Estimation and Modeling of Multidimensional Non-Stationary Stochastic Processes: Application to the Remote Sensing of Atmospheric Temperature Fields.

Vol. 1

by

Alain Charles Louis Briançon

SM, EE, Massachusetts Institute of Technology, 1983

Ingénieur, Ecole Supérieure d'Electricité, 1982

Submitted to the Department of
Electrical Engineering and Computer Science
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

August 1986

Signature of Author: _____
Department of Electrical Engineering and Computer Science
August 27, 1986

Certified by: _____
David H. Staelin
Thesis Supervisor

Accepted by: _____
Arthur Smith
Chairman, Department Committee

Estimation and Modeling of Multidimensional Non-Stationary Stochastic Processes: Application to the Remote Sensing of Atmospheric Temperature Fields.

by
Alain Charles Louis Briançon

Submitted to the Department of Electrical Engineering and Computer Science on August 27, 1986, in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in Electrical Engineering and Computer Science.

Abstract

The satellite remote sensing of atmospheric temperature is analyzed in the context of Bayesian multichannel multidimensional linear filtering. Retrievals are performed by combining radiances measured at neighboring microwave frequencies and adjacent locations, thus benefiting from the known three-dimensional correlation structure of atmospheric temperature fields.

A model representing the three-dimensional covariance of temperature from the ground up to 10 *mbar* and horizontal displacement up to 3000 *km* is developed using theoretical and empirical results from geofluid dynamics and atmospheric science. The model is fitted to and validated by National Meteorological Center observational radiosonde data.

Single- and multi-spot retrieval operators optimal in the mean square sense are examined in the natural representation of temperature profiles or by projecting them along the row and kernel subspaces associated with the weighting functions relating radiances with temperature. Lower performance bounds for the retrieval error, limiting expressions for operators, and improvement provided by the inclusion of horizontal correlation are analytically and heuristically examined: Multi-spot operators improve error by low-pass filtering of the sensor noise for the row subspace, and use vertical variations of the horizontal correlation in conjunction with differences in the weighting function to reduce error in the kernel space. The impact of horizontal correlation is measured *vis à vis* the number of channels of sounding systems and the sampling grid resolution. A novel interpretation of the design equation and the corresponding solutions for operators of finite and infinite support as well as Kalman based filters are presented. These operators implemented on the three-channel Microwave Sounding Unit (MSU) provide typically 15 to 25 percent improvement in rms over single spot operators. Simulations for the twelve-channel Advanced Microwave Sounding Unit (AMSU) suggests from 5 to 10 percent improvement.

Adaptation of retrieval operators to non-stationarities is considered in the context of single-spot operators. The optimal adaptation criterion from a retrieval

mean square error point of view is found to be the Euclidian norm between the operator optimum for a given dataset and the operator used over the dataset. This criterion can be approximated by the Euclidian norm between corresponding vertical covariance kernels. This yields adaptation operators which minimize the penalty on a dataset basis (Fisher) or globally (Bayesian). These adaptations are found inefficient for MSU but appropriate for AMSU. Adaptation of the mean profile on a localized scale is found to be successful for MSU.

Also considered are operators optimized for criteria derived from the fields of information theory and constrained estimation. One estimator minimizes the expected maximum square error. All these estimators are found to exhibit similarities with the linear minimum mean square estimator.

Thesis Supervisor: David H. Staelin

Title: Professor of Electrical Engineering

My Declaration of Acknowledgment

When in the course of human events, it becomes necessary for one individual to dissolve the academic bonds which have connected him with an institution of higher education, and to assume among the members of society the separate and equal station to which the laws of Nature's God entitle them, a decent respect to the opinions of mankind requires that he should declare the support which helped him to the separation.

We hold these truths to be self-evident, that all Men are created equal, that they are endowed by their creator with certain unalienable rights, that among these are Life, Liberty and the pursuit of Happiness...

...The history of the present dissertation is a history of repeated efforts and commitments, all having in direct object the establishment of a document helping science. To prove this, let facts be submitted to a candid world.

Professor Staelin supervised this work and made it so much better through his ideas and critiscisms. Without his guidance, no one knows what this research and document would look like. His intuition never failed and his dedication to this research was total. In addition to being my academic advisor, I count him as a friend who helped whenever possible, with my military service, with professional choices, with moral support, and with goodies for the baby,

Doctor Rosenkranz, who suffered through numerous exposés of my ideas (the good ones and the bad ones) was essential to the realization of this dissertation. Phil always closely examined the theoretical aspects of the research, pointed to appropriate references and reports, and maintained the NOVA computer system on which most of this work was performed. I will miss his "Force Tranquille",

Professor Schweppe, who served as a reader, got me interested in the entire issue of modeling and system identification through his course on the subject. His class was one of the bests I took at MIT. He also read the entire document at a speed approaching that of light,

My collegues Mark Colavita, Gerard DeJager, Al Gasiewski, Henrique Malvar, Jerome Shapiro, Bernard Szabo, and Gregory Wornell all helped with various theoretical aspects of this research, and contributed their expertice and points of views to the signal processing aspects of this endeavor. Moreover, Mark proofread numerous reports and proposals now scattered among this dissertation, Gregory proofread chapter VII, Henrique provided some of this document's figures, and Bernard helped me administrate the group's APOLLO network when my working load topped my working capacity,

Ashok Popat and Kris Pandava convinced me to learn TEX to produce this

document and got me hoocked on hacking. Kris sold me the VT100 which allowed me to stay at home evenings typing and watching my daughter grow up rather than having to go to an inhumane terminal cluster. Ashok was the best of officemates,

Heather Porter joined this document's editing staff by learning about my writing abilities in chapters I and VIII and helped to correct them. Cathy McCue was always available to hear my complaints and endless dithyrambes,

All members of the Radio-Astronomy group always helped and provided an optimal working environment, their help and friendship during the trying times of this thesis and Maria's pregnancies are gratefully acknowledged,

CGA@OZ, MBR@NEWT, GREG@XX, JP@ATRP, and CSTACY@MX all contributed to the production of this document by providing me with access to the required hardware and software,

My family and my in-laws encouraged me during my studies at MIT, forgave me numerous times for not writing, calling or visiting, and helped whenever possible. Thanksgiving dinner at Manchester, Serbian Christmas in Wenham, Week-ends in New-Hampshire, and Mariage in Auvillers were the best time of year. Maman, merci encore mille fois d'être venue pour le bébé,

Finally, my wife Maria proofread the entire document and learned about SNR, MSE, and other LMMSE, accepted being a thesis widow, gave birth to our wondergirl Alicia, and sacrificed many political engagements and nights of sleep so that I could finish this work. When I was about to give up on remote sensing, she showed me the light. This document should have her name as a co-author. Maria, I love you, you are the best!

... And for the support of this Declaration, with a firm reliance on the protection of divine Providence, I gratefully thank all those who gave their Lives, their Fortunes, their Knowledge and their sacred Honors to this endeavor.

A Note on the Type

The text of this thesis was set in a digitized version of a typeface called Times Roman, designed by Stanley Morison (1889-1967) for *The Times* (London) and first introduced by that newspaper in 1932.

Among typographers and designers of the twentieth century, Stanley Morison was a strong forming influence as a typographical advisor to The Monotype Corporation, as a director of two distinguished English publishing houses, and as a writer of sensibility, erudition, and keen practical sense.

After J. Anthony Lukas

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I
INTRODUCTION

"I will go far away, alone
like the echo of the hallowed bell,
up there amid the white snows;
I'll go alone and far away
up among the golden clouds!
But my foot is firm!
Let us go for the way is long.
let us set off."
Alfredo Catalani: La Wally, "Ebben? ne andro lontana"

I.1 Importance of Weather Monitoring by Satellite

And the flood was forty days upon the Earth; and the waters increased. GENESIS 8.17. Ever since that time, the monitoring and prediction of weather has been important to mankind's development. Among all the aspects of weather, temperature is one of the most important for it regulates most human activities. Until the 1960's, the monitoring of temperature could only be performed above land areas by occasional aircrafts or by using radiosondes attached to balloons and launched at regular intervals (typically 12 hours). This caused a problem for numerical weather forecasting based on such data because oceans play an important role in the evolution of weather patterns.

Satellites have the unique ability to measure atmospheric and terrestrial geoparameters around the globe in a continuous fashion, regardless of the exact location. These satellite can be equipped with optical and infrared sensors to characterize the ground, and in the case of interest to this dissertation, microwave sensors (also called radiometers) to characterize the atmosphere. Unlike radiosondes, satellite radiometers do not measure atmospheric temperature directly, rather they sample the upwelling thermally excited radiation spectrum emitted by the

earth and its atmosphere at appropriate wavelengths. By carefully choosing these wavelengths, one is able to measure radiation emitted by different regions of the atmosphere. However, these regions are relatively wide (typically on the order of miles). Moreover these measurements are corrupted by noise. They require extensive processing and the inclusion of auxiliary information about the temperatures being estimated to provide useful results about the temperature at specific altitudes. Reviews of methods used to invert radiance measurements can be found in Staelin (1982) and Fymat (1977).

In the context of this dissertation, this auxiliary information is statistical and is introduced through means and variances of temperature fields which requires the modeling of atmospheric temperature as a random process. In the later part of this document, bound information is also incorporated in the inversion process.

This extraction of information from the measurements is performed by the linear weighting of neighboring measurements (both in space and frequency). In that context, microwave remote-sensing is the application of multi-channel multi-dimensional filtering to a specific form of signal and statistics, and one can use the results developed by this important field of signal processing.

I.2 Purpose and Organization of this Dissertation

The main purpose of this dissertation, is to characterize, as thoroughly as possible, the impact of statistics, of non-stationarities in the statistics, of sounder characteristics, and of optimization criterion on the remote-sensing of atmospheric kinetic temperature from satellite-borne radiance measurements. This impact will be characterized from both an analytical and an heuristic point of view. As noted, the intended processing of the data is linear, thus only first and second order statistics are required to develop estimators and characterize retrieval errors. In this thesis, the impact of both statistical moments is investigated in both the horizontal and vertical directions.

The thesis is organized as follows:

Chapter II briefly examines the physics behind the fundamental sensing equation in the microwave region relating the measurements performed by sounding platforms to the vertical distribution of kinetic temperature and other relevant geophysical parameters.

Chapter III develops the most important element in terms of the computation of three-dimensional retrieval operators (these dimensions pertain to the sounding grid and the frequencies): a model describing the cross-covariance between temperatures at two different points in the atmosphere. This model is based on both mathematical and physical considerations, with a special emphasis on the physics. This is to ensure than (1) the model is valid from a physical point of view, and (2) that it can be modified to accommodate future improvements. Relevant theoretical and empirical knowledge derived from the field of atmospheric science are incorporated in both the vertical and horizontal directions. This model provides the means to discretize temperature profiles in the vertical direction, which is instrumental in the design of retrieval algorithms.

Chapter IV presents the mechanics associated with the model development: namely its fitting to observed measurement and its validity testing. In a later section of the chapter, applications of the model to atmospheric science (other than remote-sensing) are considered. Some of these applications will have direct bearing on the analysis of multidimensional operators. To avoid using corrupted data, direct measurements of temperature are used in this fitting. These measurements are incomplete, and vary in quality and extent from sample to sample. In order to optimally use these measurements, a novel estimate for mean and covariance based on the principle of maximum likelihood will be developed and analytically analyzed.

Chapter V provides the core of the mathematical analysis of the statistical impact. Linear retrieval operators, optimal in a minimum mean square error sense

(LMMSE), are derived. The most important tool in this analysis (performed in the Bayesian framework for the most part) is the projection of the sensing and retrieval processes into two of the observation operators' main vectorial subspaces; row space (also called "observable" space) and kernel ("unobservable") space. In this framework, the retrieval can be recast as a two-step estimator, different lower bounds on the performances of operators can be computed, different limiting forms as the statistics vary are obtained, and most notably, different behaviors relative to the inclusions of horizontal statistics may be extracted. The parameters controlling the impact of horizontal statistics on the retrieval process are accessed from an analytical and quantitative point of view. In an important appendix to this chapter, past implementations of retrieval algorithms based on variations of Kalman-Bucy filtering which were proven successful from an heuristic point of view are analytically examined, and the performances the the algorithms are confirmed. This chapter also provides the foundation for the efficient design of retrieval operators by analyzing this design equation as a generalized eigenvalue problem.

