the ISAC implementations for both Lake Mead and Cold Springs. The ISAC errors are almost exclusively due to a bias in the temperature estimate. In general, the ISAC-NEM temperature retrievals are underestimates of the true surface temperature. This is likely due to the use of unscaled parameters. By using the unscaled parameters, the surface temperature estimate is basically the observed brightness temperature at the band with the highest transmission. ISAC unscaled parameters assume that the transmission is 1.0 at this band. Normally, the transmission is less than that, causing the brightness temperature to be lower than the surface temperature.

The Lake Mead and Cold Springs reservoir cases dealt with water targets, which are almost blackbody in nature. Emissivity retrievals for the Railroad Valley playa were obtained to compare the performance of CCR and ISAC with nonblackbody targets. Figure 4.31 shows the emissivity retrievals for ISAC_{NR} (the results were nearly identical for all ISAC

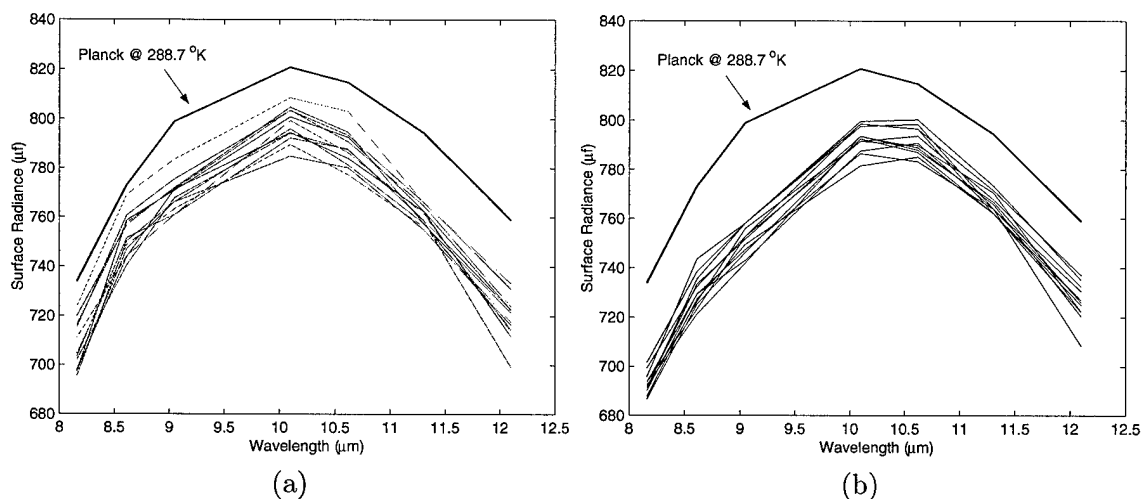


Figure 4.30: Surface radiance retrievals over Lake Mead for (a) CCR and (b) ISAC_{NR}. The Planck function for the average surface temperature is overlayed as a reference.

Method	Lake Mead RMS (bias) °K	C. Springs RMS (bias) °K
CCR	0.22 (-0.08)	1.30 (1.13)
ISAC-KS	1.66 (1.65)	1.80 (1.70)
ISAC-NR	1.25 (1.24)	1.64 (1.55)
ISAC-SITAC	1.69 (1.69)	1.98 (1.89)

Table 4.14: Comparison of CCR and ISAC surface temperature retrievals with NEM for Lake Mead and Cold Springs reservoir.

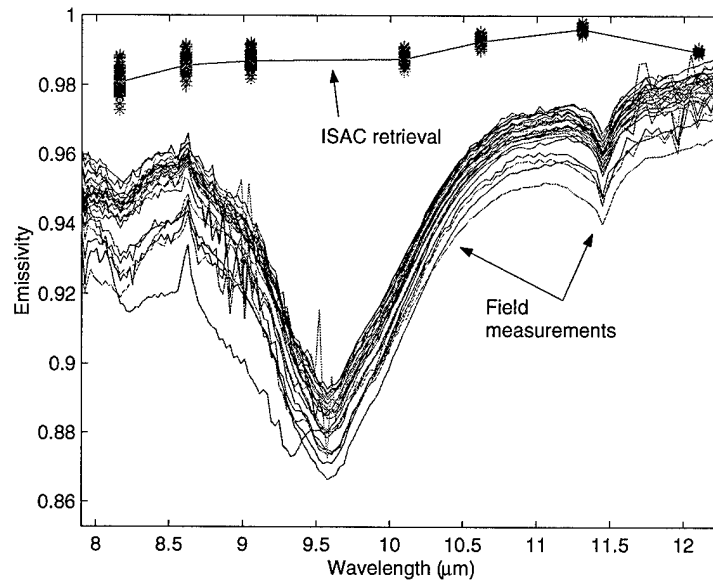


Figure 4.31: ISAC_{NR} emissivity retrievals for Railroad Valley playa.

implementations) and for CCR (with the FSL data). The ISAC-TES emissivity is biased by about 0.03 emissivity units and completely misses the feature around 9.6 μm . In contrast, the bias in CCR is between 0 and 0.01 except at the edges of the bandpass. CCR misses the center of the 9.6 μm feature but correctly estimates the emissivity along the wings of the feature. The error in the ISAC estimate is due to the atmospheric compensation. Figure 4.32 shows the ISAC surface-leaving radiance for this case. The shape is smooth and is basically blackbody. The reason why this happened is because the scene from which the ISAC transmission and upwelled radiance were derived is dominated by the composition of the playa (see Figure 3.17(c)). ISAC assumes that the majority of the pixels in the scene are blackbody. Therefore, the atmospheric spectra retrievals are contingent upon the assumption that the playa is a blackbody and contain some of the features associated with the true playa emissivity. Thus, the retrieved surface-leaving radiance resembles blackbody radiation. Figure 4.33 shows the retrieved transmission spectra for ISAC and CCR. The ISAC retrievals have a much more pronounced ozone absorption feature.

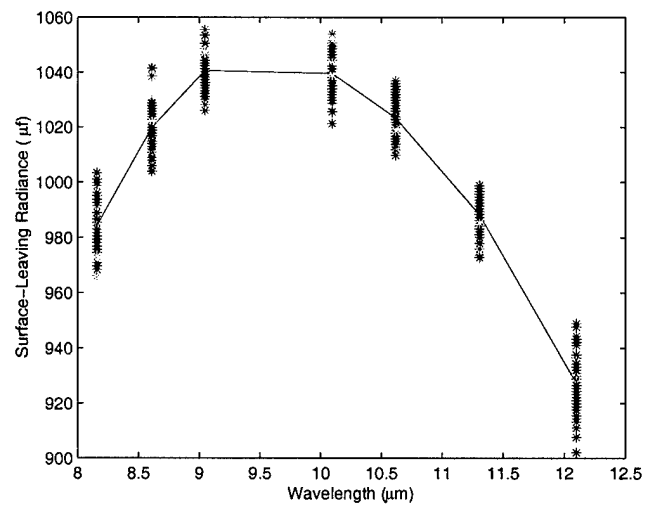


Figure 4.32: Surface-leaving radiance ISAC retrieval for Railroad Valley playa.

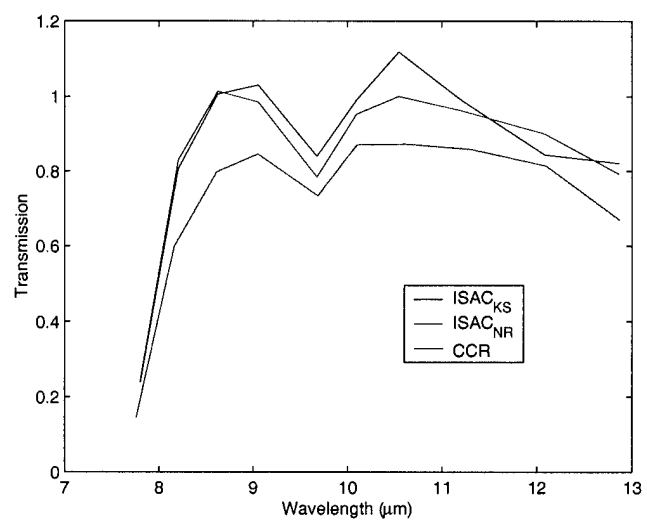


Figure 4.33: ISAC and CCR transmission spectra retrievals over Railroad Valley playa.

4.7 Appropriateness of Linear Model

The CCR inverse model is a linear transformation of the observed spectra to the space spanned by the parameters of interest. As such, nonlinear relationships between the data are not exploited. From the discussion in the Chapter 2, it is clear that the radiative transfer of radiation is nonlinear. The nonlinearities arise from the interdependence of the Planck radiation and the absorption, which is due to their joint dependence on temperature and wavelength. The linear inverse procedures described in Section 2.1.5 assumed that the Planck function dependence on wavelength and the absorption dependence on temperature were negligible. The CCR inverse model does not make these assumptions. For example, the linear transformation of the observed radiance to the vertical temperature profile does not make the use of a sounding weighting function. The transformation is made up of three steps: (1) the decomposition of the observed spectra into canonical variables via canonical weights which form an orthogonal basis, (2) the mapping between the observations and the parameter space via the canonical correlations, and (3) the reconstruction of the parameters based on the basis (canonical weights) derived from the parameter space. Thus, the linear mapping is based on the canonical variables; *not* on the original variables. While the canonical correlations will be weak if there is no linear relationship between the data, they maximize the amount of linear relationships that do exist. Therefore, if the linear relationships dominate, the first few canonical correlations will effectively “summarize” these relationships while nonlinear relationships are ignored (and contained in the weaker canonical correlations and corresponding variables). Also, the canonical relationships are based on a “steady-state” atmosphere where the true temperature of the atmosphere is known. In contrast, traditional atmospheric sounding algorithms use weighting functions derived from an initial estimate of the vertical temperature profile and adjust the weighting functions iteratively as the algorithm converges to a temperature solution. In this section, the CCR linear assumption is explored by analysis of the canonical variables and correlations and analysis of residuals.

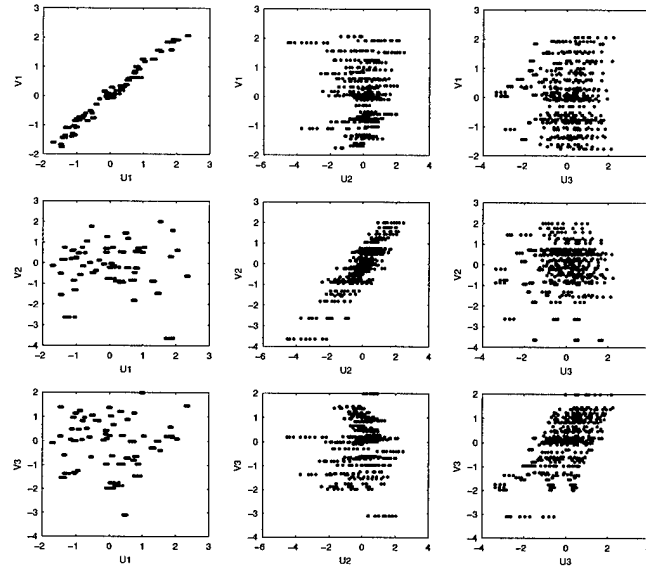


Figure 4.34: Matrix plot of canonical variables for MASTER simulations using Run 23072 of Experiment #4. The canonical correlations relate observed brightness temperatures to temperature and water vapor profiles.

4.7.1 Canonical Variables and Correlations

In this research, there were typically 3 to 5 significant canonical correlations. The first canonical correlation was always greater than 0.97. The correlations were slightly higher for the estimation of the atmospheric optical parameters (e.g., transmission) than for the physical parameters (e.g., water vapor profile). Figure 4.34 is a “matrix” plot for the first 3 canonical variables for MASTER simulations using Run 23072 of Experiment #4. Figure 4.35 is the matrix plot of the first 5 canonical variables for SEBASS simulations using Run 23072 of Experiment #4. The canonical variables shown in these plots resulted from the regression of the brightness temperatures to the vertical temperature and water vapor profiles. The diagonal plots are those corresponding to the canonical correlations and demonstrate the strong linear relationship in the data. The off-diagonal scatter plots show how the canonical variables are uncorrelated with variables corresponding to other dimensions. The plot of the canonical variables also demonstrates that there are no outliers

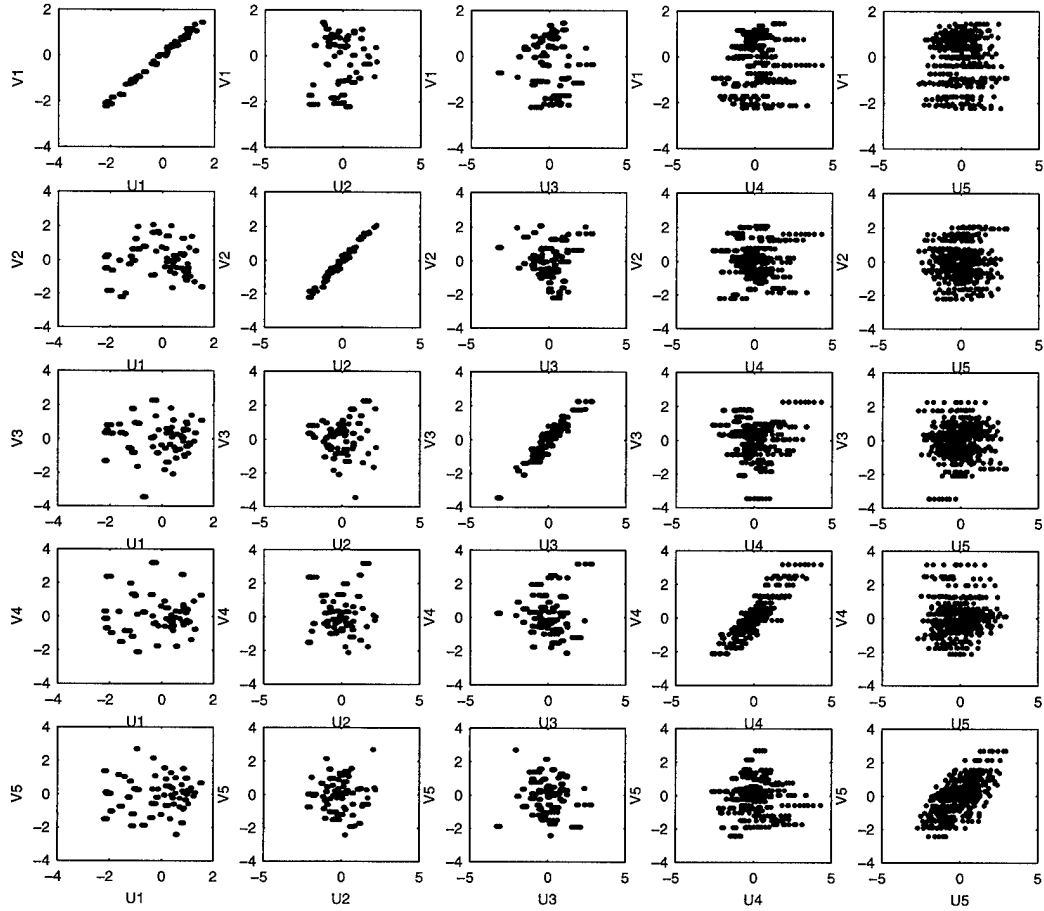


Figure 4.35: Matrix plot of canonical variables for SEBASS simulations using Run 23072 of Experiment #4. The canonical correlations relate observed brightness temperatures to temperature and water vapor profiles.

in the data. It is evident that the canonical correlations are stronger for SEBASS than for MASTER. This is due to the higher spectral resolution of SEBASS, which allows the separation of atmospheric and surface effects. This separation leads to improved correlations with the atmospheric parameters. Nevertheless, both cases result in at least one large canonical correlation, thus demonstrating the dominance of the linear relationships.

One of the appealing advantages of the CCR inverse model is that it is not based on any assumptions of the underlying probability distributions of the data. However, if the

distributions are Gaussian, the canonical correlations are Maximum Likelihood estimates and maximize mutual information (Kullback 1997; Akaho et al. 1999). In this ideal case, the linear model is optimal from an *information theory* perspective. The Central Limit Theorem states that the probability distribution of a random variable resulting from the sum of several independent and identically distributed random variables converges to a Gaussian distribution as the number of variables goes to infinity (Johnson and Wichern 1992). Thus, the maximization of mutual information is more likely with the CCR model because the probability distributions of interest are based on the canonical variables, which are sums of the original variables.

Figures 4.36(a) and 4.36(b) show the histograms for the first two canonical variables obtained with the NAST-I data at the maximum MODTRAN resolution (3,310 observations with 751 bands in the LWIR bandpass between 650 and 1400 cm^{-1}). The canonical variables were obtained from analysis relating the observed brightness temperatures to the temperature and water vapor profiles. The histograms are overlaid with a plot of the Gaussian probability distribution function. This qualitative look reveals that the first canonical variable does not seem to follow a Gaussian distribution while the second canonical variable closely matches the Gaussian distribution except at the extreme right tail. A more quantitative analysis involves the use of the Kolmogorov-Smirnov statistic (see Appendix E). A small D-statistic means that the cumulative distribution of the observations matches the normal cumulative distribution. The p-value is the probability that a distribution with the calculated D-statistic is a normal distribution. Therefore, the p-value ranges between 0 and 1.0 with 1.0 signifying 100% confidence that the observations are normally distributed. There were 8 significant canonical correlations for these data. These are listed in Table 4.15 with the associated D-statistics and p-values. The last three canonical variables are normally distributed. The second canonical variable is barely normally-distributed. Figures 4.36(a) and 4.36(b) show the normal probability plots for the first two canonical variables. These plots have the normal quantiles as the ordinate and the canonical variables as the abscissa.

If the observations follow a normal distribution, the scatter plot would fall along a perfect line. These results are consistent with the histogram plots and demonstrate that the first canonical variable is not normally distributed while the second one is. The normal probability plots also show that the deviation from normality in the second canonical variable occurs at the tails. The observations leading to these deviations were 323, 332, and 2310 in the NAST-I data. There were no obvious problems with these points.

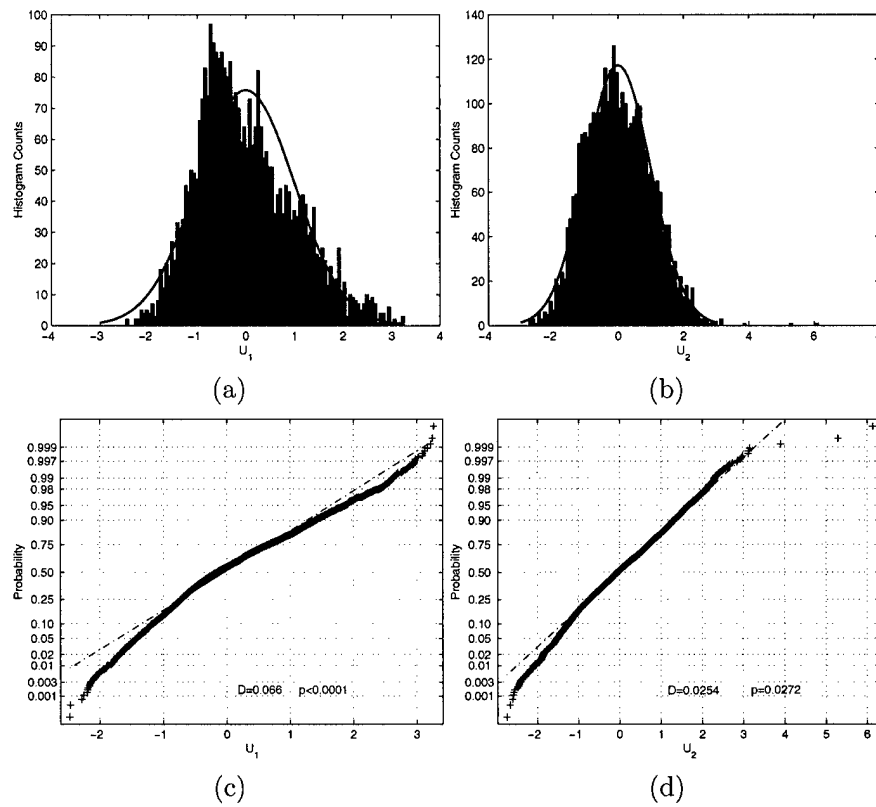


Figure 4.36: Histogram plots for (a) u_1 and (b) u_2 canonical variables with Gaussian pdf overlay. (c) and (d) are the normal probability plots for u_1 and u_2 , respectively.