Chapter VI considers the retrieval operators derived from the previous chapters, re-analyses their design equation from a computational point of view, and implements single-spot and multi-spot operators on synthetic data (measurements computed from the ground truth), and then on real data (measurements performed by a low-orbit polar satellite). These results are presented for the operational three-channel Microwave Sounding Unit (MSU) and the proposed twelve-channel Advanced MSU (AMSU). These analyses are performed in three different representation bases for the temperatures profiles: one "user-oriented", one "signal-oriented", and the most important one "system-oriented". The differences between the multidimensional retrievals for MSU and AMSU are considered.

Chapter VII considers the impact of inaccurate statistics on the retrieval process. This problem is inherent in remote sensing because of the seasonal and

regional trends in temperature which translate into non-stationarities in the temperature time series. The first analysis develops tools to allow the model for the covariance (as well as the retrieval operators derived from it) to be tuned and characterized for non-stationarities. A novel quantitative characterization of the specificity of climatologies is developed for this purpose. The adaptation (or correction) criterion for multidimensional retrieval operators is derived and examined in the Fisher and Bayesian frameworks, for both observable and unobservable subspaces. Sub-optimal analytical adaptation operators for mean and covariance are derived for different types of a-priori information. These operators are implemented for MSU and AMSU. The impact of the number of channels, observational sounding angle, and climatologies are assessed for the adaptation of first and second moment statistics.

Chapter VIII complements the estimation of temperature profiles using linear processing of the data by changing the error criteria for which the processing is optimized. The sounding equation is recast as a communication problem through a noisy channel and the concepts of entropy and mutual information are used to characterize sounders. The retrieval operators are also optimized using maximum-likelihood and maximum power criteria. In the later part of chapter VIII, the use of bounds on the derivative of the profile or on its maximum range of values are considered to yield new operators. The inclusions of constraints on the retrieval process is relevant when these constraints are easy to obtain and incorporate.

Chapter IX summarizes the main results of this thesis and points to possible improvements and extensions of this work.

This document is by all standards long, way too long. This was not my intention when I started typing it in more than 18 months ago. Because of the scope and length of the dissertation, it is at times difficult to see the forest for the trees, that is to distinguish the train of thought from the arguments. I am sorry

for that, but have decided to err in the direction of too much information. Section IX.2 presents the main ideas investigated in this work and serves as a guide to this thesis

CHAPTER II

PRINCIPLES OF RADIATIVE TRANSFER

"The Proof is complete, if only I've said it thrice." Lewis Carrol

II.1 Introduction

In this chapter, the principles of radiative transfer are briefly exposed. The derivation of the so called forward problem, which consists of the computation of the observations from the relevant geophysical temperature, is not novel and is presented here merely for the sake of completeness. After sketching the computation of the absorption coefficient for O_2, the sounding equation and equivalent weighting functions are derived. The impact of ground effects is also discussed, with most of the emphasis on the reflection properties of sea versus land. Finally, a novel analysis of the impact of non-linearities in the sounding process and a method to correct for them is presented in the appendix of this chapter.

Temperature remote sensing platforms (SCAMS, NEMS, MSU, DMSP) use the 60 GHz resonance line of oxygen to infer the vertical distribution of temperature. Other atmosphere absorbants, N_2, CO_2 and more importantly H_2O in liquid and gaseous form, can be used for remote sensing, however they are not relevant to the estimation of atmospheric temperature through observation of the radiated spectrum around 60 GHz. For more details, the interested reader is referred to chapters III, IV, and V in the American Society of Photogrammetry Manual of Remote Sensing (1983) and references therein as well as the special June 1985 issue of *Proceedings of the IEEE* on "Perceiving Earth's Resources from Space".

II.2 Oxygen Absorption

The presentation of this dissertation follows closely the one of Rosenkranz (1975).

In the microwave spectrum, the absorption line of diatomic oxygen arises from transitions of the angular momentum of the molecule. The angular momentum is the sum of the molecule angular momentum and the spins of two unpaired electrons. The resonant frequencies are directly proportional to the difference between the original and final states' energies (Wilheit and Barrett, 1971).

The absorption coefficient γ for $^{16}O_2$ can be expressed as the product (Rosenkranz, 1971)

$$\gamma = 1.434 \frac{P^2}{T^2} \nu^2 F \qquad (dB/km) \qquad (II.2.1)$$

where T denotes the temperature and P the pressure, ν the operating frequency and F is the shape factor. This shape factor can be written as the sum (Rosenkranz, 1975)

$$F = \sum_N \Phi_N \left(f_N^+(\nu) + f_N^-(\nu) + f_N^+(-\nu) + f_N^-(-\nu) \right) + \frac{0.70 w_b}{\nu^2 + P^2 w_b^2} \qquad (II.2.2)$$

This summation considered six branches of line spectrum: The branches noted f_N^+ and f_N^- corresponds to transitions with increasing or decreasing total angular momentum respectively The right-most term in the summation is a weak branch at zero frequency which accounts for nonresonant absorption. This line spectrum is symmetric with respect to zero and lines are reflected at negative frequencies (Chahine *et al.*, 1983).

The term w_b has the unit of $Hz/mbar$ and controls the linewidth's dependence of the nonresonant absorption on the pressure. The linewidth decreases with temperature. Each resonant line in the above summation is a function of the form

$$f_N^{\pm}(\nu) = \frac{d_{N\pm}^2 w_N + (\nu - \nu_N^{\pm}) y_N^{\pm}}{(\nu - \nu_N^{\pm})^2 + P^2 w_N^2} \qquad (II.2.3)$$

This lineshape correspond to two poles at $P^2 w_N^2 \pm j\nu_N$ and two zeroes at $\nu_N^{\pm} - d_{N\pm} w_N / y_{N\pm}$ and at infinity. As the pressure decreases, the poles departs from the

imaginary axis, the resonance becomes sharper and the absorption becomes weaker. This lineshape corresponds to a collisional or pressure broadening of the resonance and a temporal correlation in the form of decaying exponentials. The frequency ν_N is the N^{th} oxygen line resonance frequency. These frequencies are in the range 50 to 60 GHz with the exception of a single line at 118.7503 GHz, plus their negative counterparts. Wilheit and Barrett (1970) provide a list of the resonance frequencies. $d_{N\pm}$ is the amplitude of the line and solely depends on the quantum number N. At 300 K, the line with the largest strength and has for resonance frequency 60.4348 GHz. The overall absorption will also peak around that frequency.

Similar to the nonresonant absorption, the quantity Pw_N controls the line width of the resonance. This quantity is inversely proportional to the mean time between the collisions that broaden the line. The quantity $y_n^{\pm}$ is called the interference factor; it accounts for the correlation between the states as characterized by the angular momentum quantum number N of the molecule before and after collision. A non-zero interference factor yields an asymmetric lineshape. Tables of values for $y_N^{\pm}, w_N^{\pm}$ and $\nu_N^{\pm}$ as well as a more detailed analysis of oxygen absorption can be found in Rosenkranz (1971).

Because of its permanent magnetic moment, the oxygen molecule interacts with the terrestrial DC magnetic field and the resonance lines are subjected to Zeeman splitting (Lenoir, 1968). However, this splitting is apparent only for small resonance halfwidths, that is for low pressures. This splitting becomes significant at high altitudes ($h \geq 40km$). This altitude is beyond the sounding capabilities of current sounding systems but within the range of proposed sounding platforms. At higher altitudes ($h \geq 70km$), Doppler broadening of the line, which produces Gaussian lineshapes, occurs. Zeeman splitting and Doppler broadening will not be considered in this dissertation.

Other than oxygen, water plays a mild role in the microwave radiative bud-

get. Water vapor has two resonances, 22.235 and 183.3 GHz, which change the tropospheric attenuation near those frequencies. Liquid water in droplet form in clouds is the most important non oxygen absorbant. In this dissertation, clouds are treated as a disturbance and contribute extra absorption. Staelin (1966) provide an empirical fit to the absorption of droplet γ_d valid up to 80 GHz and small drop-size (Rayleigh scattering)

$$\gamma_d = \frac{M}{\lambda^2} 10^{0.0122(291-T)-6} \qquad (II.2.4)$$

where M is the mass of water per unit volume of air and λ the operating wavelength.

II.3 Radiative Transfer: From Planck's Law to the Sounding Equation.

All matter emits energy in the form of electromagnetic radiation because of the motion of its charged particles. The so called blackbody radiation has the remarkable property that it is solely a function of the temperature T of the body and is independent of the specific distribution of energy and particles in the body (Bekefi and Barrett, 1977)

The amount of radiation emitted at wavelength λ when the body's temperature is T follows Planck's law

$$E(\lambda, T) = \frac{2hc^3}{\lambda^3 (e^{\frac{hc}{k\lambda T}} - 1)} \qquad (II.3.1)$$

where c is the velocity of light, k the Boltzmann's constant ($k = 1.3806 \times 10^{-23} JK^{-1}$), and h the Planck's constant ($h = 6.626 \times 10^{-34} Js$). At low frequencies (up to the microwave region), a Taylor expansion of Planck's law yields the so-called Rayleigh-Jeans approximation

$$E(\lambda, T) = \frac{2kT}{\lambda^2} \qquad (II.3.2)$$

In the case of microwave remote sensing, the radiation of interest is emitted by and transmitted through the atmosphere.

As the radiation travels through the atmosphere, it gets absorbed by interaction with the matter. Also, emission from the matter, assumed in local thermodynamic equilibrium at temperature T, takes place. Let $I(s,\lambda)$ denote the radiation's intensity at wavelength λ along the radiation path. Over the incremental distance ds, the variation of intensity obeys (Chandrasekhar, 1950)

$$\frac{dI(s,\lambda)}{ds} = -\gamma(s,\lambda)I(s,\lambda) + \gamma'(s,\lambda)E(s,\lambda,T) \qquad (II.3.3)$$

where $\gamma(s,\lambda)$ is the absorption coefficient (a function of the density of absorbants and their cross sections), $\gamma'(s,\lambda)$ the emission coefficient representing the spontaneous radiation of the matter and $E(s,\lambda,T)$ the Planck function. This equation is called the Schwartzchild equation.

Reciprocity considerations yield $\gamma(s,\lambda) = \gamma'(s,\lambda)$. Moreover, the Rayleigh-Jeans approximation to Planck law permits us to rewrite (II.3.3) as

$$\frac{dI(s,\lambda)}{ds} = -\gamma(s,\lambda)I(s,\lambda) + \gamma'(s,\lambda)\frac{2kT(s)}{\lambda^2} \qquad (II.3.4)$$

The solution to this differential equation is identical to the one developed in linear system theory. The transmission operator $\Phi_\lambda(s,s')$ associated with the equation is

$$\Phi_\lambda(s,s') = \exp - \int_s^{s'} \gamma(\sigma,\lambda)d\sigma \qquad (II.3.5)$$

Thus the radiation obeys

$$I(s',\lambda) = I(s,\lambda)\Phi_\lambda(s,s') + \int_s^{s'} \gamma(\sigma,\lambda)\frac{2kT(\sigma)}{\lambda^2}\Phi_\lambda(\sigma,s')d\sigma \qquad (II.3.6)$$

The absorption coefficient integrated between two points is called the optical thickness. The exponential of its negative is referred to as the transmittance or transmission operator. The latter corresponds to a generalized attenuation coefficient.

To be consistent, one introduces the concept of brightness temperature $T_b(s,\lambda)$ which corresponds to the temperature of an ideal blackbody radiating the same power as the one observed. namely

$$T_b(s,\lambda) = \frac{\lambda^2}{2k}I(s,\lambda) \qquad (II.3.7)$$

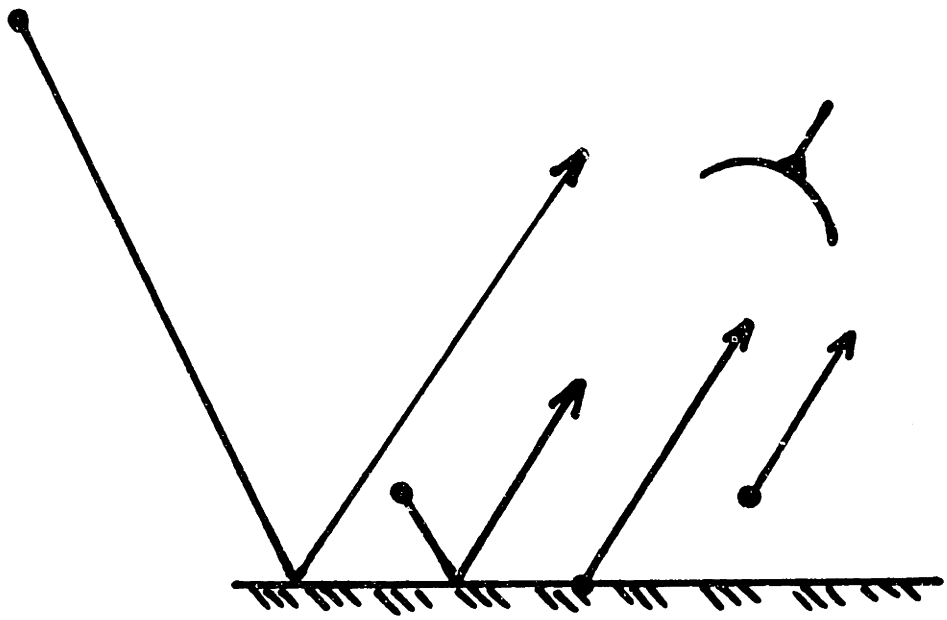


Figure II.3.1: Origin of the Upwelling Radiation Observed by a Flying Platform

Thus, the equation of radiative transfer becomes

$$T_b(s',\lambda) = T_b(s,\lambda)\Phi_\lambda(s,s') + \int_s^{s'} \gamma(\sigma,\lambda)T(\sigma)\Phi_\lambda(\sigma,s')d\sigma \qquad (II.3.8)$$

Let us compute the upwelling brightness temperature sensed by a satellite borne radiometer. This radiation has three sources: the atmosphere itself which radiates in all directions, the surface (sea or land) and the 3K big bang radiation (Staelin, 1981, Liu, 1980). Figure II.3.1 presents a schematic representation of the origins of that observed radiation. Note that the atmosphere radiation comes in two places.

The earth is not a perfect black body and thus reflects the downwelling radiation with a reflection coefficient $\rho(\lambda)$. The reflectivity is defined as the fractional power scattered back into the upper half-sphere (Tsang *et al.*, 1985). It is a function of the incident angle, polarization, roughness of the surface and correlation between the orientation of the ground at different points. More details on the ground modeling will be presented in section II.5. This radiation is proportional to its temperature and follows Planck's Law of black body radiation except for a scaling coefficient that takes into account the non perfection (Kirchoff Law). This scaling factor is called the emissivity and is denoted by $\epsilon(\lambda)$. For a non transparent media, one can write on the basis of conservation of energy

$$\epsilon(\lambda) + \rho(\lambda) = 1 \qquad (II.3.9)$$

To compute the upwelling brightness temperature, follow the radiation from the cosmic background down to the earth surface and back. At the top of the atmosphere, one has

$$T_b^*(\infty, \lambda) = T_{SKY}(\lambda) \qquad (II.3.10)$$

The asterisk symbolizes the brightness temperature traveling down. Just before getting reflected, the downwelling radiation is

$$T_b^*(0, \lambda) = T_{SKY}(\lambda)\Phi_\lambda(\infty, 0) + \int_\infty^0 \gamma(\sigma, \lambda)T(\sigma)\Phi_\lambda(\sigma, 0)d\sigma \qquad (II.3.11)$$

After reflection, taking into account the contribution from the ground, it becomes

$$\begin{aligned} T_b(0, \lambda) =& \rho(\lambda)T_{SKY}(\lambda)\Phi_\lambda(\infty, 0) + \rho(\lambda)\int_\infty^0 \gamma(\sigma, \lambda)T(\sigma)\Phi_\lambda(\sigma, 0)d\sigma \\ &+ (1 - \rho(\lambda))T_{surface} \end{aligned} \qquad (II.3.12)$$

Finally at the top of the atmosphere,

$$\begin{aligned} T_b(\infty, \lambda) =& \rho(\lambda)T_{SKY}(\lambda)\Phi_\lambda(\infty, 0)^2 + \rho(\lambda)\Phi_\lambda(0, \infty)\int_\infty^0 \gamma(\sigma, \lambda)T(\sigma)\Phi_\lambda(\sigma, 0)d\sigma \\ &+ (1 - \rho(\lambda))T_{surface}\Phi_\lambda(0, \infty) + \int_0^\infty \gamma(\sigma, \lambda)T(\sigma)\Phi_\lambda(\sigma, \infty)d\sigma \end{aligned} \qquad (II.3.13)$$

The four terms of the radiative transfer equation can be rewritten to separate the contribution of the atmosphere from the remaining. Namely,

$$\begin{aligned} T_b(\infty,\lambda) =& \rho(\lambda)T_{SKY}(\lambda)\Phi_\lambda^2(\infty,0) \\ &+ (1-\rho(\lambda))T_{surface}\Phi_\lambda(\infty,0) \\ &+ \int_0^\infty \gamma(\sigma,\lambda)(1+\rho(\lambda)\Phi_\lambda(0,\sigma))\Phi_\lambda(\sigma,\infty)T(\sigma)d\sigma \end{aligned} \qquad (II.3.14)$$

In the above equation, one recognizes (in order): (1) The big-bang radiation travelling twice through the entire atmosphere and reflected on the surface during the round trip, (2) the contribution from the surface, and (3) the atmospheric contribution, direct and indirect. Except for the first two terms, the brightness temperature is a weighted line integral of the temperature profile along the radiation path. Let $W(\sigma,\lambda)$ denote such a weight

$$W(\sigma,\lambda) = \gamma(\sigma,\lambda)(1+\rho(\lambda)\Phi_\lambda(0,\sigma))\Phi_\lambda(\sigma,\infty) \qquad (II.3.15)$$

and let $T_c(\lambda)$ denote the contribution from the ground and sky. One can rewrite the sounding equation as

$$T_b(\infty,\lambda) = T_c(\lambda) + \int_0^\infty W(\sigma,\lambda)T(\sigma)d\sigma \qquad (II.3.16)$$

This integral equation is except for T_c a Fredholm integral equation of the first kind (Van Trees, 1963, Twomey, 1963).

The weighting function is an obvious function of the wavelength through the absorption coefficient γ. The radiative transfer equation simplifies for so called "opaque" atmospheres where the operating wavelength is close to a resonance of one of the atmosphere's absorbers'. The transmittance of the atmosphere $\Phi_\lambda(\infty,0)$ is taken to be zero. Under such circumstances,

$$\begin{aligned} T_b(\infty,\lambda) &= \int_0^\infty \gamma(\sigma,\lambda)\Phi_\lambda(\sigma,\infty)T(\sigma)d\sigma \\ &= \int_0^\infty \gamma(\sigma,\lambda)\exp -\int_\sigma^\infty \gamma(\sigma',\lambda)d\sigma' T(\sigma)d\sigma \end{aligned} \qquad (II.3.17)$$

Noting that

$$\frac{\partial}{\partial s}\Phi_\lambda(s,s') = \gamma(s,\lambda)\Phi_\lambda(s,s') \qquad (II.318)$$

the sounding equation can be further simplified to

$$T_b(\infty,\lambda) = \int_0^\infty \frac{\partial\Phi_\lambda(\sigma,\infty)}{\partial\sigma} T(\sigma)d\sigma \qquad (II.3.19)$$

Integrating by parts and using the fact that the atmosphere is opaque yields

$$\begin{aligned} T_b(\infty,\lambda) =& \Phi_\lambda(\sigma,\infty)T(\sigma)|_0^\infty - \int_0^\infty \Phi_\lambda(\sigma,\lambda)\frac{\partial T(\sigma)}{\partial\sigma}d\sigma \\ =& T(\infty) - \int_0^\infty \Phi_\lambda(\sigma,\lambda)\frac{\partial T(\sigma)}{\partial\sigma}d\sigma \end{aligned} \qquad (II.3.20)$$

which defines the brightness temperature as a function of the lapse rate of the temperature profile (derivative of temperature with respect to height) projected along the sounding angle and the physical temperature of the atmosphere where the satellite is. This latter temperature is not of interest to the remote sensor. One can introduce the ground temperature by integrating the lapse rate variations according to

$$T(\infty) = T(0) + \int_0^\infty \frac{\partial T(\sigma)}{\partial\sigma}d\sigma \qquad (II.3.21)$$

Plugging (II.3.21) into (II.3.20) yields the desired alternate expression for the sounding expression

$$T_b(\infty,\lambda) = T(0) + \int_0^\infty (1-\Phi_\lambda(\sigma,\infty))\frac{\partial T(\sigma)}{\partial\sigma}d\sigma \qquad (II.3.22)$$

One must note that in the above equation, the ground temperature $T(0)$ is defined from the temperature at the satellite (infinity) and not vice-versa. Hence, if the lapse rate becomes null over a finite height, it does not contribute to the integral in the right hand side of (II.3.21). However, in such a case, the ground temperature is altered and the brightness temperature changes. This explanation prohibits the somewhat ambiguous equation (II.3.22) which tends to have one believe that

the brightness temperature is insensitive to constant temperature profiles. Note, however, that the ground temperature is independent, the difference in brightness temperature between two different frequencies does not respond to constant temperature profiles (or portion thereof). Namely,

$$T_b(\infty,\lambda) - T_b(\infty,\lambda') = \int_0^\infty (\Phi_\lambda(\sigma,\infty) - \Phi_{\lambda'}(\sigma,\infty)) \frac{\partial T(\sigma)}{\partial \sigma} d\sigma \qquad (II.2.23)$$

II.4 Weighting matrices

As noted in section II.2, close to resonance the oxygen absorption increases with pressure. This implies that, because of the gravitational stratification, the absorption $\gamma(\sigma,\lambda)$ will be larger close to the ground and decreases monotonically with height. The transmittance $\Phi_\lambda(\sigma,\infty)$ increases with altitude. The weighting function which is the product of these two functions will thus present a maximum at which the brightness temperature is most sensitive. The closer to a resonance the operating frequency is, the larger the absorption, and the higher the altitude of the weighting function's peak. Poon (1974) showed that if the absorption obeys the power law

$$\gamma(\sigma,\lambda) = \alpha(\lambda) p(\sigma)^{\beta(\lambda)} \qquad (II.4.1)$$

where the p is the pressure at abcisse σ, and where α and β are functions only of the operating wavelength, and if the pressure p obeys the so-called hyspometric equation and diminishes exponentially with height h according to

$$p(h) = p_o \exp(-h/H) \qquad (II.4.2)$$

then the weighting function peaks where the optical thickness $\int_\sigma^\infty \gamma(\sigma,\lambda) d\sigma$ equals unity. The weighting function's value at that location is

$$W_{max}(\sigma) = \frac{\gamma(\sigma_{max},\lambda)}{e} \qquad (II.4.3)$$

Figure II.4.1 present the weighting functions for 12 frequencies in the 60 GHz resonance region corresponding to the AMSU instrument. This instrument will be described in more details in chapter VI. The weighting functions are computed using standard US atmosphere. These functions look like skewed Gaussian bells. Note that because of their poor resolution, the weighting function sense a large area of the atmosphere at once.

The weighting matrix is defined by integrating the Schwartzchild equation along the line of sight. However, because of stratification, the absorption is only a function of height. Denoting by θ the angle from zenith the observation is performed at, the weighting matrix for an opaque atmosphere can be written as (the result extend to any atmosphere)

$$W(h,\theta,\lambda) = \gamma(h,\lambda) \exp - \int_{\frac{h}{\cos\theta}}^{\infty} \frac{\gamma(h',\lambda)}{\cos\theta} dh' \qquad (II.4.4)$$

This departure from zenith effectively increases the absorption and raises the peak of the weighting functions. Figure II.4.2 presents the impact of this lifting on the 3 channel MSU by plotting the weighting functions at nadir and at 56°. The hashed area corresponds to different reflectivities of the ground (this will be discussed in the following section). The MSU will be examined with more details in chapter V.

As noted, the sounding equation is a Fredholm equation of the first kind. By using measurements provided by different channels, one can infer the temperature profile $T(\sigma)$. This process is ill-conditioned and is usually implemented by discrete quadrature (Landweber, 1951, Strand and Westwater, 1968a, 1968b,Twomey, 1963). That is, the temperature is expressed as a function of a finite number of temperature located at various fixed pressure levels according to

$$T(\sigma) = \sum_{i=1}^{N} s(\sigma, \sigma_p(i)) T(\sigma_p(i)) \qquad (II.4.5)$$

where $\sigma_p(i)$ $i = 0, \ldots, N$ are the discrete levels and $s(.,.)$ the interpolating functions.