Canonical Correlation	D-statistic	p-value
0.99991	0.06629	<0.00001
0.99725	0.02543	0.02717
0.99242	0.05679	<0.00001
0.9847	0.02940	0.00640
0.97367	0.03401	0.00091
0.93251	0.01544	0.40655
0.72471	0.01493	0.44898
0.655	0.01481	0.45958

Table 4.15: Canonical correlations and corresponding Kolmogorov-Smirnov statistics.

4.7.2 Analysis of Residuals

The residual plots can give an indication of the appropriateness of the linear model. For example, if there is a nonlinear relationship between the surface temperature and the observed brightness temperature, then the residual errors between the estimated and the true temperatures would have some curvature when plotted against the magnitude of the estimated temperature. Residual plots for surface temperature and column water vapor are shown in Figure 4.37. These residuals were obtained from the analysis of SEBASS simulations using Run 23072 from Experiment #4. The temperature errors appear to cluster at the center of the plot and increase at the edges of the surface temperature range. This increased error at the edges is due to the CCR model only predicting deviations from the mean surface temperature (see Section 4.7.3 for a more detailed explanation). The column water vapor residuals exhibit a series of straight lines with slopes equal to -1. This is an artifact of the experimental design. For Run 23072, each water vapor profile was included in the data set 9 times—9 repeats because there were three surface temperatures, each with three different emissivities. Draper and Smith (1998) provides an excellent discussion about this effect, which is present in all regression residuals regardless of what model and fitting technique is used (if no repeats are included in the data, then only one point of each line is observed). The effect is more obvious here because there are discrete levels of column water vapor that are sufficiently apart. The effect is also present in the surface temperature residuals

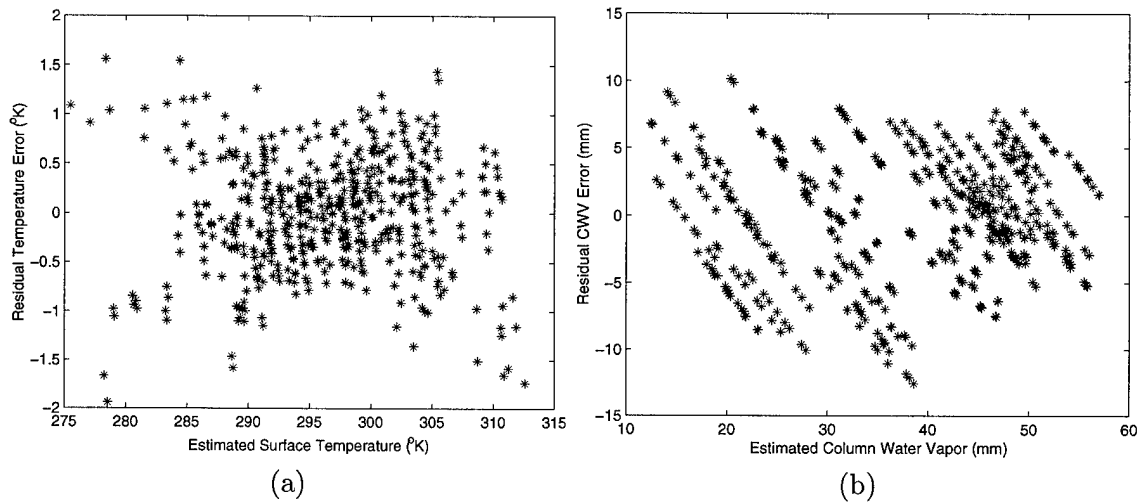


Figure 4.37: Residual plots for (a) surface temperature and (b) column water vapor with repeats.

but it is not as noticeable because the temperature values are much more closely spaced and there are only 3 repeats. In addition, the total range of temperatures is much larger than the residual magnitudes, which results in sloped lines that appear more vertical due to the aspect ratio of the plot. This effect may be compensated by only including one of the repeats in the residual analysis. Figure 4.38 shows the residual plots without repeats. The correlation of the residuals with the fitted surface temperature values is much more apparent. Again, this results from not estimating the mean surface temperature level with the model. The effect may be reduced if a smaller range of temperatures is considered. The column water vapor residual plot does not exhibit any significant patterns.

4.7.3 The CCA Regression Model

The regression model obtained with CCA is significantly different from the standard multivariate regression model. The difference stems from the symmetry of CCA. Recall that the

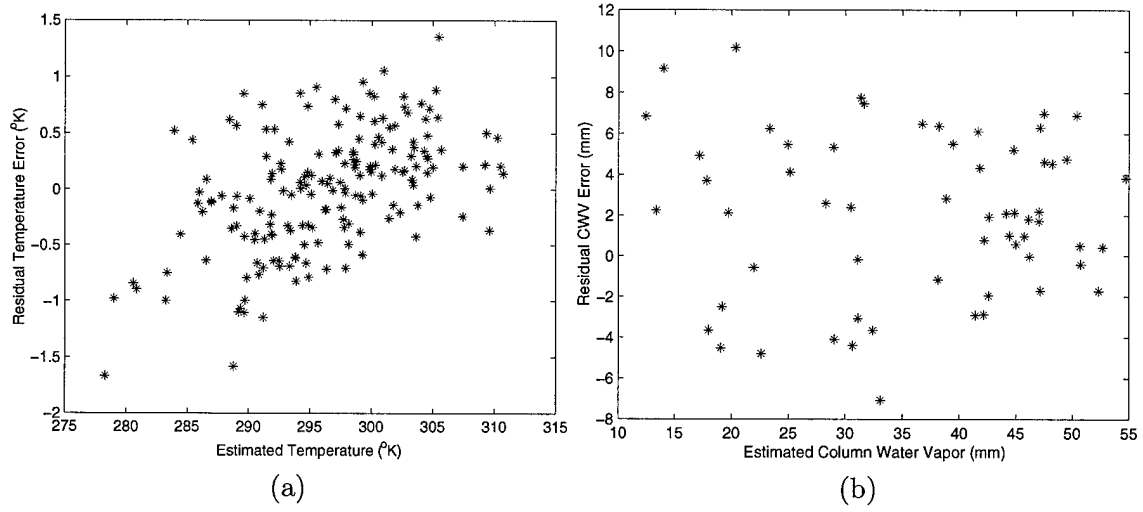


Figure 4.38: Residual plots for (a) surface temperature and (b) column water vapor without repeats.

canonical transformations are obtained from the solutions to

$$\Sigma_{xx}^{-1} \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx} \mathbf{a} = \psi_1^2 \mathbf{a}$$

$$\Sigma_{yy}^{-1} \Sigma_{yx} \Sigma_{xx}^{-1} \Sigma_{xy} \mathbf{b} = \psi_1^2 \mathbf{b}$$

The use of the covariance matrix notation partly obscures what CCR is really doing. Replacing the covariance matrices with mean-centered and scaled matrices $\mathbf{X}$ and $\mathbf{Y}$, the CCA equations become:

$$\underbrace{(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}}_{\beta_{\mathbf{X}\mathbf{Y}}} \underbrace{(\mathbf{Y}'\mathbf{Y})^{-1} \mathbf{Y}'\mathbf{X}}_{\beta_{\mathbf{Y}\mathbf{X}}} \mathbf{a} = \psi_1^2 \mathbf{a} \quad (4.5)$$

$$\underbrace{(\mathbf{Y}'\mathbf{Y})^{-1} \mathbf{Y}'\mathbf{X}}_{\beta_{\mathbf{Y}\mathbf{X}}} \underbrace{(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}}_{\beta_{\mathbf{X}\mathbf{Y}}} \mathbf{b} = \psi_1^2 \mathbf{b} \quad (4.6)$$

This clearly shows that the canonical weights are the eigenvectors which simultaneously maximize the least-squares solutions for the regression of $\mathbf{X}$ on $\mathbf{Y}$ and the regression of $\mathbf{Y}$ on $\mathbf{X}$. In a way, the CCR solutions are analogous to the geometric mean functional relationship regression coefficients. The geometric mean regression is appropriate for cases

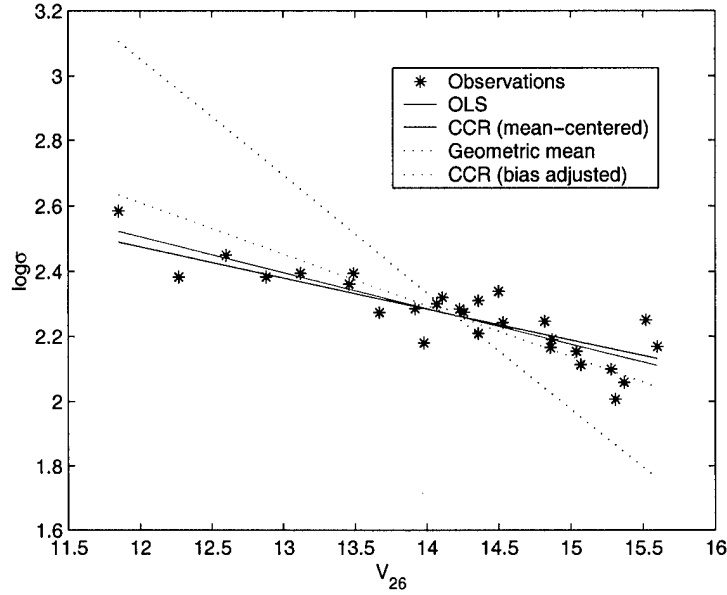


Figure 4.39: Example regression data from Draper

where there is noise in both $\mathbf{X}$ and $\mathbf{Y}$ and when the number of observations in the regression data is small compared to the number of parameters in the model (Draper and Smith 1998).

In the context of symmetry, the concept of a “y-intercept” has no meaning. The symmetry of the CCA regression model forces the intercept to the origin of the data coordinates. Therefore, the model is only appropriate when the data are mean-centered or when the true mean is zero. For zero-mean data, the CCR solutions are identical to OLS solutions if all of the canonical correlations are kept in the model. The same is not true for mean-centered data. That is, the solutions obtained from OLS after adding the mean back to the data are different than those obtained with CCR. Again, this is due to the symmetry of the CCR model which is *retained* even after the mean has been added back to the data.

Figure 4.39 shows an example of the differences between the OLS, CCR, and geometric mean solutions for the star data example in Table 2.2 of Draper and Smith (1998). The data consist of magnitude observations (V_{26}) and the log central velocity dispersion ($\log \sigma$)

for stars in the Coma cluster. The red line is the OLS solution. The blue line is the CCR mean-centered solution after the mean is added back to the data. The dotted blue line is the geometric mean functional relationship solution. Finally, the red dotted line is the solution obtained with CCR if the data are not mean-centered but are bias adjusted by the geometric means. That is, the slope coefficient is obtained from CCR and the “y-intercept” is obtained from $b_0 = \bar{y} - \bar{x}b_1$ where b_1 is the CCR regression coefficient and b_0 is the “y-intercept”. The CCR and geometric mean solutions bias the regression line such that the slope of the line is steeper than the OLS solution. Thus, the linear relationship predicted by CCR is *stronger* than the OLS prediction. This may be more representative of the *true* relationship of the parameters when there is uncertainty in both $\mathbf{X}$ and $\mathbf{Y}$. That is, we think that the OLS solution underestimates the true relationship due to the noise in the observations. The bias is much stronger for the geometric mean solution than for CCR.

Chapter 5

Conclusions and Recommendations

The outcome of any serious research can only be to make two questions grow where only one grew before.

Thorstein Veblen

Every scientific fulfillment raises new questions; it asks to be surpassed and outdated.

Max Weber

5.1 Summary

Estimating the surface temperature and emissivity of land targets is a difficult task because there are so many variables that contribute to a single sensor observation. The atmosphere and the surface are strongly coupled, particularly for targets with emissivities less than 0.9. Thus, a suitable atmospheric compensation must consider the emissivity effects so that the atmospheric parameters may be retrieved accurately.

My approach was to build an *inverse* model by using Canonical Correlation Analysis (CCA) as a rank-reduced multivariate regression to capture all the relevant physics contained in the MODTRAN *forward* model. This approach was tested by using three separate

atmospheric databases with distinct climates. The results showed that the Canonical Correlation Regression (CCR) models did not depend highly on the geographical location of the atmospheric database. This is an indication of the robustness of the model. That is, the model is not just optimized for whatever database of observations was used to create it; it is able to handle future observations as well. This was corroborated through cross-validation of the model and by the physical interpretation of the canonical weights and loadings. However, the models *do* depend on the amount of variability in the data. In general, parameters estimated with the SSEC (global) database had larger errors than those obtained with the other two databases. This is primarily due to the large amount of variability relative to the average atmospheric state in these data. In addition, discretization of the atmospheric profiles minimizes the strength of the canonical correlations. Thus, I recommend that separate CCR models be built for each distinct climate of interest with high vertical-resolution profiles.

When the quality of the atmospheric database is high, the CCR inverse model can yield very accurate results. It was demonstrated that it is feasible to estimate water and land surface temperature and emissivity to within 1.0 °K and 0.01 accuracies, respectively. It is also possible to retrieve atmospheric parameters with high accuracy depending on the available spectral resolution of the sensor. Multispectral thermal sensors may retrieve atmospheric temperatures to within 2 to 3 °K error RMS and column water vapor to within 20%. Hyperspectral and ultraspectral sensors may achieve errors on the order of 1 °K and 10% for temperature profiles and column water vapor, respectively. Thus, the CCR inverse model is a good alternative to standard statistical methods used in traditional atmospheric sounding applications.

The CCR inverse model is versatile. It can be used for the direct estimation of surface temperature, temperature and water vapor vertical profiles, atmospheric transmission, upwelled radiance, and downwelled radiance. The canonical correlations relating the observed radiance to the atmospheric spectra were strong enough to build an accurate inverse

model. For the retrieval of temperature and water vapor, the sensor brightness temperatures yielded stronger correlations than the observed spectral radiance. This is because the relationship between temperature and radiance is nonlinear. Including both radiance and brightness temperature measurements as predictors in the model would not add more information. However, it may lead to a slightly better fit if scatter plots of the canonical variables reveal any nonlinearity.

The CCR inverse models do not appear to amplify sensor noise. Even though the sensor noise was not considered, the CCR inverse models adequately predicted the surface temperature and emissivity from MASTER observations. This was true for direct temperature retrievals and for retrievals obtained through atmospheric compensation and TES. These results, however, are based on water and playa targets. The effect of sensor noise on retrievals over less emissive targets should be studied. If the sensor noise is correlated, then the observations should be preprocessed to compensate for the structure in the noise. Alternatively, the structure may be accounted in the covariance matrices used to compute the canonical correlations and weights.

CCR can be used as a sensor design tool. The canonical weights and loadings emphasize the original variables that *lead* to the highest correlations. The canonical correlations are a measure of mutual information assuming that the variables are normally distributed. Therefore, CCA may be optimal from an information theory perspective. It was shown that the analysis of the canonical weights and loadings can identify regions of the spectrum that carry the most information about the vertical structure of the atmosphere. The number of significant canonical correlations identify the smallest number of bands that can be used to retrieve parameters within a specified accuracy. The canonical loadings resemble traditional sounding weighting functions, confirming that the CCR model is physics-based. Thus, CCR is an appealing alternative to traditional sounding methods when system resources are limited and an extremely high spectral resolution sensor is not available. If the option for a high resolution *is* available, the CCR model can readily handle the large number of

bands and efficiently find the regions where narrow sensor bands and sounding weighting functions should be built.

The CCR inverse model compares well with other multivariate regression models. In general, it performs as well or better than Maximum Redundancy (MR), Partial Least Squares (PLS), and Principal Components Regression (PCR). This is particularly so for retrievals of physical parameters (i.e., surface temperature and temperature and water vapor profiles). If the number of dimensions retained in the model is low, all of the methods estimate atmospheric optical parameters with the same level of accuracy. PLS outperformed CCR, MR, and PCR when the number of dimensions was increased. Generally, the robustness of the model is inversely proportional to its complexity. However, it is possible that complex PLS models are robust and suitable for remote sensing applications. Chemometricians have found PLS consistently accurate in the estimation of material abundances from high-resolution spectroscopic data. Thus, the PLS model may be appropriate for spectral unmixing applications. More analysis with complex PLS models is recommended.

The CCR inverse model also compares well with the ISAC algorithm. Several implementations of the ISAC algorithm were developed and tested. The advantages of ISAC are that it is simple and it does not require the use of a large database. For scenes that are dominated by blackbody targets, the ISAC solutions coupled with TES may yield better surface temperature retrievals than CCR. However, if the scene is not dominated by blackbody targets and TES does not correctly compensate for the bias in the unscaled parameter solutions, then the CCR solutions will generally be more accurate.

The accuracy of the temperature retrievals was largely limited by the TES algorithm. Materials that have characteristics that do not fit the MMD regression line are biased depending on the calculated ε_{min} . Alternative emissivity-scaling methods should be explored. Despite this drawback, the algorithm performed relatively well. In particular, it was able to adequately compensate for downwelled radiance. This resulted in spectral emissivity estimates that had a shape nearly identical to that of the true emissivity. Thus, these

biased estimates could be very useful in spectral classification and target identification applications. However, large errors may be introduced when the emissivity spectral features match that of the atmosphere. Using pixel-averaged atmospheric parameter solutions may solve this problem if the spatial scale of the surface target class is smaller than the spatial atmospheric variability.

Increasing the spectral resolution from MASTER to SEBASS did not have a significant effect on the accuracy of the surface temperature retrievals when CCR was coupled with TES. The performance did not improve because the TES bias dominated the error in surface temperature estimates. However, the direct temperature retrievals were dramatically improved from those obtained with MASTER resolution. This indicates that the CCR model is able to distinguish the atmospheric features from the surface emissivity features and obtain a surface temperature estimate that is relatively unbiased by the spectral emissivity. This is feasible with higher spectral resolution because spectral atmospheric features tend to be narrower than the emissivity features. Thus, a CCR inverse model built for a high resolution sensor (e.g., ultraspectral) may be insensitive to emissivity effects. Unfortunately, the direct estimation of temperature does not yield atmospheric optical parameters. Without these, it is not possible to compensate the observations for atmospheric effects and retrieve the surface emissivity. One approach may be to estimate atmospheric profiles along with the surface temperature and use these as inputs into MODTRAN. The resulting transmission, upwelled radiance, and downwelled radiance can then be used to estimate the emissivity from the observed radiance subject to the constraint that the CCR surface temperature estimate is correct. Another possibility is to estimate the emissivities as was done here, but then use the direct temperature estimates to rescale the biased emissivities. Finally, given the CCR direct surface temperature estimate, it may be possible to indirectly use the alpha-residuals technique for the emissivity retrieval.