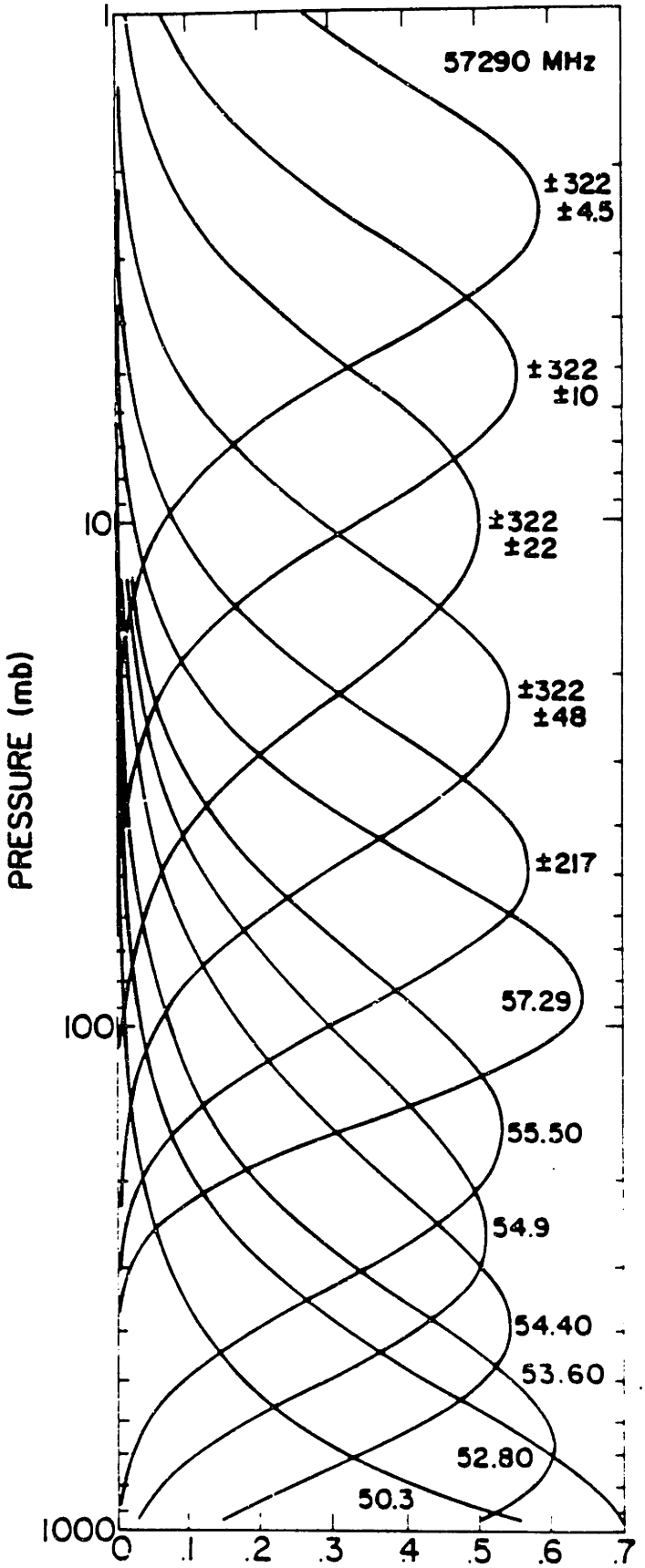


Figure II.4.1: Weighting functions for the 12 channel AMSU

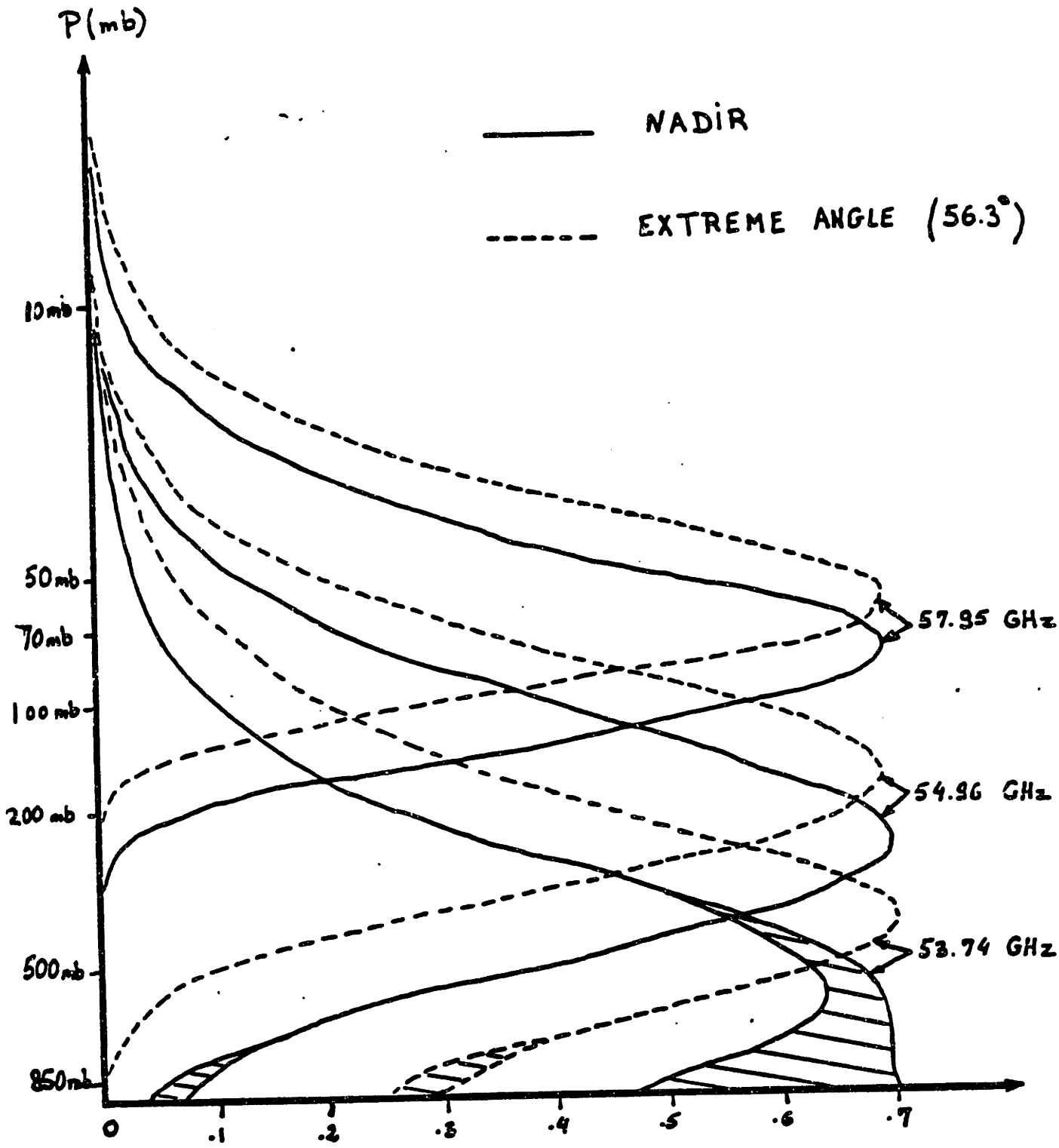


Figure II.4.2: MSU weighting functions at zenith (nadir) and 56°

The sounding equation becomes

$$T_b(\infty,\lambda) = \int_0^{\infty} w(\sigma,\lambda) \sum i = 1^N s(\sigma,\sigma_p(i))T(\sigma_p(i))d\sigma \qquad (II.4.6)$$

which permits the introduction of a discrete weighting function $W_\lambda(i)$ after inverting the order of summation:

$$T_b(\infty,\lambda) = \sum_{i=1}^{N} \underbrace{\int_0^{\infty} s(\sigma,\sigma_p(i))W(\sigma,\lambda)\, d\sigma}_{W_\lambda(i)} T(i) \qquad (II.4.7)$$

Grouping all discrete weighting functions together yields the weighting matrix which relates a discrete set of measurements with a discrete temperature profile. Note that these discrete weighting functions will change if clouds are present along the radiation path. Changes in the weighting functions created by clouds were not considered in this thesis.

The number of pressure levels, their locations, as well as the interpolating functions cannot be determined until the statistics (first and second moments) of temperature fields in the vertical direction are examined. That analysis will be performed in chapter III.

II.5 Ground Effects

To characterize the impact of the ground on the remote sensing equation, one need only to consider its emissivity or its absorptivity since the sum of the two equals one.

Emissivity varies greatly with surfaces and has been the subject of extensive research (check the references of the Manual of Remote Sensing, Chapter III). These studies pertain to the relationship between emissivity and resulting brightness temperature and relevant geophysical parameters. These studies are aimed to the remote sensing and classification of the earth's surface and use window channels

which operate far from O_2 and H_2O resonances. Because this dissertation is solely concerned with the remote sensing of atmospheric temperature, an exhaustive description of type of surfaces and corresponding emissivities would be superfluous. Let us briefly state the parameters relevant to emissivities and state the simple approximations used in this research.

For land surfaces, the emissivity increases with surface roughness (compared to wavelength). It decreases with soil moisture, and increases with green vegetation and/or dry snow cover (Staelin, 1981). Wet snow cover can increase the emissitivity. In all cases, the soil emits efficiently. For land, the reflectivity can be approximated by $\rho = 0.05$.

For ocean, the high dielectric constant of sea water dominates the emission. Thesalinity, surface roughness, the presence of ocean foam or oil slick alter the emissivity and the apparent brightness temperature. Sea ice emissivity varies from 0.8 to about 1.0 depending on frequency and the nature of the ice pack. A linear approximation to the reflectivity is (Klein and Swift, 1977)

$$\rho = 0.638 - 0.00272\nu, \qquad 10 < \nu < 200GHz \qquad (II.5.1)$$

These approximations are appropriate for the purposes of this dissertation.

II.6 Conclusions

In this chapter, we briefly reviewed the principles of radiative transfer. The sounding or observation equation shows that the sounding of temperature fields is fairly indirect and requires to be inverted the usage of statistics. The statistics of temperature fields are examined in chapters III and IV and are used in the remainder of this dissertation.

II.A Appendix A: Non Linearities in the Radiative Transfer Equation: A Correction Scheme

This appendix addresses the issue of non-linearities in the radiative transfer equation. Non-linearities appear in the sounding equation because the oxygen absorption coefficient $\gamma_{O_2}(T)$ depends on the temperature and the amount of water vapor present. The oxygen absorption depends on the temperature through the distribution of molecules in each of the states Φ_N, this implies that the weighting functions are themselves functions of temperature. This dependence yields discrepancies between the observed brightness temperature and the computed brightness temperature if the temperature profile used to compute the a-priori weighting matrix is too different from the true profile. This appendix quantifies the impact of this difference in temperature profiles, then proposes and analyzes a correction scheme based on a Taylor expansion of the absorption coefficient.

The integral form of the radiative transfer equation is solved by discretizing the atmosphere into slabs which may be expressed in a matrix form as

$$\overline{T}_b = W\overline{T} + \overline{\epsilon} \qquad (II.A.1)$$

where $\overline{T}_b$ is a vector composed of the brightness temperatures retrieved by each of the channels considered in the retrieval process, $\overline{\epsilon}$ is a noise vector associated with the observation (its thermal origin allows for a Gaussian white noise representation with covariance matrix N), and $\overline{T}$ is the discretized temperature profile. The number and locations of the slabs required to represent the continuous temperature profile can only be determined once the correlation structure of temperature in the vertical direction has been established. This task will be performed in Chapter III. For the moment, it suffices to know that the temperature has been sampled fast enough in the vertical direction.

The weight matrix W has temperature and humidity dependencies in ad-

dition to dependencies on other atmospheric parameters. Our a-priori knowledge of the temperature and humidity profile (from neighboring probed points, climatological arguments, etc) is used to compute the weighting matrix. Since this a-priori knowledge is not perfect, only an estimate of the true weighting function can be generated and used in the inversion of the radiative transfer equation.

Correction between the estimated and the true matrix can be performed using an iterative method which alternates between estimating the profile and the weighting matrix (Chahine, 1969, 1970). At each step, the retrieved temperature profile is used in the computation of the weighting matrix, a new retrieval operator is constructed, and a new estimate of the profile produced. The process is repeated until the difference between two successive retrieved temperature profiles is less than some threshold. Although convergence of such a method seems certain, its computational burden can be excessive. Furthermore, the method requires improved humidity profile retrieval schemes to compute the exact weighting functions for low-altitude probing channels. For these reasons and for the reason that such iterative algorithms do not allow for the incorporation of horizontal correlation, we will concentrate on statistical linear estimation.

However, when studying the impact of error in the weighting matrix on the retrieved temperature profiles, one must incorporate into the analysis the region of support of the retrieval operator, the number of channels and the statistics of the retrieval dataset. This makes the analysis operator dependent, which is an undesirable feature.

Let us then approach the issue of error in the weighting matrix from a purely forward-problem point of view. Let $\overline{T}_o$ and $\overline{T}_i$ denote the true and a-priori guesses of the temperature profile respectively. Using $\overline{T}_i$ as an estimate for the true temperature profile, we assume that the discretized radiative transfer equation is

$$\overline{T}_b = W(\overline{T}_i)\overline{T}_o + \overline{\epsilon} \qquad (II.A.2)$$

when, in fact, we are observing

$$\overline{T}_b = W(\overline{T}_o)\overline{T}_o + \overline{\epsilon} \qquad (II.A.3)$$

Rewriting the first equation,

$$\overline{T}_b = W(\overline{T}_o)\overline{T}_o + (W(\overline{T}_o) - W(\overline{T}_i))\overline{T}_o + \overline{\epsilon} = W(\overline{T}_o)\overline{T}_o + \overline{\epsilon}' \qquad (II.A.4)$$

we introduce a additional noise term to the observation noise to obtain $\overline{\epsilon}'$. Assume that the algorithm for choosing the profile $\overline{T}_i$ is such that $W(\overline{T}_o) - W(\overline{T}_i)$ has zero mean and is uncorrelated with $W(\overline{T}_i)$. These assumptions can be interpreted as requirements for that algorithm to be unbiased, i.e. centered around the actual profile (a desired feature in general), and that its departures from the actual profile are small. Under these assumptions, one can consider $\overline{\epsilon}'$ as a structured noise observation vector with zero mean and covariance matrix N' given by

$$N' = <\overline{\epsilon}'\overline{\epsilon}'^t> = N + <(W(\overline{T}_o) - W(\overline{T}_i))T_o T_o^t (W^t(\overline{T}_o) - W^t(\overline{T}_i))> \qquad (II.A.5)$$

where $< . >$ denotes the expectation over all actual and a-priori temperature profiles. The covariance matrix N' is no longer diagonal. Note that since the intended retrieval operation is linear, only the first and second moments of the signals and noises are required to design and characterize the retrieval operation, and that, under these working assumptions, the impact of non-linearities can be characterized using solely N'. The above equation allows us to interpret the non-linearities in the radiative transfer equation in terms of induced noise. Too large an error in the computation of the weight matrix will result in data embedded in a noise with a large variance. Such noisy measurements are unreliable.

The determination of the amount of error tolerable in the first guess of the temperature profile can now be seen as equivalent to the determination of the amount of noise tolerable in the measurement process. Note that the induced covariance matrix depends on the distribution of profiles used. To avoid being statistics

dependent, let us consider the deviation $(W(\overline{T}_o) - W(\overline{T}_i))\overline{T}_o$ for a specific deviation between a-priori and true profiles.

Because an error in the a-priori temperature profile will not have the same impact on the weighting matrix for every altitude along the profile, a worst case analysis is performed. Since channels are very sensitive to the physical temperature at altitudes corresponding to the peak of their respective weighting functions, an error at that particular level will be sensed with the most effect in terms of changes in the observed brightness temperature. Starting from a standard temperature profile (the a-priori profile), a new profile (the actual one) is generated by shifting the profile by a fixed amount at the location of the peak of the weighting function. Using 128 levels between the ground and 1 *mbar*, such a shift is applied over an altitude of about 500 meters.

Several compiled temperature profiles are used in the simulations. These profiles are chosen to be as different as possible. In the study, three different profiles are utilized: the US standard atmosphere, 30° north January average, and 60° north January average. In all cases, simulations are performed for both dry and wet atmospheres. In the latter case, the water vapor density is taken to be $10\exp(-h/2)g/cm^3$ where h is the height expressed in km. This analysis assumes a positive correlation of the difference $W(\overline{T}_o) - W(\overline{T}_i)$ over neighboring levels, an assumption that can be translated as a requirement that the estimated temperature profile to be smooth.

We study the robustness of the discretized radiative transfer equation by comparing the deviation $\sigma = (W(\overline{T}_o) - W(\overline{T}_i))\overline{T}_o$ with two bounds: a loose one, $\sigma < 0.2K$, and a tight one, $\sigma < 0.1K$. These bounds are typical of the standard deviations of present sounding systems. For instance, the Microwave Sounding Unit (MSU), which is the system most studied in this dissertation, has observational noise with a nominal standard deviation on the order of $0.2K$.

channel (MHz)	2K	5K	10K
50300	+	+	+
52800	+	+	+
53600	+	+	+
54400	*	*	*
54900	*	*	*
55500	*	*	*
57290	*	o	+
f_o ±217	*	o	+
f_o ±322±48	*	o	+
f_o ±322±22	*	+	+
f_o ±322±10	o	+	+
f_o ±322±4.5	o	+	+

where $f_o = 57290.334 MHz$, $*$ means that $\sigma < 0.1K$, o that $\sigma < 0.2K$, and $+$ that $\sigma > 0.2K$

Table II.A.1: Deviation between the true brightness temperature and the brightness temperature computed using a constant weighting matrix as a function of the temperature shift in the profile

Table II.A.1 presents the deviations introduced for different amount of temperature shift. These simulations are performed for the 12 channel AMSU whose weighting functions are presented in Figure II.4.1. The deviations in brightness temperature are the maximum deviations encountered over the three standard atmospheres (worst case analysis). The shifts applied to the profile are 2, 5, and 10 K respectively. $5K$ is the typical temperature deviation associated with typical datasets used for retrievals whereas a $10K$ deviation, is sometimes encountered close to the ground in winter situations. For half of the channels, a temperature profile deviation of 5K produces a difference in brightness temperature smaller or equal to the loose bound. For the highest and lowest probing channels, the deviations exceed that value: the maximum difference for the 5K deviation is 0.7K for the 50.3

GHz channel whereas for the 10 K deviation, the largest difference is 1.1 K (also for 50.3 GHz). The differences in brightness temperature for the six highest probing channels are all less than 0.6 K for a profile deviation equal to 10K.

Let us now consider a correction scheme for the weighting matrix that will extend the range of tolerable error between the true profile and the a-priori profile. As seen in section II.2., the dominating elements in the radiative transfer equation are the oxygen absorption coefficient $\gamma(h, \lambda, T)$ and its integrated value Φ_λ. The proposed correction scheme considers a first-order Taylor expansion of the transmittance. This scheme was first introduced by Philip Rosenkranz (Personal Communication). The presented calculations assume an opaque atmosphere, that is the contributions from the ground and from the reflected downward radiation by the ground are neglected. The principles of calculation presented below can easily be extended to transparent atmospheres.

Recall the simplified sounding equation

$$T_b(\infty, \lambda) = \int_0^\infty \frac{\partial \Phi_\lambda(\sigma, \infty, T)}{\partial \sigma} T(\sigma) d\sigma \qquad (II.A..6)$$

where the dependence of the transmittance on the temperature is written explicitely. The difference in brightness temperature caused by a change $\delta T(\sigma)$ in the temperature profile is

$$\delta T_b(\infty, \lambda) = \int_0^\infty \frac{\partial \Phi_\lambda(\sigma, \infty, T + \delta T)}{\partial \sigma} \big(T(\sigma) + \delta T(\sigma)\big) d\sigma - \int_0^\infty \frac{\partial \Phi_\lambda(\sigma, \infty, T)}{\partial \sigma} T(\sigma) d\sigma \qquad (II.A..7)$$

The Taylor expansion of the vertical spatial derivative of the transmittance is

$$\frac{\partial \Phi_\lambda(\sigma, \infty, T + \delta T)}{\partial \sigma} = \frac{\partial \Phi_\lambda(\sigma, \infty, T)}{\partial \sigma} + \delta T \frac{\partial^2 \Phi_\lambda(\sigma, \infty, T)}{\partial T \partial \sigma} \qquad (II.A..8)$$

Hence the change in brightness temperature is

$$\delta T_b(\infty, \lambda) = \int_0^\infty (T(\sigma) + \delta T(\sigma)) \frac{\partial^2 \Phi_\lambda(\sigma, \infty, T)}{\partial T \partial \sigma} d\sigma + T(\sigma) \frac{\partial^2 \Phi_\lambda(\sigma, \infty, T)}{\partial T \partial \sigma} \delta T(\sigma) d\sigma \qquad (II.A..9)$$

The expansion of the transmittance is related to the expansion of the absorption coefficient according to

$$\frac{\partial^2 \Phi_\lambda(\sigma,\infty,T)}{\partial T \partial \sigma} = \frac{\partial \gamma(\sigma,\lambda,T(\sigma))}{\partial T}\Phi_\lambda(\sigma,\infty,T) + \gamma(\sigma,\lambda,T(\sigma))\frac{\partial \Phi_\lambda(\sigma,\infty,T)}{\partial T} \qquad (II.A..10)$$

Plugging (II.A.10) into (II.A.7) yields

$$\delta T_b(\infty,\lambda) = \int_0^\infty \delta T(\sigma)\gamma(\sigma,\lambda,T(\sigma))\Phi_\lambda(\sigma,\infty,T)d\sigma$$
$$+ \int_0^\infty T(\sigma)\frac{\partial \gamma(\sigma,\lambda,T(\sigma))}{\partial T}\Phi_\lambda(\sigma,\infty,T)\delta T(\sigma)d\sigma$$
$$+ \int_0^\infty T(\sigma)\gamma(\sigma,\lambda,T(\sigma))\delta T(\sigma)d\sigma\left(\int_\sigma^\infty \frac{\partial \gamma(\sigma',\lambda,T(\sigma'))}{\partial T}d\sigma' \exp\left(-\int_\sigma^\infty \gamma(\sigma",\lambda,T(\sigma"))d\sigma"\right)\right) \qquad (II.A..11)$$

After factorization,

$$\delta T_b(\infty,\lambda) = \int_o^\infty \delta T(\sigma)\gamma(\sigma,\lambda,T(\sigma))\Phi_\lambda(\sigma,\infty,T)d\sigma$$
$$+ \int_0^\infty T(\sigma)\Phi_\lambda(\sigma,\infty,T)d\sigma\delta T(\sigma)\left[\frac{\partial \gamma(\sigma,\lambda,T(\sigma))}{\partial T} - \int_\sigma^\infty \gamma(\sigma',\lambda,T(\sigma'))\frac{\partial \gamma(\sigma',\lambda,T(\sigma))}{\partial T}d\sigma'\right] \qquad (II.A..12)$$

After integration by parts

$$\delta T_b(\infty,\lambda) = \int_0^\infty \delta T(\sigma)\frac{\partial \Phi_\lambda(\sigma,\infty,T)}{\partial \sigma}d\sigma$$

+

$$\int_0^\infty \frac{\partial \gamma(\sigma,\lambda,T(\sigma))}{\partial T}\Phi_\lambda(\sigma,\infty,T)(T(\sigma) - T_b(\sigma,\lambda))\delta T(\sigma)d\sigma \qquad (II.A..13)$$

The first term of the RHS is the change of brightness temperature if the weighting matrix is left unchanged (linear sensing). The second term can be interpreted as a sounding equation using an incremental weight dW equal to

$$dW(\sigma,T(\sigma)) = \Phi_\lambda(\sigma,\infty,T)\frac{\partial \gamma(\sigma,\lambda,T(\sigma))}{\partial T}(T(\sigma) - T_b(\sigma,\lambda)) \qquad (II.A.14)$$

The incremental weight is the product of three terms that do not peak at the same locations. The transmittance Φ_λ is a monotonically increasing function of altitude.

The difference $T(\sigma)-T_b(\sigma)$ will be smallest when the brightness temperature is close to the physical temperature; this occurs whenever the atmosphere is very opaque. For channels which are opaque (probing at high altitude), this is almost the case along the entire profile up to a certain altitude when the absorbant density decreases too much to emit radiation anymore. For channels which are more transparent (probing closer to the ground), this difference will be smallest close to the ground. This quantity can take either positive or negative values depending on the lapse rate. The last term in the product is the derivative of the absorption with respect to temperature, which is typically largest close to the ground. This quantity is negative. This analysis proves, to the contrary of the linear weighting matrix, the form of the incremental weight matrix is not easy to predict. The only trend is that it decays with altitude past a certain altitude, and that for opaque channels, it is close to zero at low altitudes.