One practical difficulty was the implementation of the atmospheric databases. Because the databases were generated based on a factorial design, the number of observations in-

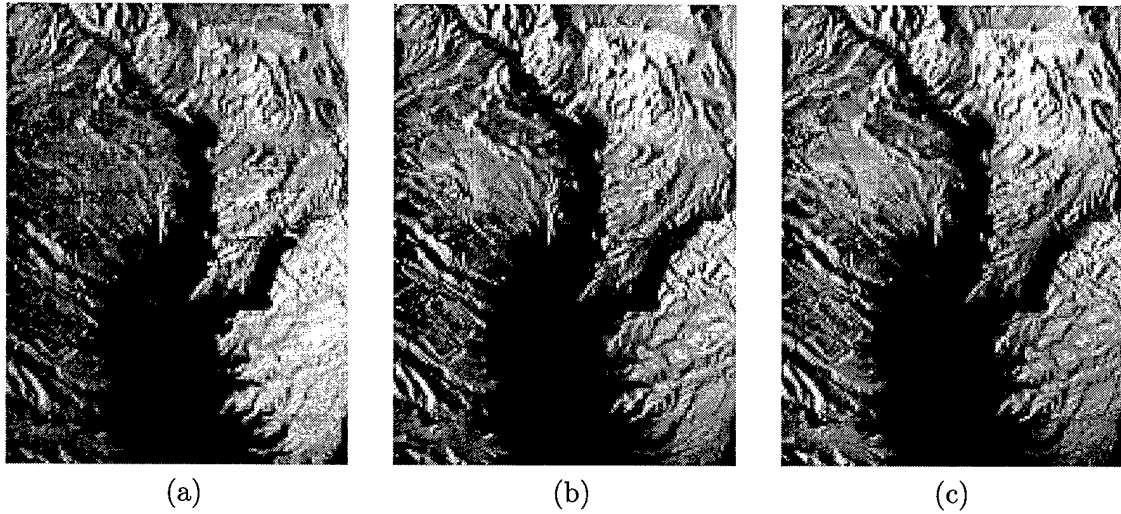


Figure 5.1: Lake Mead images obtained from MASTER: (a) CCR surface temperature map retrieval; (b) Band 46 sensor brightness temperature image; (c) CCR surface temperature map retrieval without mean-centering. The brightness scales of the images were modified through histogram equalization for better viewing.

creases exponentially as the number of factors (e.g., surface temperature and emissivity) increases. This exponential increase depends on the number of levels that each factor is allowed to vary. This may lead to a prohibitive computational burden. Therefore, a judicious choice of factors and levels is essential. Although building the database is computationally intensive, the database has to be generated only once. After that, a simple matrix of coefficients is all that is needed to process new data. This is a big advantage considering the large size of hyperspectral data because the operation is fast and may be implemented on hardware onboard the sensor platform. As an example, Figure 5.1(a) is a temperature map of Lake Mead generated by applying a simple linear transform (the CCR regression coefficients) to a $400 \times 300 \times 10$ (number of lines by number of samples by number of bands) MASTER multispectral image cube. The operation was done with MATLAB and lasted 0.35 seconds (including mean-centering operations) on an AMD Athlon 600 MHz PC with 128 MB SDRAM running Windows 2000 Professional. Figure 5.1(b) is the sensor brightness temperature for MASTER band 46. The CCR surface temperature map is considerably

“noisier” and has less contrast. The same thing was observed for the other images tested in this research. The exact reason why this happens is not known, but it seems to be related to the symmetry of the CCR model (Section 4.7.3). That is, the CCR solutions are based on mean-centered data because the “y-intercept” cannot be included in the model. The data in the regression analysis are centered about a *scalar* value. It may be that the true mean of the image lies in a plane with dimensions dictated by the context of the scene. If the CCR inverse model is built without mean-centering, it may be possible to recover this mean. Figure 5.1(c) is the temperature map estimated with CCR without mean-centering. The visual quality of the image is much better than the mean-centered solution. Unfortunately, the RMS errors in surface temperatures are much larger (by 2.0 °K or more) than the mean-centered solutions. The disparity between the visual and RMS results may be due to the psychophysics associated with the human visual system. This phenomenon is also seen in error diffusion theory where halftone images with lower RMS errors do not necessarily reproduce a better quality image. The image histograms are shown in Figure 5.2. These plots quantify the visual effects seen in the images. The evidence suggests that direct inversions to surface temperature with CCR should be done *without* mean-centering for *qualitative* or *relative* image analysis only. Otherwise, mean-centering should be used. More research on this issue is recommended.

5.2 Extensions of the CCA Approach

The CCA approach described in this research may be enhanced to potentially obtain more accurate estimates of atmospheric and surface parameters. This section describes some ideas on how this might be done. In addition, other uses of CCA for remote sensing applications are discussed.

One potential extension is to use CCA to retrieve the surface spectral emissivity directly from the observed radiance. The results of this research show that CCA is able to find spectral features in the observed radiance that are highly correlated with atmospheric

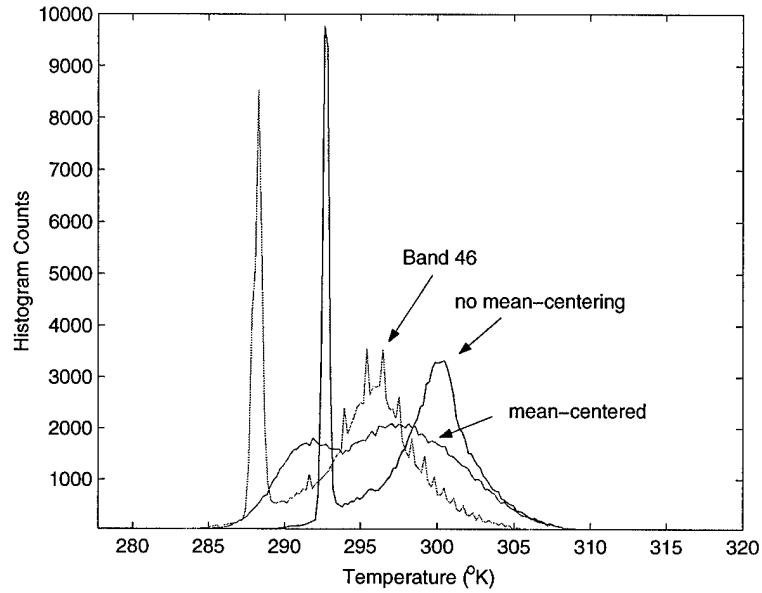


Figure 5.2: Histogram of Lake Mead temperature maps

parameters. CCA was forced to find these features despite biases from surface emission and reflection. Estimates of the atmospheric parameters were obtained with the *highest* canonical correlations. However the complete set of canonical correlations span all the data. Thus, the data are projected to dimensions that range between highest to lowest correlation. It may be possible to establish a separability of atmospheric and surface effects by separating the dimensions that are associated with the atmosphere and those that are *least* associated. That is, if there are $k = \min(p, q)$ canonical correlations, then the first r dimensions are highly correlated with the atmosphere and the last $k - r$ dimensions have almost no correlation with the atmosphere. The canonical variables $\mathbf{U}_{n \times (k-r)}$ associated with these dimensions could then be used as efficient predictors of surface emission and reflectance parameters. Because the CCR inverse model is generated with modelled data, the modelled observed radiance has no noise. Therefore, all of the variability in the set $\mathbf{U}_{n \times (k-r)}$ is due to surface effects. However, to relate $\mathbf{U}_{n \times (k-r)}$ to the surface emission, the ensemble must include surface emission variability. It might be tempting to consider including the spectral

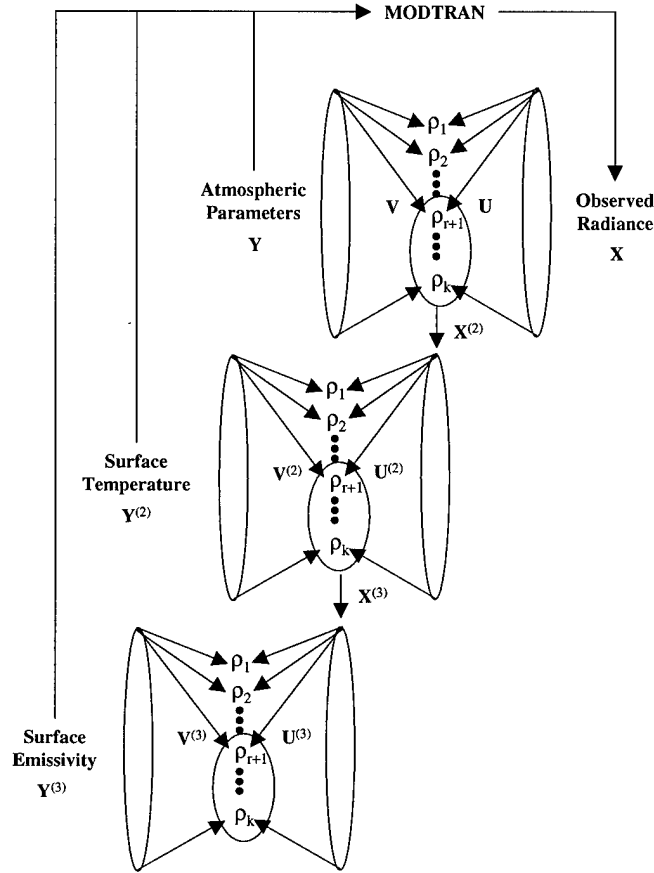


Figure 5.3: Diagram of the cascaded CCA approach

emissivity as another set of variables in the analysis. Unfortunately, this is impractical because the factorial design would result in a prohibitive number of MODTRAN runs. It may be possible to circumvent this problem by implementing a set of *cascaded* CCA's as shown in Figure 5.3. First, an ensemble X is built using a database of atmospheres Y with no variability in the surface emission and reflectance. The resulting canonical correlations can then be segmented and the $k - r$ *smallest* correlations and associated weights retained. A second ensemble $Y^{(2)}$ with variability *only* in surface temperature can then be used to generate new observed radiances, which are in turn projected onto the $k - r$ canonical weights derived from the previous analysis. It is only necessary to use any one

of the atmospheres used to generate $\mathbf{X}$ to make the new observed radiances comparable to the original observations and related canonical weights. The choice of the atmosphere is not crucial because the subsequent operations are atmosphere-independent by design. The resulting variables $\mathbf{X}^{(2)}$ are then used in a new CCA with $\mathbf{Y}^{(2)}$. As before, the $k - r$ smallest correlations and associated weights are retained. The $\mathbf{U}_{n \times (k-r)}^{(2)}$ are the least correlated with the surface temperature. Finally, a third ensemble of observed radiances is generated with a set of spectral emissivities $\mathbf{Y}^{(3)}$. These can be synthetic or from a spectral library of measurements (e.g., ASTER library). The same atmosphere used to generate $\mathbf{X}^{(2)}$ could be used to generate the new observed radiances. Using the weights from the previous step, the variables $\mathbf{X}^{(3)}$ are constructed. These are then used in a third CCA, this time relating the variables to the spectral library $\mathbf{Y}^{(3)}$. Once the analysis is complete, a new observation can be processed through the set of cascaded CCA rotations generating different parameter-estimates at each level. One potential problem with this approach is that the canonical weights are derived based on the geometric properties of the data. Therefore, linear combinations that result in separability with one set of variables may not do the same with a different set of data with new geometric properties. This problem may be addressed by adding the ensemble from the previous CCA to the new observations.

Another extension of CCA is to maximize the *statistical dependence* between the canonical variables $\mathbf{u}_i$ and $\mathbf{v}_j$, subject to the constraint that $\mathbf{u}_i$ and $\mathbf{v}_j$ are *statistically independent* for all $i \neq j$. In this research, I have stated that the orthogonality of the canonical variables implies independence. This is only true when the joint probability distributions are Gaussian, which was implicitly assumed. If the joint probability distribution is not perfectly symmetric (e.g., Gaussian distribution), it is possible to have orthogonal variables that are statistically dependent. For two variables to be statistically independent, their joint probability distribution must be able to be represented as a product of the individual probability distributions (i.e., $P(x, y) = P(x)P(y)$). Statistical independence implies orthogonality and imposes a stronger restriction than orthogonality. In fact, if two vari-

ables are independent, *any* transformation of the variables results in variables that are also independent. An alternative definition of independence is that all of the moments of the individual probability distributions are orthogonal. An extension of PCA that looks for independent components is Independent Component Analysis (ICA) (Hyvärinen 1999). In remote sensing, ICA has been applied to spectral unmixing problems where the endmembers and abundances are unknown (Bayliss et al. 1997; Tu 2000). The ICA approach can be extended to CCA except that the independent components are found so that the mutual dependence is maximized, which implies a maximization of mutual information. Recently, this approach was implemented as a neural network (Akaho et al. 1999). To my knowledge, this is the only work that has been done with this approach so this is a ripe area of research. For that matter, extensions of all of the multivariate regression models used in this research based on independent component analysis should be considered. Results from Section 4.7.1 showed that the canonical variables did not always have a Gaussian distribution. Thus, the independent CCA approach may result in improved estimates.

In this research, the focus was placed on the information content in the observed radiance spectra. However, there is also *contextual* information in images. As the spatial resolution of hyperspectral sensors improve, the use of contextual information in addition to spectral information may lead to dramatic improvements in the accuracy of retrieval algorithms. CCA lends itself very easily to the inclusion of contextual information. One approach may be to create parameter-based texture models (e.g., Markov Random Fields) and include the texture parameters as another set of variables in CCA. Another approach is to create a database of textures which form a basis for scenes of interest. The textures represent surface emission components, which are processed through MODTRAN to generate ensembles of observed radiances. CCR inverse models built with these textures would include context in the definition of the canonical correlations. Because MODTRAN does not have any scene generation capability, a different model would have to be employed. The Digital Imaging and Remote Sensing Image Generation (DIRSIG) model merges ray-tracing

calculations with CAD models and MODTRAN to generate scene simulations (Schott 1997). Thus, DIRSIG may be an appropriate tool for this CCA extension.

Extensions of the CCA approach may be associated with other remote sensing applications. For example, CCA may be useful as a change detection algorithm. Nielsen (1995) developed a Multivariate Alteration Detection (MAD) algorithm for change detection based on CCA. The concept is similar to the cascaded CCA procedure described earlier. CCA is used to relate two (or more) multispectral images taken at different times. Areas leading to high canonical correlations are associated with no change while those leading to the smallest canonical correlations are associated with change. Because CCA does not restrict the $\mathbf{X}$ and $\mathbf{Y}$ to have the same number of variables, images acquired with different sensors may be used. Another example is data compression. The set of significant canonical correlations are typically much lower than the number of dimensions of hyperspectral data. If the purpose of the data is to extract a particular parameter, only the canonical variables need to be transmitted. The receiver station would then use stored canonical correlations and loadings to reconstruct the parameters of interest.

5.3 Outlook on the future

These are exciting times for remote sensing and Earth science. With the launch of Terra on December, 1999, the NASA Earth Observing System (EOS) is well on its way to achieving many objectives supporting fundamental research areas in Earth science. The MODIS sensor onboard Terra has already acquired spectacular imagery of the Earth (<http://earthobservatory.nasa.gov>) and will continue to do so over 36 spectral bands covering the entire globe every 1-to-2 days (King and Greenstone 1999). ASTER, also onboard Terra, has also begun to acquire multispectral thermal images at 90 m spatial resolution. The information provided in the imagery from these sensors alone will have a tremendous impact on our understanding of the environment.

The next generation of sensors will continue to push technology toward higher spectral and spatial resolution. As the volume of data increases, the timeliness of specialized algorithms cannot be overstated. It is in the context of the evolution of these algorithms that the results, insights, and conclusions obtained from this research should have the largest impact.

Appendix A

Water Vapor Units and Conversions

Beware of the man who won't be bothered with details.

William Feather, Sr.

There are several measures of water vapor that can be used to characterize its content in the atmosphere. In this appendix, relevant units used in this research are discussed. A more thorough discussion on atmospheric chemistry may be found in (Seinfeld and Pandis 1998) and (Saucier 1989).

One way of expressing concentration is in terms of moles per volume of air. The concentration describes how many molecules are in the volume based on Avogadro's number (i.e., $N_A = 6.022 \times 10^{23} \text{ mol}^{-1}$). The ideal gas law states

$$PV = NRT \quad (\text{A.1})$$

where P is the pressure in Pascal (Pa^1), V is the volume (m^3), N is the number of moles, R is the molar gas constant ($8.314 \text{ Nm mol}^{-1} \text{ K}^{-1}$), and T is the temperature (K). The

¹1 hPa = 10000 Pa = 1 millibar (mb). Also, a standard atmosphere (atm) is defined as 1013.25 mb.

concentration depends on pressure and temperature. Because these physical parameters are dynamic, a better measure of concentration for atmospheric studies is the *mixing ratio*:

$$\xi_i = \frac{c_i}{c} \quad (\text{A.2})$$

where ξ_i is the mixing ratio of the i^{th} constituent, c_i is the concentration of the i^{th} constituent, and c is the total concentration of all the constituents. This is a measure that is independent of pressure and temperature. The measure can also be expressed in terms the ratio of *partial* pressure:

$$\xi_i = \frac{p_i}{p} \quad (\text{A.3})$$

where p_i is the partial pressure of the i^{th} constituent and p is the total (ambient) pressure.

The *mass* mixing ratio of water is the ratio of the partial pressures weighted by the ratio of molecular weights:

$$r = \frac{m_w e_w}{m_d p_d} \quad (\text{A.4})$$

where m_w is the molecular weight of water (18.015 g/mol), m_d is the molecular mass of dry air (28.966 g/mol), and p_d is the partial pressure of dry air. The mixing ratio is usually expressed in units of g/kg and can referenced to either dry or wet air. When compared to wet air, the mixing ratio of water is also called the *specific humidity*. In this research, all mixing ratios are expressed relative to dry air. The mixing ratio can be expressed in terms of water vapor and total pressure:

$$r = 1000 \frac{m_w}{m_d} \frac{e_w}{p - e_w^{(s)}} = 621.97 \frac{e_w}{p - e_w^{(s)}} \quad (\text{A.5})$$

where the factor of 1000 is introduced to make the ratio units of g/kg. $e_w^{(s)}$ is the *saturation* water vapor pressure. The saturation water vapor pressure is the maximum pressure that can be exerted by water at a given ambient temperature. Thus, the partial pressure of dry air is the total pressure less the *potential* partial pressure of water. The saturation mixing ratio is

$$r_s = 621.97 \frac{e_w^{(s)}}{p - e_w^{(s)}} \quad (\text{A.6})$$

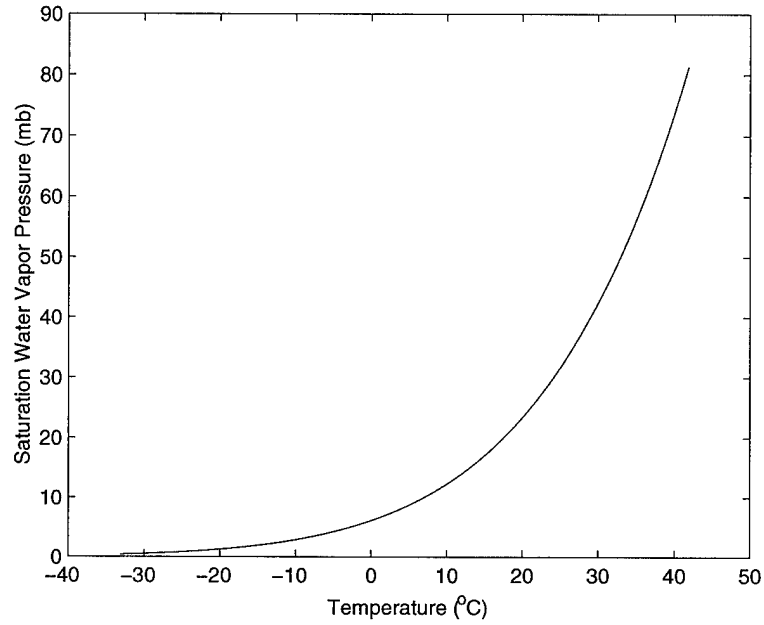


Figure A.1: Water vapor saturation pressure

The saturation water vapor pressure is a function of temperature *only* (see Figure A.1). The functional relationship is the Goff-Gratch formula with reference over liquid water, which has been standardized by the International Meteorological Organization (List 1951). The formula is $e_w^{(s)}(T) = 10^{\alpha(T)}$, where

$$\begin{aligned} \alpha(T) = & -C_1 \frac{T_s}{T-1} + C_2 \log\left(\frac{T_s}{T}\right) - C_3 \left[10^{C_4 \left(1 - \frac{T}{T_s}\right)} - 1 \right] \\ & - C_5 \left[1 - 10^{-C_6 \frac{T_s}{T-1}} \right] + \log(1013.25) \end{aligned} \quad (\text{A.7})$$

where T is the ambient temperature (K), T_s is the temperature of boiling water, and 1023.25 mb is the water vapor saturation pressure at T_s . The coefficients C_1 through C_6 are the elements of the array:

$$\mathbf{C} = [7.90298, 5.02808, 1.3816 \times 10^{-7}, 11.344, 8.1328 \times 10^{-3}, 3.49149]$$

The relative humidity (RH) is the percentage of the partial pressure of water referenced to the water vapor saturation pressure. Thus,

$$\text{RH} = 100 \frac{e_w}{e_w^{(s)}} \quad (\text{A.8})$$

where RH is in %. Alternatively,

$$\text{RH} = 100 \frac{r}{r_s} \quad (\text{A.9})$$

Thus, the RH is an indication of how much water vapor is in the air relative to how much there can be without condensation.