The incremental weight for a regular atmosphere is presented here for completeness. With the notations of section II.3,

$$dW(\sigma,T(\sigma)) = \Phi_\lambda(\sigma,\infty,T)\frac{\partial\gamma(\sigma,\lambda,T(\sigma))}{\partial T}(T(\sigma)-T_b(\sigma,\lambda))$$
$$+\rho(\lambda)\Phi_\lambda(0,\infty,T)\left[T(\sigma)\Phi_\lambda(\sigma,0,T)-T_b^*(\sigma)\right]\Phi_\lambda(\sigma,\infty,T)\frac{\partial\gamma(\sigma,\lambda,T(\sigma))}{\partial T} \qquad (II.A.15)$$

which becomes identical to the opaque atmosphere incremental weight regardless of opacity of the atmosphere. In the above expression, $T_b^*(\sigma)$ denotes the downward propagating brightness temperature at altitude σ. If the ground reflectivity $\rho(\lambda)$ becomes zero, the second term in the summation becomes zero. Such a low reflectivity is encountered above land (notably tropical forests). In such cases, the expression in equation (II.A.14) can be used.

Figures II.A.1, II.A.2 and II.A.3 present the incremental weights for a low-altitude, medium-altitude, and high-altitude probing channels for a dry atmosphere. The lowest probing channel has an incremental weight which looks like a decaying

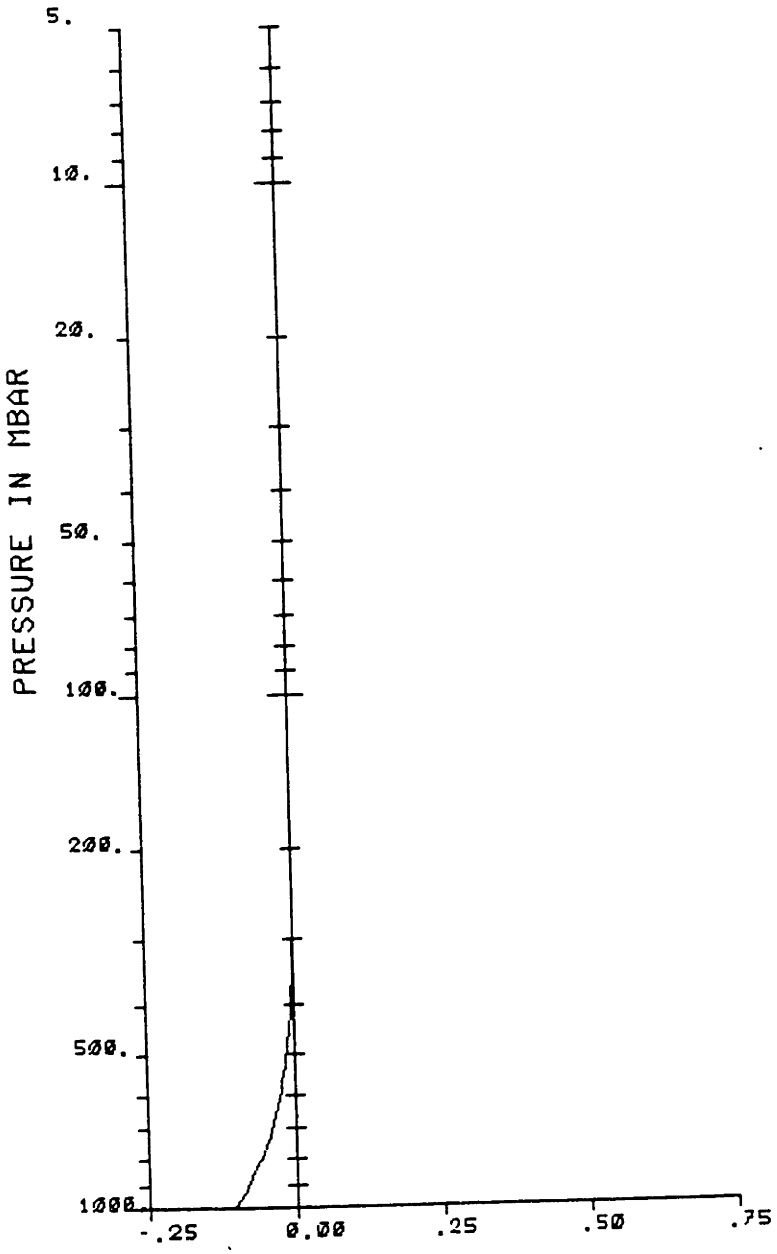


Figure II.A.1: Incremental weighting function with respect to temperature for f=50.3 GHz

exponential, whereas the other two channel incremental weights take on positive and negative values. The higher probing the channel, the higher the incremental weight is.

The implementation of the Taylor correction scheme and the analysis of its performances was carried out as before. Results of the simulations are presented in Table II.A.2. The improvement over the first case (no correction) is remarkable.

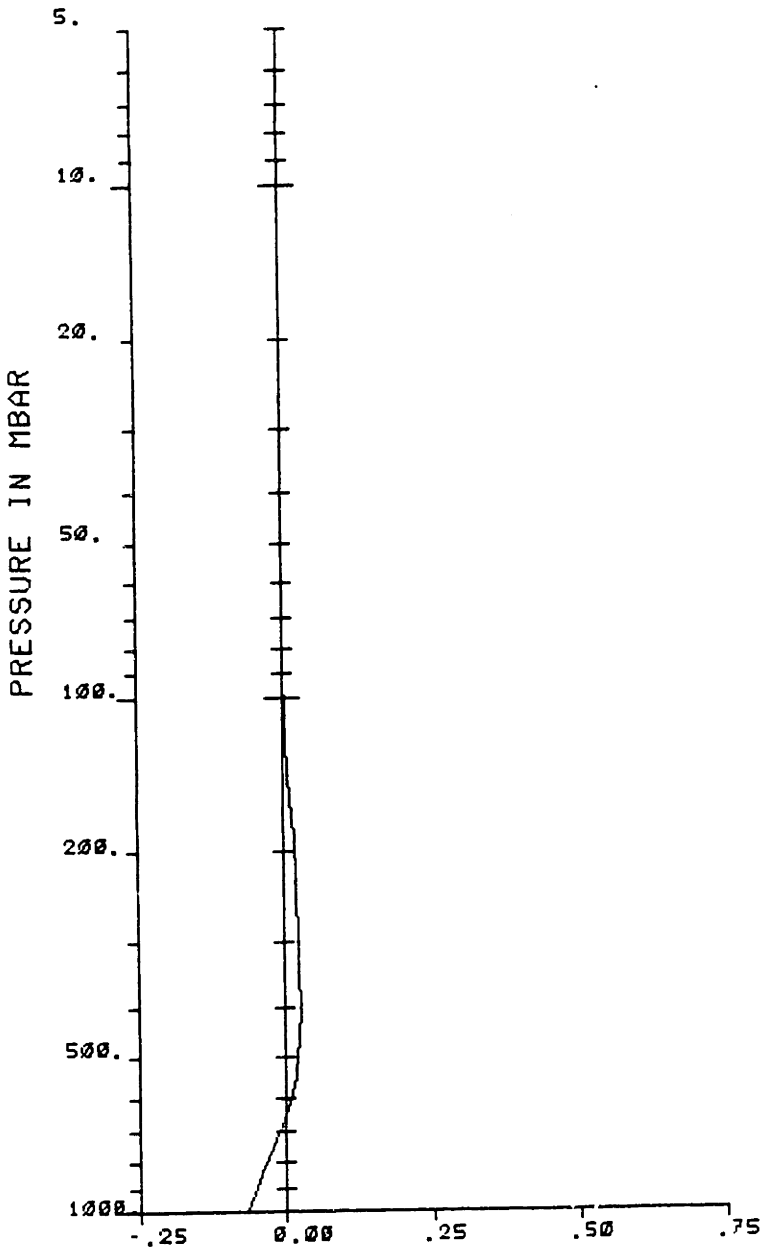


Figure II.A.2: Incremental weighting function with respect to temperature for f=53.6GHz

If one allows errors bounded by $5K$ between actual and a-priori profiles, then the Taylor expansion will correct the error within the tight bound for all but two of the AMSU channels. The brightness temperature deviations for the first two channel are about 0.5K. For the 10K deviation, the maximum difference is 0.8K for the 50.3 and 52.8 GHz channels. Note that these two channels have a significant contribution from the ground (check Figure II.4.1), hence they are more susceptible to errors in

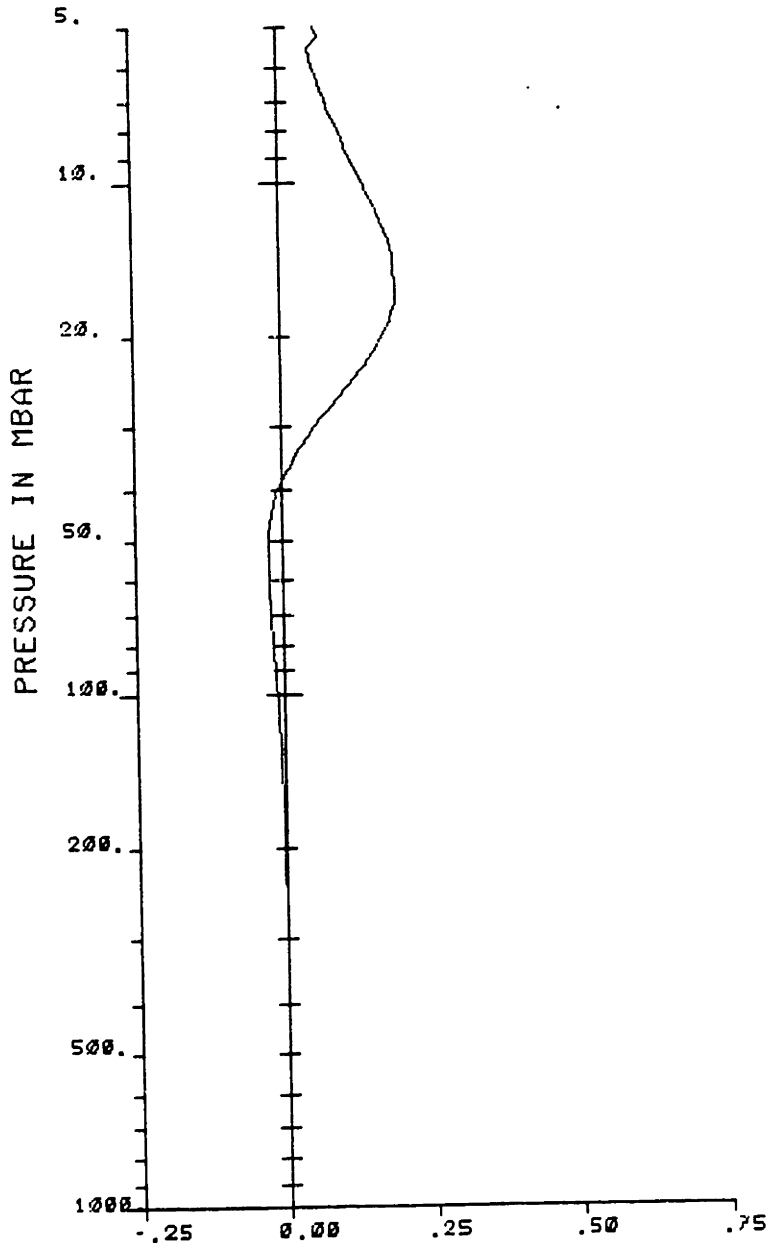


Figure II.A.3: Incremental weighting function with respect to temperature for f=57.290334±0.322±0.01 GHz

the ground's reflectivity. For these channels, the proposed bounds may be too tight.

In conclusion, the impact of error in the weighting matrix due to nonlinearities has been investigated as an increase in the noise associated with the measurement process. A simple mathematical operation allowed the interpretation of the error in the weighting functions as a structured additive noise uncorrelated with the true observation noise. This interpretation relied on the assumption that

channel	2K	5K	10K
50300	o	+	+
52800	o	+	+
53600	*	*	*
54400	*	*	*
54900	*	*	*
55500	*	*	o
57290	*	*	+
$f_o \pm 217$	*	*	+
$f_o \pm 322 \pm 48$	*	*	+
$f_o \pm 322 \pm 22$	*	*	+
$f_o \pm 322 \pm 10$	*	*	+
$f_o \pm 322 \pm 4.5$	*	*	+

where $f_o = 57290.334MHz$, $*$ means that $\sigma < 0.1K$, o that $\sigma < 0.2K$, and $+$ that $\sigma > 0.2K$

Table II.A.2: Deviation between the true brightness temperature and the temperature computed using the proposed first order correction scheme versus deviation in temperature

the temperature profile's estimate is unbiased, has a small variance and does not present sharp variations in the vertical direction. Under these assumptions, the effect of error has been shown to be important if no corrective scheme is applied to the retrieval process. The a-priori profile has to be within a couple of degrees of the actual profile for the level of noise introduced by the non-linearities to be small compared to the brightness temperature's requirements. A correction scheme based on a Taylor expansion of the weighting functions presented dramatic improvements; It allowed departures of $5K$ even $10K$ between actual and a-priori profiles with only errors of the order of 0.2 K in the brightness temperatures.

CHAPTER III

THE THREE-DIMENSIONAL MACROSTRUCTURE OF ATMOSPHERIC TEMPERATURE FIELDS: MODEL STRUCTURE

"All things are artificial, for nature is the art of God." Thomas Browne

III.1 Introduction

The terrestrial atmosphere can be modeled as a stochastic system governed by the laws of thermodynamics and fluid mechanics (Eskinazi, 1975). It is thus characterized by a probability distribution or equivalently by its characteristic moments.

The atmosphere typically presents scales of phenomena whose distribution (or "state") varies with geographic location and time of year (Lorenz, 1969, Charney, 1971, Palmen and Newton, 1969). Large scale phenomena vary more slowly with time than do small scale phenomena. Because of these different time scales, the definition of the probability distribution characterizing the atmospheric flow or equivalently its characteristic moment is a difficult task. This difficulty is inherent to the characterization of non-stationary processes. In all cases, these moments vary with time and space.

One can think of a climate as a probability distribution which is stationary (North and Cahalan, 1981). For a specific climate, one can estimate the climate's moments by taking an ensemble of independent measurements of the atmospheric state and by computing the moments based on this ensemble of measurements. However, the number of measurement used or equivalently the number of degrees of freedom in the compilation must be selected with care. If the number of degrees of freedom used in the evaluation of these moments is small, one will characterize the distribution around a specific realization rather than the climatology itself.

This is true because of the temporal and spatial correlation between measurements. To avoid this problem, one must ensure that the number of degrees of freedom is large enough so that the set of measurements characterizing a specific realization of atmospheric temperature field characterizes the desired climate.

The purpose of this model is the mathematical representation of the three dimensional covariance (second moment of the probability distribution) of atmospheric temperature fields as a function of pressure and isobaric displacement when the number of degrees of freedom approaches infinity. The intended range of validity of the model is from the ground up to pressure levels of the order of a millibar (troposphere, tropopause, and the lower part of the stratosphere) and isobaric displacements at mesoscale and larger spatial scales. Because of the large number of degrees of freedoms it models and because it only has a finite number of parameters itself, the model will not characterize the temperature structure of specific phenomena such as storms, either within or across the boundary of the storm.

At the spatial scales of interest, atmospheric flow and temperature fields are characterized by a three-dimensional structure that exhibits the features associated with two-dimensional flow (Eskinazi, 1975). This two-dimensional aspect is caused by the evolution of temperature fields controlled to a great extent by turbulent flow in the isobaric domain (Charney, 1971) and by the global balance of solar heating by poleward transport of heat by baroclinic eddies (Charney, 1973). This difference in the behavior in the vertical and horizontal domains can also be explained by the ratio of horizontal to vertical scales in the atmosphere (Boer and Sheperd, 1983). This phenomenon forces a separate modeling of the vertical domain (sometimes referred to as the pressure domain) and the horizontal domain (thereafter referred to as the isobaric domain).

The basis for this modeling is mathematical in nature with the adjunction of constraints and properties derived from the physics of the atmosphere. The

development of the model will begin with an analysis of the revelant physical properties of temperature fields in both the vertical and the isobaric directions. The three dimensional model will be derived as a convolutional combination of models for both domains.

The identification and analysis of the model will be conducted in Chapter IV. The model will be fitted to observed data. Its validity will be assessed by comparing the observed covariances with the one derived from the model. The impact of the number of degrees of freedom used in the evaluation of the model will be assessed in terms of confidence intervals for the different parameters of the model.

Strictly speaking, the word structure denotes the expected value of the squared difference between two temperatures at different locations (Boltenkov, 1966, Gage, 1979). Under the assumption of stationarity, the structure is directly related to the familiar covariance and by Fourier transform to the spectrum. In this analysis, the word structure will denote any of these quantities.

Before proceeding, one last word is required. As it will soon appear, the analysis presented here of the properties of atmospheric flow and its impact on temperature is quite extensive. It is so for three principal reasons: First, the author wished there was such an analysis when he developed the model. Second, such an analysis has never been performed in this context and for the audience of remote sensing and objective analysis researchers. Third, it is the author's strong belief that the understanding of the physics of the phenomena being studied (in this case, atmospheric temperature) is essential to the understanding of the performances of the mathematical techniques used (in this case, multidimensional filtering).

III.2 Physical and mathematical constraints on the isobaric covariance kernel

The most important aspect of isobaric temperature fields is their local homogeneity. Homogeneity is the extension of wide sense stationarity of one-dimensional time series. It is important for operational purpose (Rutherford, 1971) and because non homogeneity in the temperature field can yield cases where the covariance K_{TT} between two points is not symmetric (Gandin 1963).

The homogeneous covariance function depends only on the relative position of the two points considered. That is

$$K_{TT}(\vec{x}, \vec{x}') = K_{TT}(\vec{x} - \vec{x}') \qquad (VIII.2.1)$$

where $\vec{x} - \vec{x}'$ is the lag vector. The computation of the covariance can be decomposed into two steps. The first one is the computation of a scalar metric measure $|(\vec{x}-\vec{x}')|$ which accounts for the correlation structure of the temperature field. It expresses if the correlation is isotropic, ellipsoidally isotropic, direction-dependent of separable. When producing a contour plot of the covariance function as a function of displacement along the latitude and longitude axes, the metric $|.|$ controls the shape of the isopleths. The second part of this computation is the evaluation of the covariance function which becomes a one-dimensional problem. The characterization of this covariance is performed either in the spatial or in the frequency domain. When producing a cut across the contour plot through the zero-displacement, the one-dimensional function controls the shape of the covariance. These two steps are separate and they will be analyzed in turns.

III.2.1 Correlation structure of isobaric temperature fields

The analysis of homogeneity depends on the specific geometry used for the calculation. In the case of planar geometry, the theory (Charney, 1971) predicts a locally homogeneous atmosphere in horizontal planes. This property has experimentally been observed for displacements up to 3 Mm (Gandin, 1963, Boltenkov,

1966). Although, for the purpose of the intended modeling, planar geometry is appropriate, the analysis of homogeneity in spherical geometry is important because of its consequences.

Let us consider the terrestrial sphere and let $\vec{r}$ denote a vector from the center to the surface of the sphere. Homogeneity can be expressed by stating that the covariance between two points characterized by the vectors $\vec{r}$ and $\vec{r}'$ depends solely on the inner product between the two vectors. That is, denoting the covariance by $K_{TT}(\vec{r},\vec{r}')$

$$K_{TT}(\vec{r},\vec{r}') = \rho(\vec{r}.\vec{r}') \qquad (III.2.2)$$

This hypothesis has important consequences. The empirical orthogonal expansion or Karhunen-Loeve expansion of a stochastic field satisfying (III.2.2) is (Van Trees, 1968)

$$T(\vec{r}) = \sum_{l=0}^{\infty} \sum_{m=-l}^{m=l} T_{lm} Y_l^m(\vec{r}) \qquad (III.2.3)$$

where the Y_l^m are spherical harmonics. As for the expansion of homogeneous temperature field in planar geometry, the coefficients T_{lm} are mutually orthogonal, and are found by taking the inner product of the fields with the basis functions (North and Cahalan, 1981).

Large scale phenomena correspond to small l and and do not exhibit rapid variations with respect to azimuth. Spherical harmonics are also the solution to the wave equation in spherical coordinates which plays an important role in non-divergent barotropic flow (Thomson, 1982). This expansion of the temperature field in spherical harmonics derived from statistical considerations coincides with the expansion along the eigenvectors of the deterministic linearized operator describing the atmospheric flow (North and Cahalan, 1981). Hence, it is an appropriate expansion.

This expansion for the temperature field can now be used to model an-

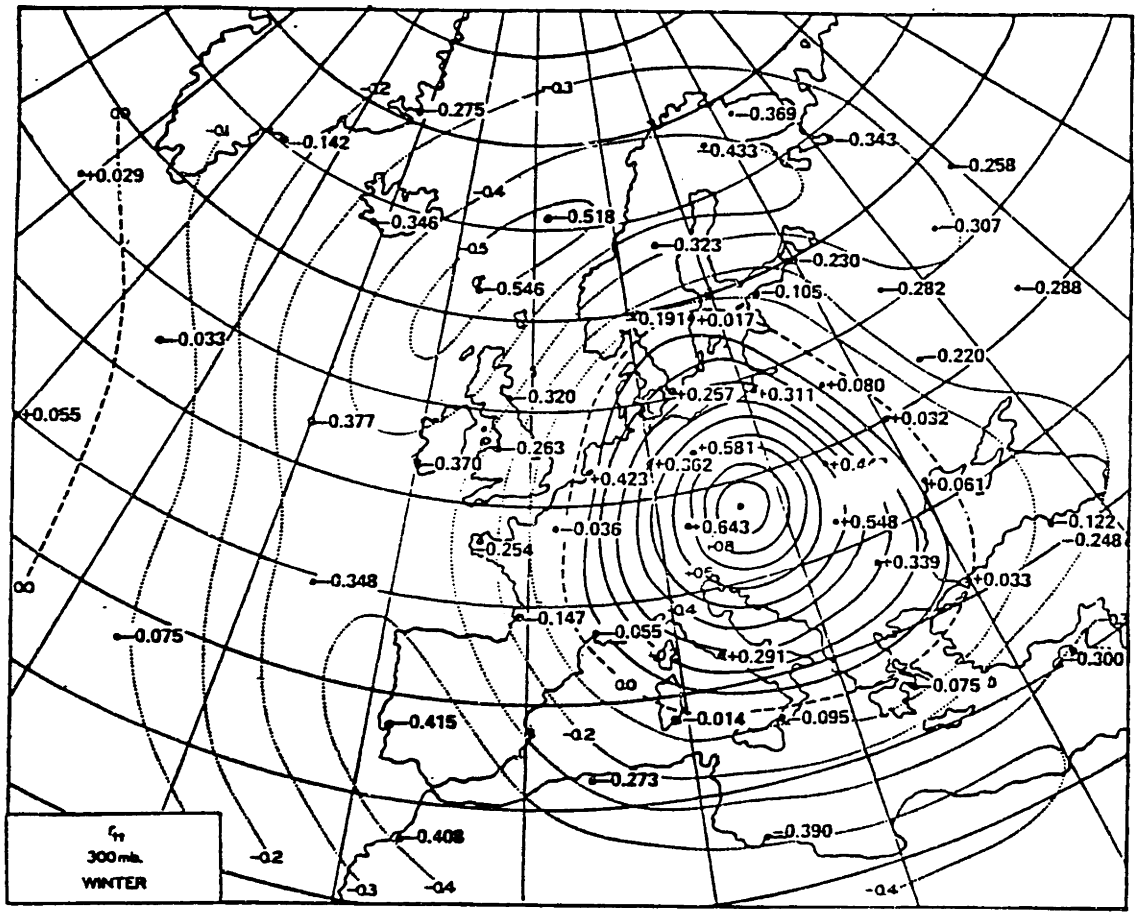


Figure III.2.1: Observed correlation for temperature field (from Buel, 1959)

alytically the isobaric covariance kernel as a function of distance between points, orientation with respect to the earth rotation axis, and topology.