The relative humidity is a simple parameter to understand, but it can be intractable as a measure of water vapor because RH is measured with respect to water vapor saturation, which is highly dependent on temperature. Thus, a small decrease in ambient temperature can result in a high increase in relative humidity if the water vapor content remains constant. A better measure is the *dew point* temperature, which is the temperature to which the atmosphere would have to be cooled for saturation to occur. This is the standard measurement of water vapor in radiosonde data. The water vapor pressure can be estimated with Tetens' formula:

$$e_w = 6.11 \times 10^{at/(t+b)} \quad (\text{A.10})$$

where $a = 7.5$ and $b = 237.3$ °K over water (i.e., liquid-phase temperatures). The coefficients are $a = 9.5$ and $b = 265.5$ °K over ice.

All of the measurements described so far can be used to describe the amount of water vapor in a layer of the atmosphere. The vertical profiles of water vapor used in this research are presented in units of mixing ratio. There are times, however, when it is useful to quantify the *total* amount of water vapor in a column of air. The total column water vapor (CWV) can be directly calculated from the water vapor profiles. The CWV is equivalent to precipitable water, which is measured in millimeters. Suppose all of the water vapor in the column is condensed at the surface, then the precipitable water measures the height of the condensed water in a hypothetical 1 m² surface area. The total CWV is a function of

the amount of water and atmospheric pressure:

$$\text{CWV} = \frac{1}{g \rho_{H_2O}} \int_0^p \gamma(p) dp \quad (\text{A.11})$$

where ρ_{H_2O} is the water density at standard pressure (1000 kg/m³), g is the gravitational constant (9.807 m/s²), p is the ambient pressure (i.e., the "pressure altitude" of the atmospheric layer with thickness dp)(Pa), and $\gamma(p)$ is the water mixing ratio (g/g) expressed as function of pressure altitude.

MODTRAN accepts user-defined atmospheres (i.e., vertical profiles of concentrations and temperature). Several units may be specified for water vapor, including mixing ratio, relative humidity, and dew point. However, the output files are less flexible. Older versions of MODTRAN generated vertical water vapor profiles in units of g/cm². This is a density-dependent measure of concentration and is related to the mixing ratio by

$$c_o = \frac{1}{10} \frac{\gamma}{g} dp \quad (\text{A.12})$$

where c_o is the concentration in g/cm², γ is the mixing ratio (g/g), g is the gravitational constant, and dp is the pressure of the atmospheric layer. Because of the density dependence, these values were referenced to sea-level density. Newer versions of MODTRAN generate vertical water vapor profiles in atm-cm. This is a pressure and temperature independent measure because it is referenced to STP (i.e., $T = 273.15$ °K and $P = 1$ atm). The molar volume of *any* gas at STP is 22,413.83 cm³. Thus, the pressure-volume is 22,413.83 atm-cm³. Since the molecular weight of water is 18.015 g/mol, the conversion from c_o to the new MODTRAN units is

$$c_n = \frac{22413.83}{18.015} c_o = 124.42 \frac{\gamma}{g} dp \quad (\text{A.13})$$

Appendix B

Singular Value Decomposition

It is often useful to decompose a matrix into a set of basis vectors. The basis are like “building blocks” that can be used to reconstruct the matrix when the appropriate weighting of the basis is applied. The decomposition of a matrix is often called a *factorization*. Ideally, the matrix is decomposed into a set of factors (often orthogonal or independent) that are optimal based on some criterion. For example, a criterion might be the reconstruction of the decomposed matrix. The decomposition of a matrix is also useful when the matrix is not of full rank. In these cases, the rows or columns of the matrix are linearly dependent and do not form an orthogonal basis for the matrix. In theory, a rank-deficient matrix may be decomposed into a smaller number of factors than the original matrix and still preserve all of the information in the matrix.

If $\mathbf{A}$ is a square symmetric matrix, then a useful decomposition is based on its eigenvalues and eigenvectors. That is,

$$\mathbf{A}\mathbf{E} = \mathbf{E}\mathbf{\Lambda} \tag{B.1}$$

where $\mathbf{E}$ is the matrix of eigenvectors and $\mathbf{\Lambda}$ is the diagonal matrix of eigenvalues. The eigenvectors have the convenient mathematical property of orthogonality (i.e., $\mathbf{E}'\mathbf{E} = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix) and span the entire space of $\mathbf{A}$. The eigenvalues are the largest values of $\mathbf{A}$ and form a *spectrum* of orthogonal values. For that reason, this procedure is often referred to as a *spectral decomposition* (Johnson and Wichern 1992).

The limitation of this approach is that $\mathbf{A}$ must be square and symmetric. In many cases, we wish to decompose a matrix that is not square nor symmetric. Consider the matrix $\mathbf{A}$ that is $n \times p$ where $n \neq p$. In this case, the rank of the matrix is $r \leq \min(n, p)$ and the matrix is defined by row and column spaces that have different rank. However, $\mathbf{A}'\mathbf{A}$ is square and symmetric (it is also positive semi-definite). The same is true for $\mathbf{A}\mathbf{A}'$. The former is a *inner* product of the matrix and results in a matrix that is spanned by the column space (i.e., the *range*) of $\mathbf{A}$. The latter is a *outer* product of the matrix and results in a matrix that is spanned by the row space of $\mathbf{A}$. Refer to Trefethen and III (1997) for more on principles of matrix algebra.

The nonzero eigenvalues of $\mathbf{A}'\mathbf{A}$ and $\mathbf{A}\mathbf{A}'$ are called *singular* values. However, the corresponding eigenvectors are different. The eigenvectors of $\mathbf{A}\mathbf{A}'$ are called the “left” singular vectors while the eigenvectors of $\mathbf{A}'\mathbf{A}$ are the “right” singular vectors. By retaining the nonzero eigenvalues $k = \min(n, p)$, a singular value decomposition (SVD) can be constructed:

$$\mathbf{U}\mathbf{S}\mathbf{V}' = \mathbf{\Sigma} \quad (\text{B.2})$$

where $\mathbf{U}$ is a $n \times k$ matrix of left singular vectors, $\mathbf{V}$ is a $k \times p$ matrix of right singular vectors, and $\mathbf{S}$ is a diagonal matrix of singular values. The singular values are the *squared* singular values of $\mathbf{A}$, making the decomposition positive semi-definite.

The SVD is a powerful tool for linear algebra. The left and right singular vectors form a basis of the row and column spaces of $\mathbf{A}$ that are orthogonal. As such, the SVD can be used to compute the generalized inverse of $\mathbf{A}$ by using the *reciprocal* singular values in the decomposition. That is,

$$\mathbf{A}^\dagger = \mathbf{V}\mathbf{S}^{-1}\mathbf{U}' \quad (\text{B.3})$$

where $\mathbf{A}^\dagger$ is the generalized inverse and is equivalent to the Moore-Penrose pseudoinverse. The SVD provides a convenient and flexible framework for computing the generalized inverse because the rank of the singular value matrix can be controlled. By eliminating null singular values, the inverse operation is numerically stabilized. This *truncation* is a form of

regularization. In addition, the singular vectors point in the directions of maximum variance of the column and row spaces. In terms of signal processing, the variance is a measure of *information*. Thus, analysis of the singular vectors can provide insight into the inverse operation. A detailed description on the use of the SVD for inverse-problems may be found in Hansen (1998).

Appendix C

Simultaneous Retrieval of Temperature and Concentration Profiles

Before deriving the simultaneous retrieval approach, it is beneficial to review some basic definitions and fundamental atmospheric physics. The definition of optical depth was introduced in section 2.1.3. The optical depth can also be expressed in terms of the absorption *mass coefficient* or *mass cross-section*:

$$\delta = \int_0^z \beta_{abs} dz = \int_0^z C_\alpha \rho dz \quad (\text{C.1})$$

where C_α is the absorption mass cross-section in cm^2/g and ρ is the air mass density. The optical depth can also be represented in terms of the *mixing ratio* of gases of interest. The mixing ratio is defined as the ratio of the gas mass density to the air mass density. The relationship is easily derived from the hydrostatic equation:

$$\rho dz = -\frac{q}{g} dp \quad (\text{C.2})$$

where q is the mixing ratio, g is the acceleration due to gravity, and dp is some small incremental change in pressure. Multiplying both sides of the hydrostatic equation by the

absorption mass cross-section results in

$$C_\alpha \rho dz = -C_\alpha \frac{q}{g} dp \quad (\text{C.3})$$

Integrating both sides and using eq. (C.1) yields

$$\delta = - \int_{p_s}^0 C_\alpha \frac{q}{g} dp = \int_0^{p_s} C_\alpha \frac{q}{g} dp \quad (\text{C.4})$$

where p_s is the pressure at the surface boundary layer. Similarly, we can define the *optical mass*:

$$U(p) = \frac{1}{g} \int_0^p q(p') dp' \quad (\text{C.5})$$

This definition follows directly from the hydrostatic equation by setting the incremental optical mass dU equal to ρdz and integrating over a pressure altitude range. Note that in these equations we have assumed that altitude z of the sensor high enough that the ambient pressure is approximately zero.

With these definitions, we can now derive the perturbed form of the radiative transfer equation. We can rewrite the radiative transfer equation in (2.36) as

$$L(\lambda) = \tau(\lambda) \varepsilon(\lambda) L_{BB}(\lambda, T_s) - \int_0^z L_{BB}(\lambda, T_z) d\tau(\lambda, z) \quad (\text{C.6})$$

where the emissivity and the Planck function are used to represent the surface emission L_s . The explicit dependence of the Planck function on wavelength is noted because we need to distinguish between the spectral *regions* over which the weighting functions will be built. In other words, while the Planck function is nearly constant in the *neighborhood* of the center absorption line of a particular constituent, it will change considerably for different constituents. We can also write the equation in terms of pressure altitude

$$L(\lambda) = \tau(\lambda) \varepsilon(\lambda) L_{BB}(\lambda, T_s) - \int_0^{p_s} L_{BB}(\lambda, T_p) d\tau(\lambda, p) \quad (\text{C.7})$$

Hereafter, the notation will be simplified by denoting the wavelength dependence as a subscript and labelling the Planck profile as a function of pressure (i.e., $L_{BB}(T_p) = L_{BB}(p)$).

Thus, the expected spectral radiance from an initial estimate is

$$L_\lambda^o = \tau_\lambda^o(p_s) \varepsilon_\lambda^o L_{BB_\lambda}^o(p_s) - \int_0^{p_s} L_{BB_\lambda}^o(p) d\tau_\lambda^o(p) \quad (C.8)$$

To simplify the analysis, we will assume a blackbody surface (i.e., $\varepsilon_\lambda=1$). Subtracting the estimated initial spectrum from the true spectrum results in

$$\begin{aligned} \Delta L_\lambda = L_\lambda - L_\lambda^o &= \tau_\lambda(p_s) L_{BB_\lambda}(p_s) - \tau_\lambda^o(p_s) L_{BB_\lambda}^o(p_s) \\ &\quad - \int_0^{p_s} L_{BB_\lambda}(p) d\tau_\lambda(p) + \int_0^{p_s} L_{BB_\lambda}^o(p) d\tau_\lambda^o(p) \end{aligned} \quad (C.9)$$

We now define the following perturbations

$$\begin{aligned} \Delta L_{BB_\lambda}(p) &= L_{BB_\lambda}(p) - L_{BB_\lambda}^o(p) \\ \Delta \tau_\lambda(p) &= \tau_\lambda(p) - \tau_\lambda^o(p) \end{aligned} \quad (C.10)$$

Rewriting eq. (C.9) in terms of these perturbations yields

$$\begin{aligned} \Delta L_\lambda &= L_{BB_\lambda}(p_s) \Delta \tau_\lambda(p_s) + \Delta L_{BB_\lambda}(p_s) \tau_\lambda^o(p_s) \\ &\quad - \int_0^{p_s} L_{BB_\lambda}(p) d[\Delta \tau_\lambda(p)] - \int_0^{p_s} \Delta L_{BB_\lambda}(p) d\tau_\lambda^o(p) \end{aligned} \quad (C.11)$$

The first integral is evaluated by parts so that

$$\int_0^{p_s} L_{BB_\lambda}(p) d[\Delta \tau_\lambda(p)] = L_{BB_\lambda}(p_s) \Delta \tau_\lambda(p_s) - \int_0^{p_s} \Delta \tau_\lambda(p) dL_{BB_\lambda}(p) \quad (C.12)$$

Substituting into eq. (C.11) yields

$$\Delta L_\lambda = \Delta L_{BB_\lambda}(p_s) \tau_\lambda^o(p_s) - \int_0^{p_s} \Delta L_{BB_\lambda}(p) d\tau_\lambda^o(p) + \int_0^{p_s} \Delta \tau_\lambda(p) dL_{BB_\lambda}(p) \quad (C.13)$$

This equation depends on the perturbed values of the Planck function. We wish to show the direct dependence on temperature so that we can solve for a perturbed temperature directly. To do this, the Planck function is expanded with a Taylor series about T_p^o :

$$L_{BB_\lambda}(T_p) = L_{BB_\lambda}(T_p^o) + \left. \frac{\partial L_{BB_\lambda}(T_p)}{\partial T_p} \right|_{T_p^o} (T_p - T_p^o) + \left. \frac{\partial^2 L_{BB_\lambda}(T_p)}{\partial T_p^2} \right|_{T_p^o} (T_p - T_p^o)^2 \dots \quad (C.14)$$

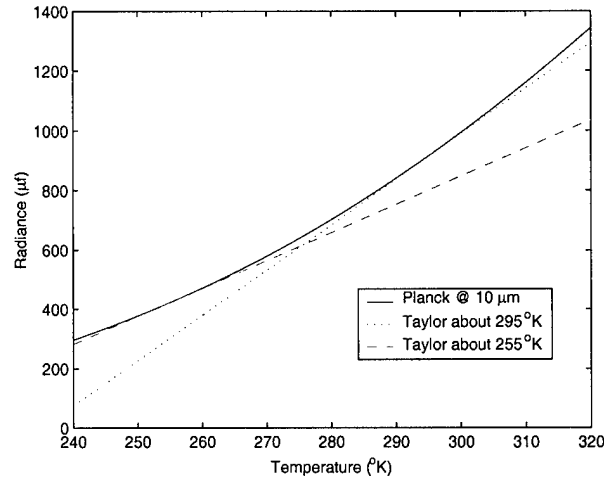


Figure C.1: Taylor approximation of the Planck function.

To make the notation more concise, let

$$k_{\lambda}(p) = \frac{\partial L_{BB_{\lambda}}(T_p)}{\partial T_p} = \frac{c_1 c_2 e^{c_2/\lambda T}}{\lambda^6 T^2 (e^{c_2/\lambda T} - 1)^2} \quad (\text{C.15})$$

and $\Delta T(p) = T_p - T_p^o$. Because the Planck function is smooth, it can be reasonably estimated by a truncated Taylor series expansion. Thus, the function is “linearized” by using the first-order Taylor approximation resulting in

$$\Delta L_{BB_{\lambda}}(p) = k_{\lambda}^o(p) \Delta T(p) \quad (\text{C.16})$$

In the limit $\Delta T(p) \rightarrow 0$, the finite difference approximation results:

$$\Delta T(p) \approx dL_{BB_{\lambda}}(p) = k_{\lambda}^o dT(p) \quad (\text{C.17})$$

Figure C.1 shows a plot of the Planck function evaluated at 10 μm and two first-order Taylor series approximations centered about 295 °K and 255 °K, respectively. Note that the approximation is very good when $\Delta T(p)$ is less than about 10 °K. Initial temperature estimates that are this close to the true profile should not be difficult to obtain.

Now we can substitute eqs. (C.16) and (C.17) into eq. (C.11) so that

$$\begin{aligned} \Delta L_\lambda = k_\lambda^o(p_s) \Delta T(p_s) \tau_\lambda^o(p_s) &- \int_0^{p_s} k_\lambda^o(p) \Delta T(p) \frac{d\tau_\lambda^o(p)}{dp} dp \\ &+ \int_0^{p_s} k_\lambda^o(p) \Delta \tau_\lambda(p) \frac{d\tau_\lambda^o(p)}{dp} dp \end{aligned} \quad (C.18)$$

The only term that still contains an “unknown” is the perturbation of the transmission $\Delta \tau_\lambda(p)$. The goal of the following development is to represent this quantity in terms of the optical mass, which can then be recast in terms of the perturbed temperature profile. We begin by using the definition of the total transmission given by eq. (2.81) and using the definition for the optical mass so that

$$\tau_\lambda(p) = \prod_i \tau_{\lambda_i}(p) = \exp \left[- \sum_i \int_0^p C_{\alpha_\lambda} \frac{dU_i(p')}{dp'} dp' \right] \quad (C.19)$$

A small incremental change would then be

$$\begin{aligned} d\tau_\lambda(p) &= d \left[- \sum_i \int_0^p C_{\alpha_\lambda} \frac{dU_i(p')}{dp'} dp' \right] \exp \left[- \sum_i \int_0^p C_{\alpha_\lambda} \frac{dU_i(p')}{dp'} dp' \right] \\ &= \tau_\lambda(p) d \left[- \sum_i \int_0^p C_{\alpha_\lambda} \frac{dU_i(p')}{dp'} dp' \right] \end{aligned} \quad (C.20)$$

We can use eq. (C.19) to redefine $d\tau_\lambda(p)$ in terms of the natural logarithm of the transmission profiles of the individual constituents:

$$\ln \tau_{\lambda_i}(p) = - \int_0^p C_{\alpha_\lambda} \frac{dU_i(p')}{dp'} dp' \quad (C.21)$$

which implies that

$$d\tau_\lambda(p) = \tau_\lambda(p) \sum_i d \ln \tau_{\lambda_i}(p) \quad (C.22)$$

Thus, the initial estimate is

$$d\tau_\lambda^o(p) = \tau_\lambda^o(p) \sum_i d \ln \tau_{\lambda_i}^o(p) \quad (C.23)$$

and using the finite difference approximation $d\tau_\lambda^o(p) \approx \Delta \tau_\lambda(p)$ then

$$\Delta \tau_\lambda(p) = \tau_\lambda^o(p) \sum_i \Delta \ln(\tau_{\lambda_i}(p)) \quad (C.24)$$

The perturbation of the logarithmic transmission is obtained from the relationship established in eq. (C.22) so that

$$\Delta \ln \tau_{\lambda_i}(p) = - \int_0^p C_{\alpha_\lambda} \frac{d \Delta U_i(p')}{dp'} dp' \quad (C.25)$$

Now we integrate this equation by parts to get

$$\Delta \ln \tau_{\lambda_i}(p) = -C_{\alpha_\lambda} \Delta U_i(p) + \int_0^p \Delta U_i(p') \frac{dC_{\alpha_\lambda}}{dp'} dp' \quad (C.26)$$

But the second term goes to zero because C_{α_λ} is constant with respect to pressure. Also, from eq. (C.21) we can define

$$C_{\alpha_\lambda} = - \frac{d \ln \tau_{\alpha_i}^o(p)}{dU_i^o(p)} \quad (C.27)$$

This assumes that the initial estimates are formulated with a model that employs the correct absorption mass cross-sections. Thus, we end up with

$$\Delta \tau_\lambda(p) = \tau_\lambda^o(p) \sum_i \Delta U_i(p) \frac{d \ln \tau_{\lambda_i}^o(p)}{dU_i^o(p)} \quad (C.28)$$

Now we can substitute this expression into equation (C.18) to get

$$\begin{aligned} \Delta L_\lambda &= k_\lambda^o(p_s) \Delta T(p_s) \tau_\lambda^o(p_s) - \int_0^{p_s} k_\lambda^o(p) \Delta T(p) \frac{d \tau_\lambda^o(p)}{dp} dp \\ &+ \int_0^{p_s} k_\lambda^o(p) \left[\tau_\lambda^o(p) \sum_i \Delta U_i(p) \frac{d \ln \tau_{\lambda_i}^o(p)}{dU_i^o(p)} \right] \frac{dT(p)}{dp} dp \end{aligned} \quad (C.29)$$

and rearranging

$$\begin{aligned} \Delta L_\lambda &= k_\lambda^o(p_s) \Delta T(p_s) \tau_\lambda^o(p_s) - \int_0^{p_s} k_\lambda^o(p) \Delta T(p) \frac{d \tau_\lambda^o(p)}{dp} dp \\ &+ \sum_i \int_0^{p_s} k_\lambda^o(p) \Delta U_i(p) \frac{dT(p)}{dU_i^o(p)} \tau_\lambda^o(p) \frac{d \ln \tau_{\lambda_i}^o(p)}{dp} dp \end{aligned} \quad (C.30)$$

After substituting eq. (C.23) we get

$$\begin{aligned} \Delta L_\lambda &= k_\lambda^o(p_s) \Delta T(p_s) \tau_\lambda^o(p_s) - \sum_i \int_0^{p_s} \left[k_\lambda^o(p) \Delta T(p) \tau_\lambda^o(p) \frac{d \ln \tau_{\lambda_i}^o(p)}{dp} \right. \\ &\quad \left. - k_\lambda^o(p) \Delta U_i(p) \frac{dT(p)}{dU_i^o(p)} \tau_\lambda^o(p) \frac{d \ln \tau_{\lambda_i}^o(p)}{dp} \right] dp \end{aligned} \quad (C.31)$$

Finally, we let

$$\Delta T_i(p) = \Delta T(p) - \Delta U_i(p) \frac{dT(p)}{dU_i^o(p)} \quad (\text{C.32})$$

That is, each constituent contributes to the perturbed temperature profile based on the temperature dependence on that constituent (the derivative term) and how much of the constituent is present ($\Delta U_i(p)$). Substituting eq. (C.32) into eq. (C.31) yields

$$\Delta L_\lambda = k_\lambda^o(p_s) \Delta T(p_s) \tau_\lambda^o(p_s) - \sum_i \int_0^{p_s} k_\lambda^o(p) \Delta T_i(p) \tau_\lambda^o(p) \frac{d \ln \tau_{\lambda_i}^o(p)}{dp} dp \quad (\text{C.33})$$

where the perturbed temperature profile of the i^{th} constituent is

$$\Delta T_i(p) = T_i(p) - T^o(p) \quad (\text{C.34})$$

Thus, the same initial estimate of the temperature profile $T^o(p)$ is used for *all* atmospheric constituents. Equation (C.33) is the final form of the perturbed radiative transfer equation. The surface temperature differential can be approximated by making a measurement in a highly transmissive region of the spectrum. The perturbed temperature profile can be solved for by applying one of the inversion techniques discussed in Section 2.1.5. The retrieval of temperature and concentration profiles can be summarized in three steps:

1. Build weighting functions based on an initial estimate of the atmospheric profiles and solve for $\Delta T_i(p)$ using a linear or nonlinear inversion technique.
2. Let the true temperature profile be $T(p) = T^o(p) + \Delta T_j(p)$ where j denotes a constituent that is well mixed (i.e., CO_2).
3. Solve for the concentration of other constituents by equating eqs. (C.32) and (C.34) and solving for the optical mass:

$$U_i(p) = U_i^o(p) + \frac{dU_i^o(p)}{dT(p)} [T(p) - T_i(p)] \quad (\text{C.35})$$

or in terms of the mixing ratio

$$q_i(p) = -g \frac{dU_i(p)}{dp} \quad (\text{C.36})$$

Appendix D

Other Multivariate Regression Methods

This appendix provides a description of two alternative multivariate regression models. The concept is similar to CCR and PCR in that a multivariate regression relating $\mathbf{X}$ and $\mathbf{Y}$ is based on a latent lower-dimensional space.

D.1 Partial Least Squares

Another approach that is similar to canonical correlations is Partial Least Squares (PLS). The technique was originated by Herman Wold for the determination of latent paths and the creation of simplified models (Geladi 1988). Further development of PLS was led by research in chemometrics. In this field, the goal is to quantify the concentration of a particular absorber based on the spectroscopic measurement. The dataset used to build the regression is known as a “calibration” set. The path model for PLS is shown in Figure D.1. Note that the arrows flow in only one direction. This is because the PLS analysis is not symmetric and is biased toward predicting one set from the other. The linear combinations obtained from PLS maximize the *covariance* between data sets. This is done subject to the constraint that the scores from the linear transformations have maximum variance (just as

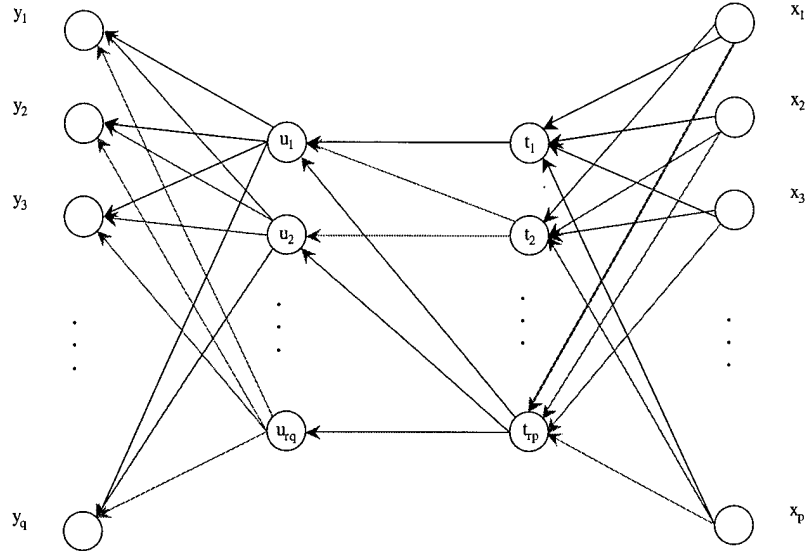


Figure D.1: Path model for partial least squares.

in PCA). Therefore, PLS can be viewed as a “compromise” between CCA and PCA because the solution must maximize variance and covariance simultaneously. This constraint also results in a latent correlation space that is *not* orthogonal. This is depicted in Figure D.1 by the cross-paths between the latent variables $\mathbf{u}$ and $\mathbf{t}$, which are analogous to the CCA and PCA scores. These latent cross-paths occur between the same or lower dimensions and not between lower and higher dimensions (from the perspective of mapping $\mathbf{T}$ to $\mathbf{U}$). This results in an upper-triangular matrix $\mathbf{T}'\mathbf{U}$ (i.e., $\mathbf{t}_i'\mathbf{u}_j = 0$ for $i > j$).

The original PLS algorithm, known as the Nonlinear Iterative Partial Least Squares (NIPALS) algorithm, employs a numerical technique for the computation of eigenvectors (Wold 1984). A popular implementation of the PLS algorithm is based on the NIPALS approach. One unique aspect of this approach is that the latent dimensions (i.e., the eigenvectors) are computed iteratively one at a time. The algorithm computes a “dominant” eigenvector from the covariance matrix $\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yx}}$ and attempts to predict $\mathbf{Y}$ just using this linear combination. The predicted $\hat{\mathbf{Y}}$ is then compared to the true $\hat{\mathbf{Y}}$. Likewise, the original

$\mathbf{X}$ is compared to an estimate $\hat{\mathbf{X}}$ obtained with just the first eigenvector. The next eigenvector is computed from the *residual* $\mathbf{X} - \hat{\mathbf{X}}$ and regressed to the residual of $\mathbf{Y}$. This process is repeated until convergence or until the number of eigenvectors is equal to the rank of $\mathbf{X}$ (i.e., $r \leq \min(p, n)$). The operation on the residuals ensures the orthogonality of the weights. It also provides an inherent method for finding the latent rank of the data. The analysis has the effect of “shrinking” the data until there is no more information to be exploited. One way to determine whether all the information has been used is to compute the determinant of the residual $\mathbf{X}$ matrix. The determinant can be thought of as a measure of the *mass* of a matrix. Thus, the algorithm stops when the mass of the matrix has dropped below some cutoff point r . This point then defines the rank. Another advantage of PLS is that it can be implemented so that no matrix-inverse operations need to be made. The PLS method presented in this section is often called “two-block” PLS because it shrinks two blocks of data (i.e., $\mathbf{X}$ and $\mathbf{Y}$), making it suitable for multivariate regression.

Two aspects make the PLS technique difficult to interpret: (1) the iterative approach used to find the eigenvectors; and (2) the way the regression coefficients are computed. The work of Helland (1988), (Höskuldsson 1988), (Stone and Brooks 1990), and (Phatak 1993) have done much to explain PLS and to put the approach in the same context as other multivariate regression methods. Another attempt is made here. The algorithm is initialized by letting $\mathbf{E}_0 = \mathbf{X}$ and $\mathbf{F}_0 = \mathbf{Y}$, where $\mathbf{X}$ and $\mathbf{Y}$ are $n \times p$ and $n \times q$ matrices, respectively. Then for $i = 1$ to k :

1. Let $\mathbf{u}_i$ be any column of $\mathbf{Y}$. At this point, $\mathbf{u}_i$ is an $n \times 1$ column vector estimating the score obtained from the transformation $\mathbf{Y}\mathbf{c}$, where $\mathbf{c}$ is a $q \times 1$ column vector defining the latent path from $\mathbf{Y}$ to the first dimension of $\mathbf{U}$.
2. $\mathbf{w}'_i = (\mathbf{u}'_i\mathbf{u}_i)^{-1}\mathbf{u}'_i\mathbf{E}_{i-1}$. The $\mathbf{w}$ vector is a $p \times 1$ vector of weights that describe the mapping from the original $\mathbf{X}$ space to the latent variable space. By definition, it is the same as the regression coefficients (least-squares solution) relating $\mathbf{u}$ to $\mathbf{E}$.

3. $\mathbf{t}_i = \mathbf{E}_{i-1} \mathbf{w}_i$ performs the mapping to the latent space defined by $\mathbf{w}_i$.
4. $\mathbf{c}'_i = (\mathbf{t}'_i \mathbf{t}_i)^{-1} \mathbf{t}'_i \mathbf{F}_{i-1}$. The $\mathbf{c}$ is the least-squares solution relating the latent variables $\mathbf{t}$ to the original variables in $\mathbf{Y}$. It is therefore a $1 \times q$ vector.
5. $\mathbf{c}_i = \mathbf{c}_i / (\mathbf{c}'_i \mathbf{c}_i)$ for unity normalization.
6. $\mathbf{u}_i = \mathbf{F}_{i-1} \mathbf{c}_i$ maps the original $\mathbf{Y}$ variables to the latent variable $\mathbf{u}$.
7. Iterate on steps 2-6 until convergence. During the iteration, several "partial" least-squares regressions are performed. The regressions are done in a criss-cross pattern going back and forth between the latent variables and the original variable spaces. These iterations converge on solutions for $\mathbf{w}$ and $\mathbf{c}$ that are the basis of the latent variables.
8. $\mathbf{p}'_i = (\mathbf{t}'_i \mathbf{t}_i)^{-1} \mathbf{t}'_i \mathbf{E}_{i-1}$ is the least-squares solution relating the latent variable $\mathbf{t}_i$ to the $\mathbf{X}$ data. $\mathbf{p}$ is a $p \times 1$ vector representing an inverse mapping from the latent space to the original space (i.e., *loading*).
9. $\mathbf{q}'_i = (\mathbf{u}'_i \mathbf{u}_i)^{-1} \mathbf{u}'_i \mathbf{F}_{i-1}$ is the loading for $\mathbf{Y}$.
10. $b_i = (\mathbf{t}'_i \mathbf{t}_i)^{-1} \mathbf{t}'_i \mathbf{u}_i$ is the regression coefficient relating the latent variables.
11. $\mathbf{E}_i = \mathbf{E}_{i-1} - \mathbf{t}_i \mathbf{p}'_i$. When $i = 1$, this computes the residual between the $\mathbf{X}$ data and an estimate of $\mathbf{X}$ based on the first loading. Subsequent analysis is based on the residual matrix, which contains all information not spanned by the first loading. This has the effect of "shrinking" the matrix.
12. $\mathbf{F}_i = \mathbf{F}_{i-1} - b_i \mathbf{t}_i \mathbf{c}'_i$ does the same as the previous step for the $\mathbf{Y}$ data. The $\mathbf{F}$ matrix is reconstructed with the information spanned by the $\mathbf{t}$ latent variables. It is a least-squares estimate of $\mathbf{F}$ weighted by the regression coefficient b relating the latent variables. It is also an estimate of the bilinear expansion $\mathbf{u}_i \mathbf{q}'_i$ defining $\mathbf{F}_i$.

13. The procedure is done all over again for the next dimension and is based on the *residual* data matrices from the previous iteration. The operation on the residuals ensures that the latent variables are *linearly independent*. However, the relationship between the latent variables is not orthogonal.

The resulting k dimensions can then be grouped into matrices so that

$$\hat{\mathbf{Y}}_{PLS} = \mathbf{T}\mathbf{B}\mathbf{C}' \quad (\text{D.1})$$

where $\mathbf{T}$ is $n \times k$, $\mathbf{B}$ is a $k \times k$ diagonal matrix, and $\mathbf{C}$ is $q \times k$. In terms of the original variables, $\mathbf{T} = \mathbf{E}\mathbf{W}$ where $\mathbf{W}$ is a matrix whose columns are the vectors $\mathbf{w}'_i$. The PLS estimate of $\mathbf{Y}$ can therefore be rewritten as

$$\begin{aligned} \hat{\mathbf{Y}}_{PLS} &= \mathbf{E}\mathbf{W}\mathbf{B}\mathbf{C}' \\ &= \mathbf{E}\mathbf{W}[(\mathbf{T}'\mathbf{T})^{-1}\mathbf{T}'\mathbf{U}]\mathbf{C}' \\ &= \mathbf{E}\mathbf{W}[(\mathbf{W}'\mathbf{E}'\mathbf{E}\mathbf{W})^{-1}\mathbf{W}'\mathbf{E}']\mathbf{F}\mathbf{C}\mathbf{C}' \end{aligned} \quad (\text{D.2})$$

Thus, the PLS regression coefficients relating $\mathbf{X}$ to $\mathbf{Y}$ can be defined as

$$\beta_{PLS} = \mathbf{W}(\mathbf{W}'\mathbf{X}'\mathbf{X}\mathbf{W})^{-1}\mathbf{W}'\mathbf{X}'\mathbf{Y}\mathbf{C}\mathbf{C}' \quad (\text{D.3})$$

Since $\mathbf{Y} = \mathbf{U}\mathbf{C}'$ then

$$\beta_{PLS} = \mathbf{W}(\mathbf{W}'\mathbf{X}'\mathbf{X}\mathbf{W})^{-1}\mathbf{W}'\mathbf{X}'\mathbf{Y} \quad (\text{D.4})$$

If the inverse operation needs to be avoided, PLS regression can be implemented so that the prediction of $\mathbf{Y}$ is built iteratively. The procedure is similar to the computation of the eigenvectors. For $i = 1$ to k , map the new data $\mathbf{E}$ with $\mathbf{w}_i$ to get $\mathbf{t}_i$. At each step, let $\mathbf{F}_i = \mathbf{F}_{i-1} + b_i\mathbf{t}_i\mathbf{c}'_i$. Here, b and $\mathbf{c}$ are obtained from the model built with the regression data and $\mathbf{t}$ is obtained with the new observations. The new data $\mathbf{X}$ is also “shrunk” based on $\mathbf{t}_i\mathbf{p}'_i$.

The formulation of the regression coefficients given in eq. (D.4) is useful to interpret the optimization achieved by PLS. The linear combinations $\mathbf{W}$ and $\mathbf{C}$ result in latent variables that have maximum covariance, subject to the constraint of maximum variance.

Proof. Upon convergence,

$$\begin{aligned}\mathbf{W} &= \mathbf{E}'\mathbf{U}(\mathbf{U}'\mathbf{U})^{-1} \\ &= \mathbf{E}'\mathbf{F}\mathbf{C}(\mathbf{C}'\mathbf{F}'\mathbf{F}\mathbf{C})^{-1}\end{aligned}\tag{D.5}$$

Multiplying both sides by $\mathbf{C}'\mathbf{F}'\mathbf{F}\mathbf{C}$ yields

$$\mathbf{W}(\mathbf{C}'\mathbf{F}'\mathbf{F}\mathbf{C}) = \mathbf{E}'\mathbf{F}\mathbf{C}\tag{D.6}$$

Substituting $\mathbf{C} = \mathbf{F}'\mathbf{T}(\mathbf{T}'\mathbf{T})^{-1}$ and using $\mathbf{T} = \mathbf{E}\mathbf{W}$ results in

$$\mathbf{W}(\mathbf{C}'\mathbf{F}'\mathbf{F}\mathbf{C}) = \mathbf{E}'\mathbf{F}\mathbf{F}'\mathbf{E}\mathbf{W}(\mathbf{W}'\mathbf{E}'\mathbf{E}\mathbf{W})^{-1}\tag{D.7}$$

Multiplying both sides by $\mathbf{W}'\mathbf{E}'\mathbf{E}\mathbf{W}$ yields

$$\mathbf{W}(\mathbf{C}'\mathbf{F}'\mathbf{F}\mathbf{C})(\mathbf{W}'\mathbf{E}'\mathbf{E}\mathbf{W}) = \mathbf{E}'\mathbf{F}\mathbf{F}'\mathbf{E}\mathbf{W}\tag{D.8}$$

Therefore, the $\mathbf{W}$ are the eigenvectors resulting from

$$\mathbf{W}\mathbf{\Lambda} = \mathbf{E}'\mathbf{F}\mathbf{F}'\mathbf{E}\mathbf{W}\tag{D.9}$$

where $\mathbf{\Lambda} = (\mathbf{C}'\mathbf{F}'\mathbf{F}\mathbf{C})(\mathbf{W}'\mathbf{E}'\mathbf{E}\mathbf{W})$. Therefore, PLS maximizes the squared covariances between $\mathbf{X}$ and $\mathbf{Y}$. This is in contrast to CCA where the *correlations* are maximized. Thus, the PLS results depend on the scaling of the data whereas CCA results do not. The maximization of covariance in PLS is done subject to the constraint of maximum variance.

From the definition of $\mathbf{\Lambda}$,

$$\mathbf{W}'\mathbf{E}'\mathbf{E}\mathbf{W} = (\mathbf{C}'\mathbf{F}'\mathbf{F}\mathbf{C})^{-1}\mathbf{\Lambda}\tag{D.10}$$

Multiplying by $\mathbf{W}$ and using the orthonormality property from eq. (D.9) (i.e., $\mathbf{W}'\mathbf{W} = \mathbf{I}$ and $\mathbf{W}' = \mathbf{W}^{-1}$) results in

$$\mathbf{E}'\mathbf{E}\mathbf{W} = \mathbf{W}(\mathbf{C}'\mathbf{F}'\mathbf{F}\mathbf{C})^{-1}\mathbf{\Lambda}\tag{D.11}$$

So that

$$\mathbf{E}'\mathbf{E}\mathbf{W} = \mathbf{W}\mathbf{\Lambda}_E\tag{D.12}$$

where $\Lambda_E = (\mathbf{C}'\mathbf{F}'\mathbf{F}\mathbf{C})^{-1}\Lambda$. Thus, the $\mathbf{W}$ maximize the variance in $\mathbf{X}$. This also results in

$$\mathbf{C}'\mathbf{F}'\mathbf{F}\mathbf{C} = \Lambda\Lambda_E^{-1} \quad (\text{D.13})$$

which results in the maximization of variance of $\mathbf{U}$, subject to the constraint of maximum covariances and maximum variance of $\mathbf{T}$. $\square$

D.2 Maximum Redundancy

In the previous section, it was shown that PLS maximizes the covariance between the latent variables. The PLS latent variables are derived subject to the constraint that they represent the maximum variance in $\mathbf{X}$ and $\mathbf{Y}$. This in contrast to PCR where the sole criterion is to maximize the $\mathbf{X}$ covariance matrix. The other extreme would be to find the latent variables that maximize the $\mathbf{Y}$ covariance matrix. This, of course, could be accomplished via a standard PCA of the $\mathbf{Y}$ data. However, the purpose of the analysis is to estimate $\mathbf{Y}$ with $\hat{\mathbf{Y}}$ based on the observations $\mathbf{X}$. Therefore, the alternative is to maximize the covariance matrix of $\hat{\mathbf{Y}}$ from linear combinations of $\mathbf{X}$ (i.e., latent variables of $\mathbf{X}$). One way of doing this is to find the linear combinations of $\mathbf{X}$ that maximize the *Redundancy Index* (RI):

$$\text{RI} = \frac{\text{tr}(\hat{\mathbf{Y}}'\hat{\mathbf{Y}})}{\text{tr}(\mathbf{Y}'\mathbf{Y})} \quad (\text{D.14})$$

where $\text{tr}(\cdot)$ is the trace operator. The trace is the sum of the diagonal elements of a matrix and is equal to the sum of the eigenvalues. This “Maximum Redundancy” (MR) approach was developed by Van den Wollenberg in 1977. It turns out, however, that the same multivariate method had already been derived under different names and using different objective functions (Merola 1998). It can be viewed as a “Reduced Rank Regression” where the OLS estimates of $\mathbf{Y}$ are derived from rank-reduced regression coefficients (Izenman 1975) or as a “principal components of $\mathbf{Y}$ relative to $\mathbf{X}$ ” (Merola 1998). As such, the path model for MR is identical to that of PCR (Fig. 3.2) except that the latent variables are different. Regardless of the derivation, the optimal linear combinations of $\mathbf{X}$ are found from

the eigenvalue/eigenvector solution to

$$\Sigma_{xx}^{-1}\Sigma_{xy}\Sigma_{yx}\mathbf{A} = \mathbf{A}\Phi \quad (\text{D.15})$$

where $\mathbf{A}$ are the MR *weights* and Φ is the diagonal matrix of eigenvalues. The eigenvalues turn out to be the maximum variances in $\hat{\mathbf{Y}}$.

The derivation presented here is based on the maximum redundancy concept. Since the trace of the covariance matrix is equal to the sum of the eigenvalues, the maximization can be expressed in terms of the solution to the eigenvalue problem

$$\hat{\mathbf{Y}}'\hat{\mathbf{Y}}(\mathbf{U}_{q \times q}) = (\mathbf{U}_{q \times q})\Phi \quad (\text{D.16})$$

where $\mathbf{U}_{q \times q}$ is a matrix of eigenvectors, which are the principal components of $\hat{\mathbf{Y}}$. Thus, the eigenvalues Φ are the largest variances in $\hat{\mathbf{Y}}$. Our goal, however, is to force the *scores* of $\mathbf{X}$ (i.e, $\mathbf{U} = \mathbf{X}\mathbf{A}$) to be the principal components of $\hat{\mathbf{Y}}$. The scores are $n \times k$ matrices, where the latent rank $k \leq \min(n, q)$. This requires a modification of eq. (D.16) so that it uses the *outer-product* covariance matrix $\hat{\mathbf{Y}}\hat{\mathbf{Y}}'$. This is not a problem since the nonzero eigenvalues of $\hat{\mathbf{Y}}\hat{\mathbf{Y}}'$ and $\hat{\mathbf{Y}}'\hat{\mathbf{Y}}$ are identical. Thus, the scores can be made to be the principal components by using

$$\hat{\mathbf{Y}}\hat{\mathbf{Y}}'(\mathbf{U}_{n \times k}) = (\mathbf{U}_{n \times k})\Phi \quad (\text{D.17})$$

If $n < q$, then $\mathbf{U}$ is a square orthogonal matrix of eigenvectors. Otherwise, $\mathbf{U}$ is $n \times q$. Letting $\mathbf{U} = \mathbf{X}\mathbf{A}$, eq. (D.17) reduces to eq. (D.15).

Proof. The OLS estimate of $\mathbf{Y}$ is

$$\hat{\mathbf{Y}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \quad (\text{D.18})$$

Substituting this equation for $\hat{\mathbf{Y}}$ in eq. (D.17) and letting $\mathbf{U} = \mathbf{X}\mathbf{A}$ results in

$$\hat{\mathbf{Y}}\hat{\mathbf{Y}}'(\mathbf{U}_{n \times k}) = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\mathbf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\mathbf{A} = \mathbf{X}\mathbf{A}\Phi \quad (\text{D.19})$$

Cancelling like-terms and simplifying yields

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\mathbf{Y}'\mathbf{X}\mathbf{A} = \mathbf{A}\Phi \quad (\text{D.20})$$

If the data is mean-centered and scaled by the number of observations, then $\mathbf{X}'\mathbf{X} = \Sigma_{\mathbf{xx}}$. Therefore,

$$\hat{\mathbf{Y}}\hat{\mathbf{Y}}'(\mathbf{U}_{\mathbf{n} \times \mathbf{k}}) = \Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yx}}\mathbf{A} = \mathbf{A}\Phi \quad (\text{D.21})$$

□

While a latent subspace of $\mathbf{Y}$ does not need to be defined in MR, it may be useful to do so for interpretation, outlier tests, and discriminant analysis of the subspace. The linear combinations $\mathbf{V} = \mathbf{Y}\mathbf{B}$ that maximize the covariance with the linear combinations $\mathbf{U} = \mathbf{X}\mathbf{A}$, subject to the maximum redundancy constraint, are obtained from

$$\mathbf{B} = \mathbf{Y}'\mathbf{X}\mathbf{A} \quad (\text{D.22})$$

which is simply a projection of the $\mathbf{Y}$ data onto the latent (redundancy) variables from the $\mathbf{X}$ data. Since the redundancy variables of $\mathbf{X}$ are the principal components of $\hat{\mathbf{Y}}$, then the redundancy weights defining the basis of $\mathbf{V}$ are *scores*. This offers an interpretation of the $\mathbf{Y}$ redundancy variables as the $\mathbf{Y}$ data rotated by the scores derived from its OLS subspace. Therefore, $\mathbf{V}$ is *proportional* to the variance in $\mathbf{Y}$.

The Maximum Redundancy solutions can be thought of as CCR solutions biased toward dimensions that maximize the variance in $\hat{\mathbf{Y}}$. Because of this bias, the correlations between the latent variables $\mathbf{U}$ and $\mathbf{V}$ are not orthogonal. The bias results in linear combinations $\mathbf{A}$ and $\mathbf{B}$ that are derived from eigenvalue equations that are similar to that of CCR. For example, the only difference between the MR and CCR eigenvalue equations for $\mathbf{A}$ is that $\Sigma_{\mathbf{yy}}^{-1}$ is missing in the MR equation. Thus, MR and CCR are the same for the special case where the variables in $\mathbf{Y}$ are uncorrelated and have equal variance. The bias makes MR optimal for the estimation of $\mathbf{Y}$ because the MR variables $\mathbf{U}$ form an orthogonal subspace of the OLS estimate of $\mathbf{Y}$ that maximizes the variance in $\hat{\mathbf{Y}}$. All other MR properties are subject to this constraint. This is in contrast to CCR where the estimate of $\mathbf{Y}$ is optimal subject to the constraint that the latent variables are maximally correlated.

D.3 Implementation

Principal Components Regression (PCR) and Maximum Redundancy (MR) were implemented in a manner similar to CCR. That is, a latent space is derived and a rank is estimated based on the properties of the eigenvalues. For these methods, the latent space is truncated to the dimension where the running sum of the eigenvalues is 99.5% of the total sum of eigenvalues. Typically, PCR is implemented so that the PC's of $\mathbf{X}$ are regressed directly on $\mathbf{Y}$. In this research, PCR is implemented as a "two-block" PCA where the PC's from $\mathbf{X}$ and $\mathbf{Y}$ are regressed. Thus, the estimate of $\mathbf{Y}$ is obtained by applying the inverse PC transform to the estimated $\mathbf{Y}$ PC's.

PLS was also a "two-block" implementation using the NIPALS or "canonical" method as described by Phatak (1993) and Höskuldsson (1988). There are two times when stopping criteria must be defined for this algorithm. One defines the number of iterations needed to converge on a vector that lies along the direction of maximum covariance. The second determines how many dimensions are retained for the regression. The iterations for the calculation of the "dominant" eigenvector were stopped based on the change in the "canonical" variate $\mathbf{u}$ determined by this eigenvector. Thus, convergence is achieved when $\|\mathbf{u}_i - \mathbf{u}_{i-1}\| \leq \delta$, where $\|\cdot\|$ is the Euclidean norm and i denotes the iteration step. $\delta = 0.01$ appears to work well for the data sets analyzed in this research. The next step is in the rank determination. As described in Section D.1, the PLS algorithm works on the residual matrices obtained from the regression of previous dimensions. The stopping criterion was $|\mathbf{E}_i| < \delta_e$, where $|\cdot|$ is the determinant and $\mathbf{E}_i$ is the residual matrix of $\mathbf{X}$. A value $\delta_e = 0.05$ resulted in latent-rank estimates that were consistent with those derived with CCR.

The algorithms for PLS, MR, CCR, and PCR were written in IDL and MATLAB and tested with the Linnerud data set used by Jackson (1991). The results were identical to those presented in the literature, thus validating the code implementation. Furthermore, the properties of the matrices involved (e.g., $\mathbf{W}'\mathbf{W} = \mathbf{I}$ for PLS) were also verified.

Appendix E

ISAC Implementation and Validation

The ISAC algorithm obtains the spectral transmission and upwelled radiance from a regression of the observed radiances and the calculated radiances. There are three approaches to the regression: (1) do a standard least-squares regression using only the blackbody pixels; (2) do a Kolmogorov-Smirnov fit across the top of a scatter plot with all or some of the pixels in the image; (3) do a “normalized” regression where outliers are automatically rejected with the goal of normalizing the residuals. Hereafter, these three approaches are referred to by ISAC_{LS}, ISAC_{KS}, and ISAC_{NR} (or simply LS, KS, and NR), respectively.

E.1 Least-Squares Maximum-Hit Method

The ISAC_{LS} method uses the maximum-hit approach to find blackbody pixels in a scene. This assumes that pixels having the maximum brightness temperature are likely to be blackbodies. This is certainly the case, provided that the brightness temperature is measured at a wavelength where the atmospheric effects are minimal. Otherwise, the detected maximum brightness temperature may be biased by clouds or by an atmosphere that is considerably warmer than the surface. For this reason, the maximum-hit method implemented in this

research uses a *fixed* reference wavelength. The reference wavelength corresponds to the SEBASS and MASTER Band 46, which are both close to about $10\ \mu\text{m}$. This wavelength was selected because it coincides with the peak of the Planck radiation curve for typical terrestrial temperatures. In addition, the atmosphere has a relatively high transmission at this wavelength. The maximum-hit method uses the wavelength at which the most number of pixels have the maximum brightness temperature. In practice, it was better to find the most number of pixels that were *close* to the maximum brightness temperature. Forcing the pixels to have *exactly* the same maximum brightness temperature typically lead to very few pixels being chosen for regression. A deviation δT away from the maximum brightness temperature may be driven by system noise. Therefore, pixels with brightness temperatures that fall within a δT equal to the sensor $\text{NE}\Delta T$ should be considered as “blackbody” for the regression. The approach to the selection of δT was heuristic. I found that $3 \leq \delta T \leq 5$ was necessary to get enough pixels for the regression.

E.2 Kolmogorov-Smirnov Method

The ISAC_{KS} algorithm is based on the Kolmogorov-Smirnov test for goodness-of-fit. It is an alternative approach to standard least-squares regression which can be more robust. The problem with the least squares method is that it depends on several assumptions about normality, homogeneity of variance, independence, and normal distribution of residuals. One solution is the use of a goodness-of-fit statistics such as the χ^2 and the Kolmogorov-Smirnov (KS) D statistics. The main advantage of the KS statistic is that it makes no assumptions about the probability distribution of the data and provides a more flexible criterion for goodness-of-fit. Section E.2.1 provides a generic description of the statistic and its computation. Section E.2.2 describes the asymptotic distribution of the KS statistic and how it can be used to test for goodness-of-fit. Finally, Section E.2.3 describes the use of KS for regression and how it is implemented in ISAC.

E.2.1 Kolmogorov-Smirnov Two-Sided Statistic

While the least squares method analyzes the squared difference of the data and the predicted value by the linear model, the Kolmogorov-Smirnov (KS) statistic compares the cumulative distribution functions of the data and the predicted model. The statistic is defined as

$$D = \max_x |S_n(x) - S_m(x)| \quad (\text{E.1})$$

where $S_n(x)$ and $S_m(x)$ are empirical cumulative distribution functions of two sets of random samples of sizes n and m that are in *ascending order*. Thus, for two sample sets $X_1, X_2, \dots, X_n$ and $Y_1, Y_2, \dots, Y_m$ the empirical distributions are:

$$S_n(x) = \begin{cases} 0 & x < X_1 \\ \frac{k}{n} & X_k \leq x < X_{k+1} \quad k = 1, 2, \dots, n-1 \\ 1 & x \geq X_n \end{cases} \quad (\text{E.2})$$

$$S_m(x) = \begin{cases} 0 & x < Y_1 \\ \frac{k}{m} & Y_k \leq x < Y_{k+1} \quad k = 1, 2, \dots, m-1 \\ 1 & x \geq Y_m \end{cases}$$

where k is an index that runs through the sequence of data points in the sample sets. These cumulative distributions are really proportions that define the fraction of observations that are less than or equal to the current x . Thus, if three of ten observations are less than or equal to some x in the data set then the value for $S(x)$ is 0.3. The two-sided KS statistic is then the maximum of the absolute difference between these two empirical cumulative distributions (Gibbons and Chakraborti 1992).

The way these two cumulative distributions are compared may not seem obvious. You cannot simply define each distribution and subtract each corresponding element. Instead, the index k for a particular distribution, say $S_n(x)$, is only advanced when the current X_k is less than or equal to the current Y_k . Otherwise, the index for $S_n(x)$ remains constant (i.e., it is evaluated at the same X) until Y_k is greater than X_k . To illustrate this, consider the sample set $X = 1, 2, 6$ and $Y = 3, 7, 8$. In this case, the first two elements of the array $S_n(x)$ would be evaluated using 1 and 2 and the first two elements of the array $S_m(x)$ would be evaluated at 3 and 3. On the third step, 6 is greater than 3 so $S_n(x)$ is evaluated again at 2 and $S_m(x)$ advances and gets evaluated at 7. The process continues until there are no more elements to evaluate in the set of observations. Once the new (“stretched”) cumulative distributions are obtained, the KS statistic can be calculated with equation E.1.

E.2.2 Kolmogorov-Smirnov Probability Distribution

Now that the D statistic has been obtained, it is necessary to determine whether it is significant or not. In other words, is the difference between the cumulative distributions only due to random variation in the data (i.e., it is not significant) or to the fact that the distributions are different? To determine this, a probability distribution function is needed. The KS function depends on the number of observations (which is related to the degrees of freedom). For the case where n and m are not equal, an effective number of observations is defined as follows:

$$N_e = \frac{nm}{n+m} \quad (\text{E.3})$$

otherwise $N_e = N = n = m$. The probability (i.e., “p-value”) that a random value from the KS distribution is greater than the observed D statistic is given by

$$P(D > D_{\text{observed}}) = Q_{KS} \left(\frac{\sqrt{N_e} + 0.12 + \frac{0.11}{\sqrt{N_e}}}{\sqrt{N_e}} D \right) \quad (\text{E.4})$$

where $Q_{KS}(\lambda)$ is a monotonic function defined as

$$Q_{KS}(\lambda) = 2 \sum_{j=1}^{\infty} (-1)^{j-1} e^{-2j^2 \lambda^2} \quad (\text{E.5})$$

The index j is arbitrary and is not related to the number of observations. In practice, the sum cannot be performed to infinity and there are numerical considerations for convergence. The stopping criteria recommended in *Numerical Recipes in C* is to stop whenever the new term in the sum is .001 of the previous term in the sum or when the new sum is 1E-8 of the previous sum Press et al..

Example

Consider the data set obtained from the following equation:

$$y' = 2x + 3 + 10\epsilon \quad (\text{E.6})$$

where ϵ is random error from a unit normal distribution. The data is essentially a line with some additive noise as illustrated in figure E.1. Now let the second set of data be the model

$$y = 2x + 3 \quad (\text{E.7})$$

This line is plotted on top of the noisy data in figure E.1. With these two data sets, it is expected that the KS statistic should be small and that the probability that a value from the KS distribution is greater than the observed statistic should be high (close to one). This would lead to the inference that the differences between the cumulative distributions of the two data sets is not significant and due only to random variation (i.e. the noise introduced). Using an IDL version of the algorithm suggested in *Numerical Recipes*, the two-sided KS statistic and probability (p-value) are $D = 0.06$ and $p = 0.992$, respectively. As expected, the statistic and probability values indicate that on the basis of this test, the model and the data have the same distribution. In other words, the model fits the data well.

E.2.3 Kolmogorov-Smirnov Regression and ISAC

The simple example in the previous section demonstrated that when a model fits the data well, the KS statistic is small and the p-value is close to one. Thus, a logical extension of KS

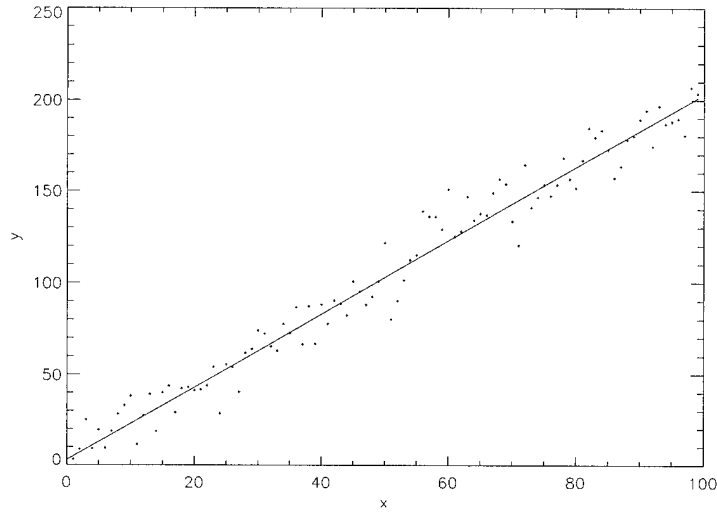


Figure E.1: Noisy data and model line plot

statistics is to implement it as a goodness-of-fit test in a regression algorithm. Ideally, such an algorithm would select the regression coefficients so that the KS statistic is minimized or the p-value maximized. This is *not* how the KS statistic is implemented in ISAC. Instead, ISAC builds the scatter plot of n observed radiance pixels ($\mathbf{y}_{n \times 1}$) vs. the estimated surface radiance ($\mathbf{x}_{n \times 1}$) and uses the KS statistic to find the points that have a distribution that is most like the Gaussian distribution. Thus, instead of comparing two data-derived cumulative distributions, ISAC compares one data-derived distribution with an “exact” analytical distribution. Actually, this is a very common implementation of the KS statistic. Many statistical software packages implement it as a test for normality by comparing the data and Gaussian cumulative distributions.

For any given bin Δx in the scatter plot, there will be a vertical spread Δy in the observations that is due to several sources of variation. Johnson and Young (1998) point out three sources: (1) surface emissivity variation, (2) surface temperature variation in the finite Δx bin, and (3) sensor noise fluctuations. Another potentially influential source

of variation is the atmosphere, particularly over wavelengths where water vapor absorption is present. The amount of variation introduced by atmospheric effects depends on the spatial heterogeneity of the atmosphere over the scene. However, one of the fundamental assumptions of the ISAC algorithm is that the atmosphere is stationary. Therefore, this error source is ignored. Of the other sources, the surface temperature variation is the easiest to compensate. For a spectral band $k \in 1 \dots p$ (where p is the number of bands), an ordinary least-squares regression may be done to estimate the slope b_1 and y-intercept b_0 of the n points. These are estimates of τ_k and L_{u_k} that *include a bias* due to errors in the surface temperature estimates. Each observation $i \in 1 \dots n$ is transformed to

$$y'_i = y_i - (b_0 + b_1 x_i) \quad (\text{E.8})$$

This removes the general linear trend of the scatter plot and the mean signal level $\bar{y}$ of each Δx bin. Thus, only variations due to sensor noise and emissivity remain (some comments and interpretations on this transformation are given in Section E.3). The emissivity variation is minimal if only the points at the top of the scatter plot are considered. In that case, the distribution of the topmost points should follow the distribution of the sensor noise. In ISAC, the noise at each band j is assumed to be Gaussian with standard deviation $\sigma = \text{NESR}$. The development is not restricted to spectrally constant NESR, but it is assumed to be so for simplicity.

Once the data are transformed, the observations are divided into N_b bins, each having N points. Each bin has a distribution of values y'_i that can be represented by a histogram. The goal is to use the points with the largest y'_i values in the analysis. These points lie on the top part of the cumulative distribution as shown in Figure E.2. ISAC_{KS} starts with the topmost pixel and builds a cumulative distribution pixel-by-pixel. At each step, the cumulative distribution of the observations is compared to the Gaussian probability distribution due to the noise. The Gaussian distribution is evaluated with the standardized

variable

$$z_i = \frac{(y'_i - y'_0)}{\sigma} \quad (\text{E.9})$$

where y'_0 is the smallest value in the selected set. The cumulative distribution of the data is based on the range between y'_0 and $\max(y'_i)$ and is calculated as shown in eq. (E.2). Because y is positive semi-definite, the z values correspond to only *half* of the Gaussian distribution (i.e., the only points considered are those with positive values, which make up only half of the distribution). Therefore, the actual distribution the data is compared to is

$$P_c(z) = -1 + \frac{2}{\sqrt{2\pi}} \int_{-\infty}^z e^{-t^2/2} dt \quad (\text{E.10})$$

where the -1 term is needed to set the probability range between 0 and 1. The D statistic is calculated at each step and the p-value recorded. As mentioned previously, the D statistic is *not* simply the maximum difference between the two distributions. The differences must be computed one at a time. Alternatively,

$$D = \max([\max |P_c(z_i) - S_n(z_i)|, \max |P_c(z_i) - S_n(z_{i+1})|]) \quad (\text{E.11})$$

where the absolute value operators are used because we are interested in deviations away from the normal distribution regardless of sign.

The calculation of the KS probability value is given in eq. (E.4), which is different than the published computation for ISAC. In this research, the ISAC_{KS} method was implemented with eq. (E.4). Once all of the pixels in the bin have been used, the list of p-values is checked, and the set of pixels that lead to the largest p-value is selected. That is, the set of topmost pixels that resulted in the lowest D statistic are the pixels at the top of the scatter with a distribution close to the sensor noise distribution. The process is repeated for the rest of the bins (N_b) in band k . This results in N_b p-values, which are then used as weights in a least-squares regression of the points that contributed to the maximum p-values.

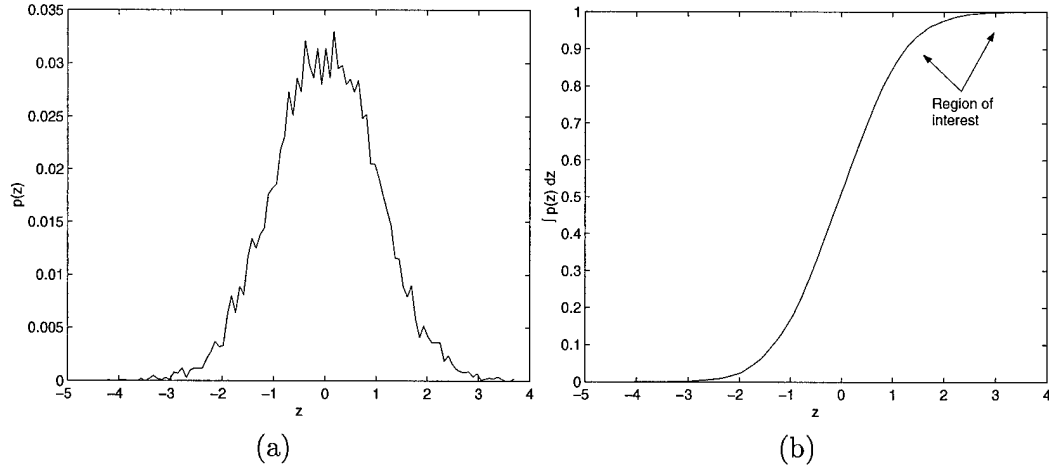


Figure E.2: (a) Histogram and (b) cumulative distribution for a hypothetical bin. The observations of interest are at the top of the distribution.

E.3 Normalized Regression

Like the Kolmogorov-Smirnov approach, the “normalized” regression attempts to find the pixels that are most likely associated with blackbody targets and have a distribution most like the normally-distributed sensor noise. However, NR avoids some of the “pitfalls” that might be encountered with KS. These include: (1) unnecessary bias introduced when least-squares parameters are correct, (2) errors due to atmospheric variations or spurious sensor response, and (3) ambiguities associated with the KS statistic. The discussion that follows expounds on these issues and develops the framework upon which NR is built.

The KS algorithm assumes that the topmost pixels will always lead to a better solution. By selecting the topmost pixels, KS introduces a bias in the least-squares estimate to compensate for observations that are due to reflective targets. However, if all of the points in a bin Δx come from blackbody targets, then only selecting the topmost points introduces an unnecessary (and erroneous!) bias. That is, in this simple case *all* of the variation is due to sensor noise and the regression line should be forced to fit through the *center* of the distribution Δy . Therefore, we need a method that will not bias the least-squares estimates

but at the same time not be affected by the presence of pixels that are not associated with blackbody targets.

Another implication of selecting only the topmost pixels is that, in certain circumstances, these pixels may actually lead to *more* error in the estimate of the regression parameters. Figure E.3(a) shows a scatter plot for SEBASS band 46 (approximately 10 μm) and a standard least-squares fit through the data. The data are observations from one of the ARM site collects (see Section E.4) that include calibration and emissivity panels. Figure E.3(b) shows the “transformed” data obtained with eq. (E.8). The transformation is simply a calculation of the residuals between the observed values and the estimated values from the regression model. The residuals are variations *unexplained* by the model. Therefore, residuals will include variations due to sensor noise, emissivity, and atmosphere. In addition, the sensor may have spurious responses that do not fall under the typical Gaussian distribution due to the NESR. Figure E.3 clearly shows that the errors may actually be larger at the top of the scatter plot. One final comment should be made about the selection of the topmost pixels. By doing this, we are concentrating on the tail of the normal distribution, which is precisely where the KS statistic is least robust (Press et al. 1992). Other statistics and tests, such as the Anderson-Darling statistic or the Wilks-Shapiro test may actually be a better measure of normality.

While KS assigns a weight to a particular bin based on the computed p-value from the KS distribution, NR attempts to keep only those pixels that appear to be normally distributed with standard deviation equal to the NESR. Figure E.4(a) shows a histogram for band 46 in the SEBASS example. While the distribution appears to be normal, it is “heavier-tailed” than it ought to be. The normal plot shown in Figure E.4(b) clearly identifies the pixels that deviate from normality. The excursion of residuals at the top of the normal plot can easily be detected and these pixels can be rejected automatically. Thus, only the pixels that have a variation on the order of the NESR are maintained. The straight line shown in the normal plot may be computed several ways. An easy approach

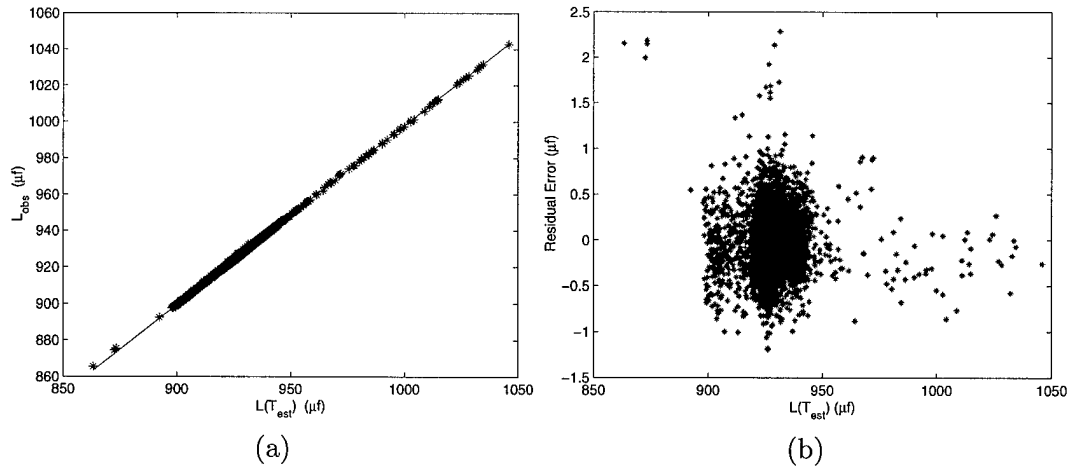


Figure E.3: Band 46 (a) Scatter plot with least-squares fit line and (b) residual error between the observed and the fitted values.

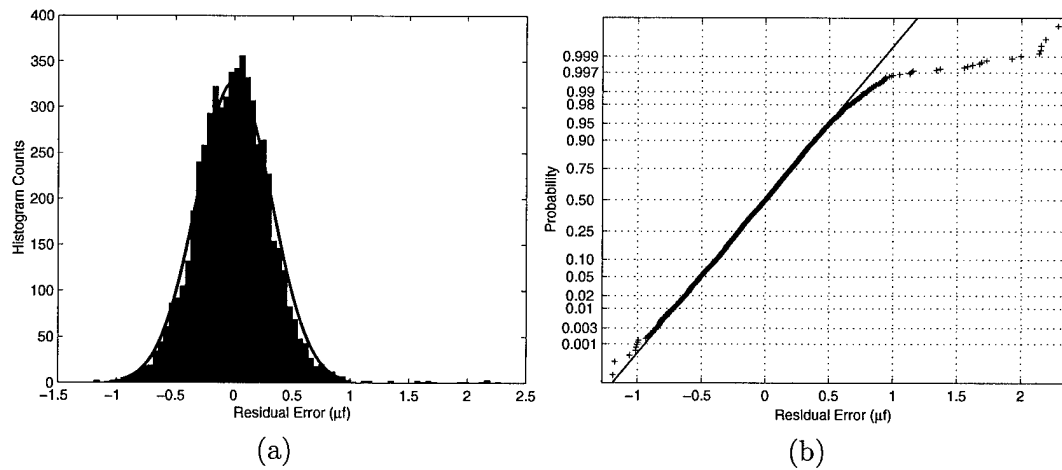


Figure E.4: Band 46 (a) Histogram of observed values with normal distribution overlay and (b) residual error normal plot showing deviation from normality

is to do a standard least-squares regression of the residuals in the normal plot. Any points that are 3σ away from the regression line are deemed to be outliers. While this may not be optimal, it avoids the calculation of a statistic and the implementation of a probability distribution for testing normality. The algorithm is also faster because the statistic and probability values do not have to be calculated iteratively as is done in KS. (The run time appears to be an issue with the KS approach; SITAC recommends the use of only 10% of the pixels in a particular bin. Unfortunately, this introduces large biased errors when the true distribution is closer to the standard regression line, in which case all of the pixels should be used in the calculation). Once the outliers have been removed, a standard least-squares regression is done again and the τ and L_u estimates recorded.

E.4 A Qualitative Validation

The following results were obtained from SEBASS data collected over the ARM site in Oklahoma on June 27th, 1997. These data were distributed with the SITAC algorithms as a test case. Flight 8, Shot 41 was collected at 0835 local standard time (LST). The altitude was 10,000 ft above sea level (ASL). The analysis was done over a segment of the image, which was a region of interest extracted from the original data cube in order to speed up processing time. The segment included the calibration panels that were placed at the ARM site. These segments were 70 samples across by 80 lines down and included all 128 spectral bands. A look at the spectral profiles in ENVI showed that there were unusual variations about the Planck spectrum. This “noise” may have been due to the present state of the atmosphere or to band-to-band sensitivity variations. In theory, the unscaled parameters cannot be used for radiometric calibration and accurate atmospheric compensation. Nevertheless, in the absence of scaled parameters, they may provide a reasonable estimate of surface temperature and emissivity. To do this, ISAC was coupled with the TES algorithm. One of the limitations of ISAC is that it does not provide an estimate of downwelled radiance. However, in certain cases, the upwelled radiance may serve as a reasonable estimate of L_d , as was done in this exercise. The ARM site calibration panel region is shown in Figure E.5.

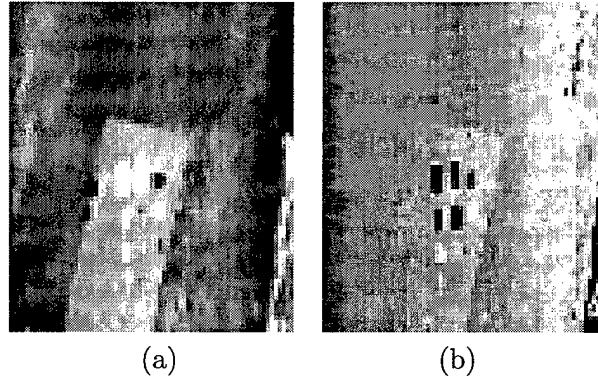


Figure E.5: (a) Temperature and (b) emissivity (band 9) maps of the ARM calibration panels. The images were generated from coupling ISAC_{KS} and TES.

Figure E.6 shows the results for the estimation of unscaled transmission and upwelled radiance with the KS algorithm. The estimates were obtained by selecting 10% of the pixels in each bin. The SITAC implementation allows “filtering” of the data via the maximum-hit method. The results shown in Figure E.6 demonstrate that the filtering can have a significant effect on the parameter estimates. In general, however, the unscaled atmospheric parameters match the MODTRAN spectra relatively well. ISAC overestimates the transmission and underestimates the upwelled radiance near the reference wavelength (i.e., band 46). This difference occurs mostly between $8.3 \mu\text{m}$ and $11.7 \mu\text{m}$ and is not uniform. This is a characteristic of the unscaled parameters because they are based on *estimated* surface temperatures, which are the sensor brightness temperatures at these wavelengths. Thus, ISAC will typically overestimate the transmission. Scaling the transmission by $\tau(\lambda_r)$ should yield a more reasonable result. The same effects are seen in the upwelled radiance as well, except reversed. The MODTRAN results were obtained with a radiosonde profile acquired in conjunction with Flight 8. The TES surface temperature retrievals for panel E1 were 297.95 and 295.98 °K for the no-filter and filter cases, respectively. The emissivity retrievals are shown in Figure E.7 and compared to a laboratory measurement of the E1 panel emissivity. The filtered KS retrieval is much smoother than the nonfiltered result. Both retrievals have a constant bias of about 0.09 emissivity units. About 0.02 emissivity units are due to

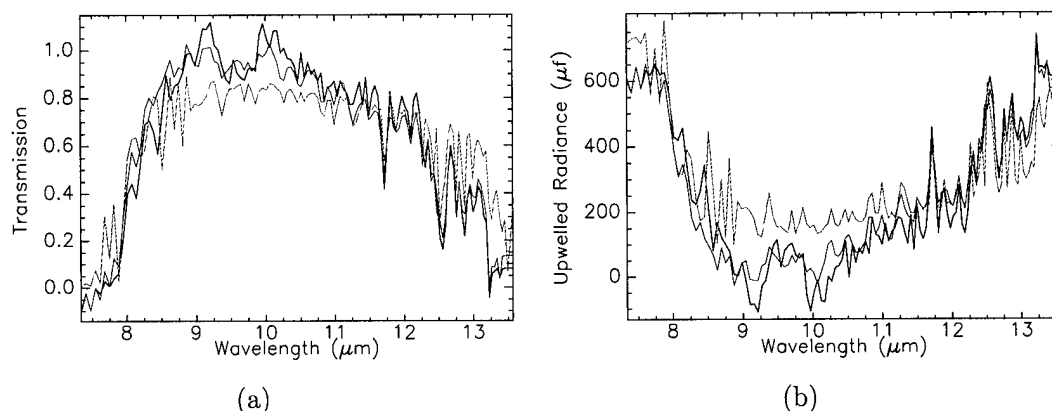


Figure E.6: ISAC_{KS} (a) transmission and (b) upwelled radiance results for no-filter (black) and filter (red) cases. The green curve is MODTRAN output with radiosonde data. 100 bins were used with NESR=0.7 μf ($\approx 3\sigma$)

bias in the TES-MMD regression line. The rest is probably due to errors in the downwelled radiance estimate. Nevertheless, the spectral shape of the curves match relatively well, particularly at the longer wavelengths. It appears that for this case, the errors in the ISAC transmission compensated for the errors in the upwelled radiance estimate. The results also show that TES is correctly accounting for the spectral structure originating from the downwelled radiance.

Figure E.8 shows the retrieved atmospheric spectra with the NR and MH approaches. The MH transmission estimates have a steep downward slope, which is manifested in the upwelled radiance as a steep upward slope. The NR solution was obtained using all pixels (except for identified outliers) and setting $\sigma = \text{NESR}$. Points beyond the 3σ variation away from the normal plot were rejected. Panel E1 temperature estimates were 297.87 and 298.49 °K for NR and MH, respectively. The emissivity retrievals are shown in Figure E.9. The NR result is very similar in shape, but has a bias of about 0.09. The same error sources affecting KS biased NR and MH estimates. Most notably, however, is the steep downward slope with increasing wavelength in the MH retrieved spectrum. This is due to errors in the transmission and upwelled radiance estimates. Increasing ΔT to 5 °K removed some of this effect.

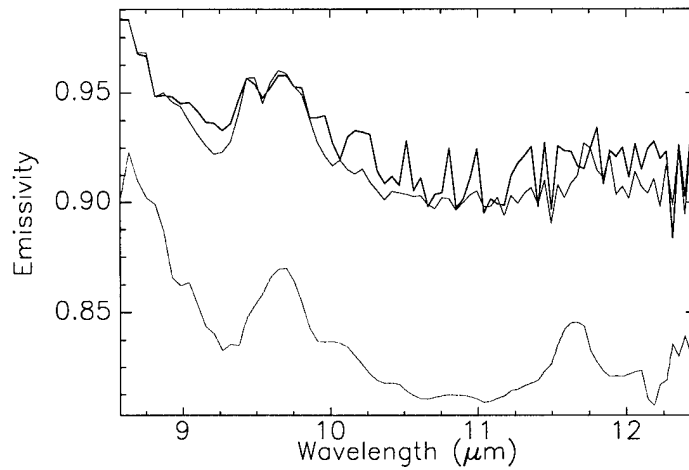


Figure E.7: ISAC and TES emissivity retrievals for panel E1 obtained with ISAC_{KS} using filter (red) and no-filter (black) settings. The green curve is the emissivity measured at the laboratory

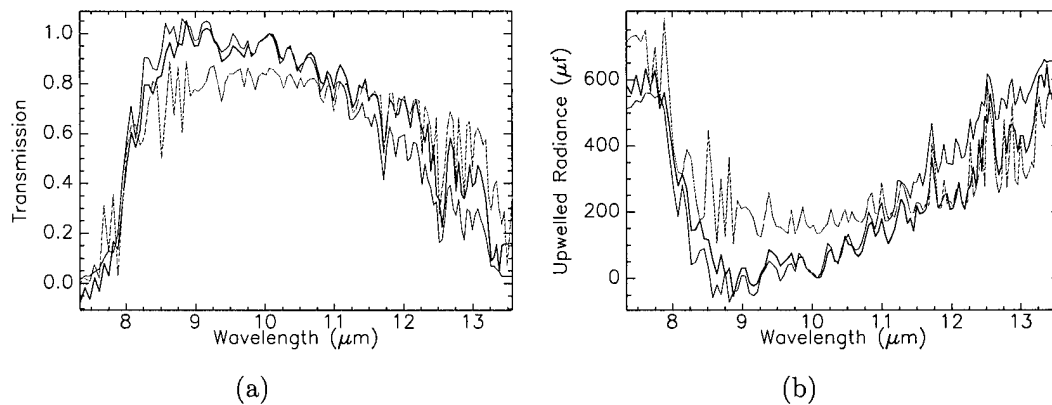


Figure E.8: ISAC_{NR} and ISAC_{MH} (a) transmission and (b) upwelled radiance results for NR (black) and MH (red) cases. The green curve is MODTRAN output with radiosonde data. The NESR was set to $0.25 \mu f$ (1σ) in NR. MH used $\Delta T = 3^\circ K$

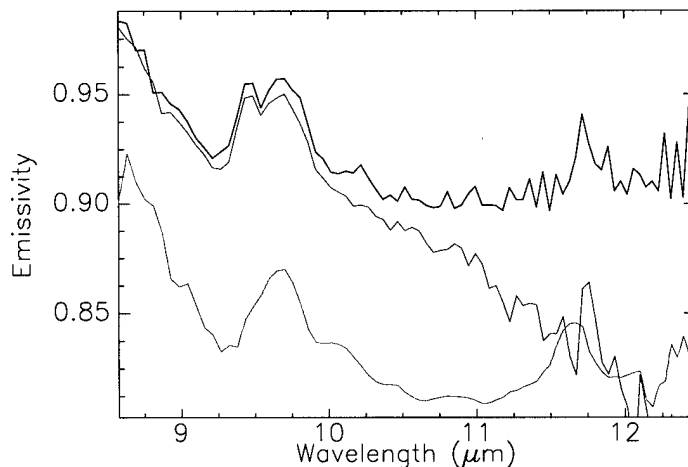


Figure E.9: ISAC_{NH} and ISAC_{MH} TES emissivity retrievals for panel E1 obtained with NR (black) and MH (red) cases. The green curve is the emissivity measured at the laboratory

Finally, Figure E.10 shows a comparison of the transmission spectra retrievals from KS and NR. They are nearly identical, as should be expected. The KS results (red) were obtained with prefiltering enabled. This prefilter step is a form of automated outlier rejection. The fact that the spectra match so well indicates that the outlier-rejection scheme used in NR correctly identifies the pixels that are not blackbodies. In general, NR is computationally more efficient than KS because it does not need to iterate on an “optimal” selection of points.

Close inspection of the atmospheric and emissivity spectra retrievals and the MODTRAN and laboratory measurements will show that there is a spectral registration error in the SEBASS-retrieved spectra, particularly at the lower wavelengths. This appears to be due to calibration errors for the ARM site collects. Because of these calibration errors, the CCR inverse model approach could not be implemented with this image set. This is because the CCR inverse model relies heavily on spectral features to characterize the atmosphere and separate atmospheric and surface emission effects. The approach implicitly assumes

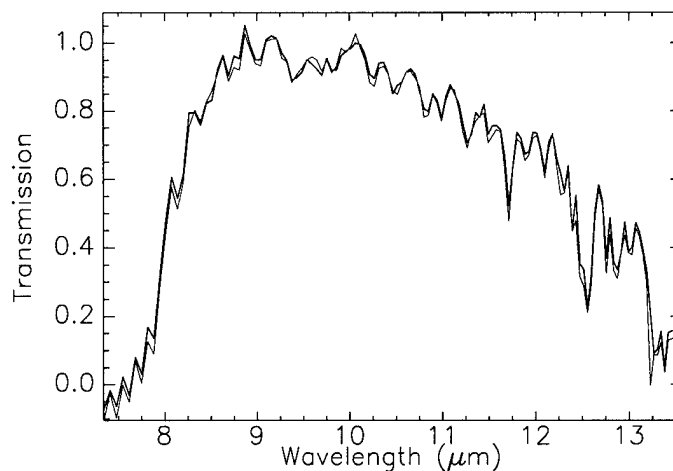


Figure E.10: Transmission spectra estimates from ISAC_NR (black) and ISAC_{KS} (red).

that the sensor is spectrally calibrated, and that the MODTRAN model spectroscopy is accurate.

The qualitative analysis presented in this section is far from rigorous. While extensive ground truth measurements were collected, none of them seemed to be reliable enough or exactly coincident with the available imagery. However, the close agreement between the different implementations provide some assurance that the algorithms were implemented correctly. The analysis illustrates that the ISAC algorithm works as it was intended. It is a relatively simple and fast approach to estimating the effects of the atmosphere (particularly the MH and NR least-squares methods). However, it is not radiometrically accurate unless some knowledge of the atmosphere exists (e.g., MODTRAN runs) that can be used to scale the estimated parameters. The use of unscaled parameters lead to a bias of about 0.07 in emissivity estimates. This error in emissivity has the potential of introducing temperature biases on the order of 2-3 °K.

Works Cited

- Akaho, S., Y. Kiuchi, and S. Umeyama. "MICA: multimodal independent component analysis." *International Joint Conference on Neural Networks (IJCNN) Proceedings* 2 (1999): 927-932.
- Anderson, T. W. *An Introduction to Multivariate Statistical Analysis*. 2 edition. Wiley Series in Probability and Mathematical Statistics. New York: John Wiley & Sons, 1984.
- Bayliss, Jessica, J. Anthony Gualtieri, and Robert F. Crompt. "Analyzing hyperspectral data with independent component analysis." *Applied Image and Pattern Recognition Workshop Proceedings, SPIE* 3240 (1997): 133-143.
- Becker, F. and Z.-L. Li. "Feasibility of Land Surface Temperature and Emissivity Determination from AVHRR Data." *Remote Sensing of the Environment* 43 (1993): 67-85.
- Becker, Suzanna. "Mutual information maximization: models of cortical self-organization." *Network: Computations in Neural Systems* 7 (1996): 7-31.
- Berk, A., G.P. Anderson, P.K. Acharya, J.H. Chetwynd, L.S. Bernstein, E.P. Shettle, M.W. Matthew, and S.M. Adler-Golden. MODTRAN User's Manual. Technical report, Air Force Research Laboratory Space Vehicles Directorate, AFMC Hanscom AFB, MA 01731-3010, June 1999.
- Better Solutions Consulting LLC, Lafayette, CO. *ENVI 3.0 Tutorials*, 1998. Internet: <http://www.envi-sw.com>.

- Borel, Christoph C. "Surface Emissivity and Temperature Retrieval for a Hyperspectral Sensor." *International Geoscience and Remote Sensing Symposium (IGARSS) Proceedings 1* (1998): 546-549.
- Caselles, Vicente, Enric Valor, César Coll, and Eva Rubio. "Thermal band selection for the PRISM instrument. 2. Analysis and comparison of existing atmospheric and emissivity correction methods for land surface temperature recovery.." *Journal of Geophysical Research* 102 (August 1997): 19,611-19,627.
- Caselles, Vicente, Enric Valor, César Coll, and Eva Rubio. "Thermal band selection for the PRISM instrument. 1. Analysis of emissivity-temperature separation algorithms." *Journal of Geophysical Research* 102 (May 1997): 11,145-11,164.
- Draper, Norman R. and Harry Smith. *Applied Regression Analysis*. Third edition. Wiley Series in Probability and Statistics. New York: John Wiley & Sons Inc., 1998.
- Geladi, Paul. "Notes on the history and nature of partial least squares (PLS) modelling.." *Journal of Chemometrics* 2 (1988): 231-246.
- Gibbons, Jean Dickinson and Subhabrata Chakraborti. *Nonparametric Statistical Inference*. 3 edition. Volume 131 of Statistics: Textbooks and Monographs. New York: Marcel Dekker, Inc., 1992.
- Gillespie, Alan R., S. Rokugawa, S. J. Hook, T. Matsunaga, and A. B. Kahle. Temperature/Emissivity Separation Algorithm Theoretical Basis Document, Version 2.3. ATBD NASA Contract NAS5-31372, NASA, August 1999.
- Goïta, Kalifa and Alain Royer. "Surface Temperature and Emissivity Separability Over Land Surface from Combined TIR and SWIR AVHRR Data." *IEEE Transactions on Geoscience and Remote Sensing* 35 (May 1997): 718-733.
- Goody, R. M. and Y. L. Yung. *Atmospheric Radiation Theoretical Basis*. 2nd edition. New York, NY: Oxford University Press, 1989.

- Green, Robert O., James E. Conel, and Dar A. Roberts. "Estimation of Aerosol Optical Depth, Pressure Elevation, Water Vapor and Calculation of Apparent Surface Reflectance from Radiance Measured by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) Using A Radiative Transfer Code." *Proceedings SPIE* 1937 (1993): 2-11.
- Green, Robert O., Dar A. Roberts, and James E. Conel. Characterization and Compensation of the Atmosphere for the Inversion of AVIRIS Calibrated Radiance to Apparent Surface Reflectance. 1996.
- Gumley, Liam. MODIS Airborne Simulator Level-1B Data User's Guide. Internet: <http://ltpwww.gsfc.nasa.gov/MAS/Home.html>, April 1994.
- Hansen, Per Christian. *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*. Philadelphia, PA: Society for Industrial & Applied Mathematics, June 1998.
- Hargreaves, J. K. *The Solar-Terrestrial Environment*. Volume 5 of Cambridge Atmospheric and Space Science. New York, NY: Cambridge University Press, 1992.
- Harsanyi, J.C. and C.-I Chang. "Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection." *IEEE Transactions on Geoscience and Remote Sensing* 32 (July 1994): 779-785.
- Hayden, Christopher M. "GOES-VAS Simultaneous Temperature-Moisture Retrieval Algorithm." *Journal of Applied Meteorology* 27 (1998): 705-733.
- Helland, Inge S. "On the structure of partial least squares regression." *Communications in Statistics-Simulation and Computation* 17 (1988): 581-607.
- HITRAN. June. 2000 "HITRAN Internet web site." <http://www.hitran.com>.
- Hook, Simon J., Jeffrey J. Myers, Kurtis J. Thome, Michael Fitzgerald, and Anne B. Kahle. "The MODIS/ASTER Airborne Simulator (MASTER)-A New Instrument for Earth Science Studies." *Submitted to Remote Sensing of Environment* (2000).

- Höskuldsson, Agnar. "PLS Regression Methods." *Journal of Chemometrics* 2 (1988): 211–228.
- Houghton, J. T. *Remote Sounding of Atmospheres*. Volume 3 of Cambridge Planetary Science Series. New York, NY: Cambridge University Press, 1984.
- Houghton, J. T. and S. D. Smith. *Infra-red Physics*. London, England: Oxford at the Clarendon Press, 1966.
- Huang, Allen Hung-Lung. *An analysis of the characteristics of atmospheric profiles obtained with the High-Resolution Interferometer Sounder (HIS)*. Ph. D. dissertation, University of Wisconsin-Madison, 1989.
- Hunter, Gordon B., C. D. Allemand, and T. W. Eagar. "Multiwavelength Pyrometry: An Improved Method." *Optical Engineering* 24 (1985): 1081–1085.
- Hyvärinen, Aapo. "Survey on Independent Component Analysis." *Neural Computing Surveys* 2 (1999): 94–128.
- Izenman, Alan Julian. "Reduced-Rank Regression for the Multivariate Linear Model." *Journal of Multivariate Analysis* 5 (1975): 248–264.
- Jackson, J. Edward. *A User's Guide to Principal Components*. Wiley Series in Probability and Mathematical Statistics. New York: John Wiley & Sons, Inc., 1991.
- Johnson, B. R. and Stephen J. Young. Inscene Atmospheric Compensation: Application to SEBASS Data Collected at the ARM Site. Technical report, Aerospace Corporation, May 1998.
- Johnson, Richard A. and Dean W. Wichern. *Applied Multivariate Statistical Analysis*. Third edition. Englewood Cliffs, NJ: Prentice-Hall, 1992.
- Jun, Li. "Temperature and Water Vapor Weighting Functions from Radiative Transfer Equation with Surface Emissivity and Solar Reflectivity." *Advances in Atmospheric Sciences* 11 (November 1994): 421–426.

- Kaplan, Lewis D. "Inference of Atmospheric Structure from Remote Radiation Measurements." *Journal of Optical Society of America* 49 (October 1959): 1004–1007.
- Kealy, Peter S. and Simon J. Hook. "Separating Temperature and Emissivity in Thermal Infrared Multispectral Scanner Data: Implications for Recovering Land Surface Temperatures." *IEEE Transactions of Geoscience and Remote Sensing* 31 (November 1993): 1155–1164.
- Keeling, C.D. and T.P. Whorf. July. 1999 "Atmospheric CO₂ concentrations (ppmv) derived from in situ air samples collected at Mauna Loa Observatory, Hawaii." . University of California.
- King, Jean I. F. "The Radiative Heat Transfer of Planet Earth." *Scientific Uses of Earth Satellites*. . University of Michigan Press, 1956. chapter 14, 133–136.
- King, Michael D. and Reynold Greenstone. 1999 EOS Reference Handbook. Technical report, NASA/Goddard Space Flight Center, 1999. <http://eospso.gsfc.nasa.gov>.
- King, Michael. D., W. P. Menzel, P. S. Grant, J. S. Myers, G. T. Arnold, S. E. Platnick, L. E. Gumley, S. C. Tsay, C. C. Moeller, M. Fitzgerald, K. S. Brown, and F. G. Osterwisch. "Airborne scanning spectrometer for remote sensing of cloud, aerosol, water vapor and surface properties." *J. Atmos. Oceanic Technol.* 13 (1996): 777–794.
- Kneizys, F.X., D.C. Robertson, L.W. Abreu, P. Acharya, G.P. Anderson, L.S. Rothman, J.H. Chetwynd J.E.A. Selby, E.P. Shettle, W.O. Gallery, A. Berk, S.A. Clough, and L.S. Bernstein. The MODTRAN 2/3 Report and LOWTRAN 7 MODEL. Technical Report F19628-91-C-0132, Phillips Laboratory, Geophysics Directorate, U.S. Air Force, Hanscom AFB, MA, November 1996.
- Kullback, Solomon. *Information Theory and Statistics*. 3 edition. Mineola, New York: Dover Publications, Inc., 1997.
- Lai, P. L. and C. Fyfe. "A neural implementation of canonical correlation analysis." *Neural Networks* 12 (1999): 1391–1397.

- Liang, Shunlin. "A New Algorithm for Retrieving Land Surface Temperature and Emis-
sivity And Applications to Airborne Hyperspectral SEBASS Data." *International
Geoscience and Remote Sensing Symposium (IGARSS) Proceedings* 1 (1998): 550-
552.
- Liou, Kuo-Nan. *An Introduction to Atmospheric Radiation*. Volume 26 of International
Geophysics Series. San Diego, CA: Academic Press, Inc., 1980.
- List, Robert J. *Smithsonian Meteorological Tables*. 6 edition. Volume 114 of Smithsonian
Miscellaneous Collections. Washington D.C.: Smithsonian Institute, 1951.
- Menzel, W. Paul and Liam E. Gumley. MODIS Atmospheric Profile Retrieval Algorithm
Theoretical Basis Document. Nasa atbd, University of Wisconsin-Madison, 1225 W.
Dayton St. Madison, WI 53706, November 1999.
- Merola, Giovanni Maria. *Dimensionality Reduction Methods in Multivariate Prediction*.
Ph.D, University of Waterloo, Waterloo, Ontario, Canada, 1998.
- Nielsen, Allan Aasbjerg. *Analysis of Regularly and Irregularly Sampled Spatial, Multivari-
ate, and Multi-temporal Data*. Ph.D., Institute of Mathematical Modelling, University
of Denmark, 1995.
- Palluconi, Frank D. Email. January 2000.
- Phatak, Alope. *Evaluation of Some Multivariate Methods and their Applications in Chem-
ical Engineering*. Ph.D. thesis, University of Waterloo, Waterloo, Ontario, Canada,
1993.
- Pieters, Carl M. and Peter A. J. Englert. *Remote Geophysical Analysis: Elemental and
Mineralogical Composition*. Volume 4 of Topics in Remote Sensing. New York, NY:
Cambridge University Press, 1993.
- Prata, A. J., V. Caselles, C. Coll, J. A. Sobrino, and C. Ottlé. "Thermal Remote Sensing
of Land Surface Temperature From Satellites: Current Status and Future Prospects."
Remote Sensing Reviews 12 (1995): 175-224.

- Press, William H., Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes in C: The Art of Scientific Computing*. Second edition. New York, NY: Cambridge University Press, 1992.
- Rodgers, Clive D. Inverse Methods for Atmospheric Sounding: Theory and Practise (Draft). February 1998.
- Sabol, Donald E., John B. Adams, and Milton O. Smith. "Quantitative Subpixel Spectral Detection of Targets in Multispectral Images." *Journal of Geophysical Research* 97 (February 1992): 2659-2672.
- Sanders, Lee C. *An Atmospheric Correction Algorithm for Hyperspectral Imagery*. Ph.D dissertation, Rochester Institute of Technology, Chester F. Carlson Center for Imaging Science, Rochester, NY, September 1999.
- Saucier, Walter J. *Principles of Meteorological Analysis*. New York, NY: Dover Publications, Inc., 1989.
- Schmugge, Thomas, Simon Hook, and César Coll. "Application of the TES Algorithm to TIMS Data Acquired in HAPEX-Sahel." *International Geoscience and Remote Sensing Symposium (IGARSS) Proceedings* 3 (August 1997): 1233-1235.
- Schott, John R. *Remote Sensing: The Image Chain Approach*. New York, NY: Oxford University Press, 1997.
- Seinfeld, John H. and Spyros N. Pandis. *Atmospheric Chemistry and Physics*. New York: John Wiley & Sons, Inc., 1998.
- Shannon, Claude E. "The mathematical theory of communication." *M.D. Computing* 14 (July-August 1997): 306-317.
- Smith, Frederick G. "Atmospheric Propagation of Radiation." *The Infrared & Electro-Optical Systems Handbook*. Ed. Joseph S. Accetta and David L. Shumaker. Volume 2 . SPIE Optical Engineering Press and ERIM, 1993.

- Smith, William L., Harold M. Woolf, and Henry E. Revercomb. "Linear Simultaneous Solution for Temperature and Absorbing Constituent Profiles from Radiance Spectra." *Applied Optics* 30 (1991): 1117-1123.
- Stephens, Graeme L. *Remote Sensing of the Lower Atmosphere: An Introduction*. New York, NY: Oxford University Press, 1994.
- Stone, M. and R. J. Brooks. "Continuum Regression: Cross-validated Sequentially Constructed Prediction Embracing Ordinary Least Squares, Partial Least Squares and Principal Components Regression." With Discussion. *Journal of the Royal Statistical Society. Series B.* 52 (1990): 237-269.
- Suplee, Curt. "El Niño/La Niña: Nature's Vicious Cycle." *National Geographic* (March 1999).
- Tang, Benyang, William W. Hsieh, Adam H. Monahan, and Fredolin T. Tangang. "Skill Comparisons between Neural Networks and Canonical Correlation Analysis in Predicting the Equatorial Pacific Sea Surface Temperatures.." *Journal of Climate* 13 (January 2000): 287-293.
- Trefethen, Lloyd N. and David Bau III. *Numerical Linear Algebra*. Philadelphia, PA: Society for Industrial & Applied Mathematics, 1997.
- Tu, Te-Ming. "Unsupervised signature extraction and separation in hyperspectral images: a noise-adjusted fast independent component analysis approach.." *Optical Engineering* 39 (April 2000): 897-906.
- Twomey, S. *Introduction to the Mathematics of Inversion in Remote Sensing and Indirect Measurements*. Mineola, NY: Dover Publications, Inc., 1996.
- Wan, Zhengming. MODIS Land Surface Temperature Algorithm Theoretical Basis Document. NASA ATBD 3.3, University of California, Santa Barbara, April 1999.
- Wan, Zhengming and Zhao-Liang Li. "A Physics-Based Algorithm for Retrieving Land-Surface Emissivity and Temperature from EOS/MODIS Data." *IEEE Transactions*

- on Geoscience and Remote Sensing* 35 (1997): 980–996.
- Wold, Svante. "PLS Regresion." *Encyclopedia of Statistical Sciences*. Ed. N. L. Johnson and S. Kotz. Volume 6 . John Wiley & Sons, 1984. 581–591.
- Yaglom, A. M. "Method of Canonical Correlations and its Applications in Meteorology." *Izvestiya, Atmospheric and Oceanic Physics* 26 (June 1990): 909–922.
- Yu, Zhi-Ping, Pao-Shin Chu, and Thomas Schroeder. "Predictive Skills of seasonal to annual rainfall variations in the U.S. affiliated Pacific Islands: canonical correlation analysis and multivariate principal component regression appraoches.." *Journal of Climate* 10 (1997): 2586–2599.