Theory (Charney, 1971) predicts that, in addition to being homogeneous, the isobaric temperature field is isotropic away from boundaries. This homogeneity is observed but only over finite distance (up to 3 Mm). Typical atmospheric covariance fields possess positive correlation region which is more or less elliptical (Wallace and Gutzler, 1981). This positive correlation is often the only significant feature in the correlation pattern. Two such patterns are presented in Figures III.2.1 and III.2.2. The correlations are analyzed over Northern Europe (Rinner and Jarvenoja,

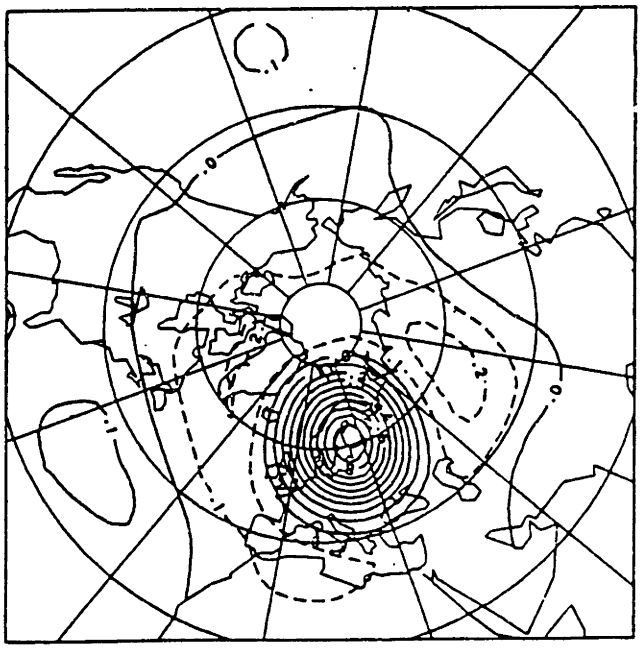


Figure III.2.2: Observed correlation for temperature field (from Rinne and Jarvenoja, 1985)

1985, Buel, 1959). In both cases, the high correlation isopleth ($\rho > 0.5$) present an ellipsoidal form. The orientation of this ellipsoidal pattern varies with location but has generally a major axis along the East-West direction, and a minor axis along the North-South direction. This pattern was observed, indirectly by analyzing wind correlation, by Dartt (1972) and by Ramanathan *et al.* (1972). Thiebaux (1976) analyzed the variations of the correlation as a function of the azimutal angle. Analysis of the zero crossing distance also indicates an E-W oriented ellipsoidal correlation pattern. These patterns correspond to east-west propagating synoptic scale phenomena. This latter analysis is corroborated by the study of potential energy by Baer (1972), which is related to temperature covariance (see section III.2.20).

As the scale increases, the correlation becomes negative and the temperature field becomes more barotropic (Herring, 1980). Typical for the strong negative

correlation region is the one of a dipole pattern with a North-South Axis (Thiebaux, 1977, Wallace and Gutzler, 1981). Such a dipole pattern has a cosine dependence in the azimutal angle.

The orography and steady flow characteristics also play an important part in the homogeneity and anisotropic behavior of temperature fields. Strong alterations to these idealized behaviors will occur whenever the topographic effects are important. This is for instance the case for the Pacific North Atlantic (PNA) oscillation or for teleconnections over the Himalaya. This effect is pressure dependent. At mid-troposphere, the regional scale is the dominant factor in the form of the correlation pattern. At the surface, the impact of the topography is more important.

Let us consider the characterization of the anisotropy of isobaric temperature fields. From an operational point of view, a simple characterization as a function of distance and of azimutal angle is preferable (Rutherford, 1972). This is not always possible given the varied behavior of temperature fields across the globe. Rinne and Jarvenoja (1985) analyzed the anisotropic behavior in the Northern hemisphere. Temperature covariance kernels are most isotropic at the north pole and worsen as one approaches the equator. At any given latitude, it is worse at the PNA.

When deriving an analytical model for the characterization of the anisotropy, one must bear in mind the need for simple variations with distance and azimutal angle. Furthermore, since isobaric temperature covariance kernels are homogeneous, they admit a spectral representation. This spectral relationship should also be simple.

The simplest characterization of anisotropy is the ellipsoidal representation which is most appropriate for small displacements. It is characterized by three parameters: orientation, length of major axis, and eccentricity. Its advantage over other characterizations of anisotropy is that it is a conformal mapping of

the isotropic (circular) case and as such preserves all angle-distance relationships between the spatial and wavenumber domains.

When the ellipsoidal characterization is not appropriate, one can consider a series expansion using any orthonormal basis. In the case of spherical geometry, which is appropriate for large displacements, the expansion takes the form

$$K_{TT}(\theta_i, \lambda_i, \theta_j, \lambda_j, p) = K_{TT}(m(\theta_i, \lambda_i, \theta_j, \lambda_j), p) \qquad (III.2.4)$$

where the metric m is defined by the expansion

$$m(\theta_i, \lambda_i, \theta_j, \lambda_j) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} [C_{ml}(i,j) \cos m(\lambda_j - \lambda_i) + S_{ml}(i,j) \sin m(\lambda_j - \lambda_i)] Y_l^m(\cos \theta_{ij}) \qquad (III.2.5)$$

where Y_l^m is the Legrendre polynomial of order l, m, C_{ml} and S_{ml} the corresponding expansion coefficients. The angle θ denotes the colatitude (difference between the two point latitudes) and λ the longitude. θ_{ij} denotes the distance between the points i and j on the sphere. Namely,

$$\cos \theta_{ij} = \cos \lambda_i \cos \lambda_j cos(\theta_i \theta_j) + \sin \lambda_i \sin \lambda_j \qquad (III.2.6)$$

In this expansion, m equal zero corresponds to the isotropic field. The main inconvenience with this expansion is that for displacements of the order of a few megameters, the usage of Legendre polynomials is not appropriate.

For this reason, it is preferable to work in a planar geometry although it is not correct for large displacements. Furthermore, the analysis of the correlation in the planar geometry will be consistent with the evaluation of the radiative transfer equation of Chapter II which is performed in that context. Two different expansions of the covariance can be found, both based on the eigenvector analysis of the homogeneous covariance kernel. In the case of homogeneous fields observed over a large domain, the eigenvectors become complex exponentials and the Karhunen-Loeve expansion becomes a Fourier transform relationship (Van Trees, 1968).

Let $K_{TT}(\sigma_x,\sigma_y)$=$K_{TT}(\sigma_r,\beta)$ denote the homogeneous isobaric covariance kernel, where σ_x,σ_y are the displacements in the zonal and meridional directions and σ_r,β are the radius and azimutal angle associated with the displacement. Let us assume that the covariance kernel is bounded, differentiable, and that its derivative is piecewise continuous. The Fourier transform relationship between covariance kernel and spectrum can be written as

$$S_{TT}(k_x,k_y)=\frac{1}{\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}K_{TT}\exp(-j(k_x\sigma_x+k_y\sigma_y))d\sigma_x d\sigma_y \qquad (III.2.7)$$

which can be reduced using the symmetry of the covariance

$$S_{TT}(k_x,k_y)=\frac{1}{\pi^2}\int_{0}^{\infty}\int_{0}^{\infty}K_{TT}\cos(-j(k_x\sigma_x+k_y\sigma_y))d\sigma_x d\sigma_y \qquad (III.2.8)$$

The Fourier transform can also be expressed in polar form as

$$S_{TT}(k_h,\beta)=\frac{1}{4\pi^2}\int_{0}^{\infty}\int_{0}^{2\pi}K_{TT}(\sigma_r,\varphi)\exp(-jk_h\sigma_h\cos(\beta-\varphi))\sigma_r d\sigma_r d\varphi \quad (III.2.9)$$

The isobaric covariance kernel can be expressed as a infinite series of kernels as

$$K_{TT}(\sigma_r,\varphi)=\sum_{n=-\infty}^{\infty}K_{TTn}(\sigma_r)\exp jn\varphi \qquad (III.2.10)$$

Plugging (III.2.10) into (III.2.9) yields

$$S_{TT}(k_h,\beta)=\frac{1}{4\pi^2}\int_{0}^{\infty}\sigma_r d\sigma_r\sum_{n=-\infty}^{\infty}K_{TTn}(\sigma_r)\int_{0}^{2\pi}\exp(jn\varphi-jk_h\sigma_r\cos(\beta-\varphi))d\varphi \qquad (III.2.11)$$

Recognizing

$$\frac{1}{2\pi}\int_{0}^{2\pi}\exp(jn\varphi-jk_h\sigma_r\cos(\beta-\varphi))d\varphi=J_n(k_h\sigma_r)(-j)^n \qquad (III.212)$$

where $J_n(.)$ denotes the Bessel function of the first kind of order n, this yields

$$S_{TT}(k_h,\beta)=\frac{2}{\pi}\exp(jn\beta)(-j)^n\sum_{n=-\infty}^{\infty}\int_{0}^{\infty}K_{TTn}J_n(k_h\sigma_r)\sigma_r d\sigma_r \qquad (III.2.13)$$

In this polar expansion, the term $n = 0$ corresponds to the isotropic covariance whereas the term $n = 1$ corresponds to the dipole pattern described by Wallace and Gutzler (1981). These two patterns are not easily described by the cartesian description. For these reasons, the polar expansion is more appropriate tc described the anisotropy of atmospheric temperature fields.

III.2.2 Spectral behavior of atmospheric temperature fields

Covariance and spectrum are related through the Fourier transform relationship. Information concerning either representation of the structure can be readily be interpreted in terms of the other representation. Incorporating information in the spectral domain is especially relevant in the modeling of isobaric fields because most of the theoretical and empirical analyses of temperature turbulences are performed in that domain. As previously stated, the Fourier transform relationship can take simpler forms depending on the metric $|.|$ used.

In the case of isotropic two dimensional planar stochastic processes, the Fourier transform adopts a simpler form which corresponds to the first order approximation of the analytical series representation (Equation III.2.13). Let $K_{TT}(\sigma_x, \sigma_y, p)$denote the isobaric kernel and $S_{TT}(k_x, k_y, p)$ denote the corresponding spectrum. The two functions satisfy

$$S_{TT}(k_x, k_y, p) = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} K_{TT}(\sigma_x, \sigma_y, p) \exp(-j(k_x\sigma_x + k_y\sigma_y)) d\sigma_x d\sigma_y \qquad (III.2.14)$$

To make use of the isotropy of the covariance, let us perform the change of variable $x = r\cos\ell$ and $y = r\sin\theta$. This angle θ has nothing in common with the colatitude of the previous section. (III.2.14) thus becomes

$$S_{TT}(k_h, p) = \frac{1}{4\pi^2} \int_0^{\infty} \int_0^{2\pi} K_{TT}(\sigma, p) \exp(-jk_h\sigma\cos\theta)\sigma d\sigma d\theta \qquad (III.2.15)$$

where k_h denotes the magnitude of the horizontal wavenumber. Integrating with respect to the phase θ and recognizing that

$$\frac{1}{2\pi}\int_0^{2\pi}\exp(-jx\cos\theta)d\theta = J_0(x) \qquad (III.2.16)$$

where J_0 denotes the Bessel function of the first kind of order 0, yields

$$S_{TT}(k_h,p) = \frac{1}{2\pi}\int_0^{\infty} K_{TT}(\sigma,p)J_0(k_h\sigma)\sigma d\sigma \qquad (III.2.17)$$

The relationship between spectrum and covariance becomes a Hankel transform relationship (Watson, 1944, Mook 1983).

Traditionally, atmospheric scientists deals with the magnitude-spectrum which corresponds to the integration of the spectrum defined above for all orientations of the wavenumber k_h. The magnitude spectrum is defined by

$$E_{TT}(k_h) = S_{TT}(k_h)k_h \qquad (III.2.18)$$

Unless otherwise indicated, we will deal with the magnitude-spectrum for the remainder of this chapter. The spectrum derived from the mathematical two-dimensional Fourier transform will be referred to as the Hankel power spectrum.

The general shape of isobaric temperature magnitude spectra $E_{TT}(k_h,p)$ as well as other geoparameters spectra such as for kinetic energy or potential energy is not theoretically determined because of the complexity of the Navier Stokes equations and their non-linearities. The analytical analyses are spectrally limited because of specific approximations or "ranges" of the Navier Stokes equations. The spectrum of isobaric temperature fields has been measured in different subranges, where the subranges are determined by the measurements' resolution.

Most studies related to spectra (temporal and spatial) of atmospheric parameters are done in terms of the kinetic spectrum, which is the spectrum of the magnitude of components of the wind. Namely,

$$E_k = < \frac{1}{2}\vec{V}.\vec{V} > \qquad (III.2.19)$$

where $\vec{V}$ denotes the wind velocity and $< . >$ denotes the expectation.

The temperature spectrum is related to the spectrum of the potential energy through appropriate scaling (Lilly, 1984). That is, denoting the potential energy by E_ϕ

$$E_\phi = < \frac{1}{2} C_p (T - < T >)^2 > \qquad (III.2.20)$$

where C_p denotes the specific heat of the air at constant pressure. The potential energy presents a shape similar to the one of the kinetic spectrum, except at small wavenumbers. As it will become apparent in the following discussion of the main concepts and models for the spectrum of energy in the atmosphere, these similarities in the spectral shapes form the basis of some of the analyses.

Some of the spectral analyses of atmospheric parameters are performed in the temporal rather than in the spatial domain. The transition from one domain to the other is performed using the frozen turbulence or Taylor assumption (Taylor, 1935). This assumption states that turbulences are convected past sensing locations frozen with the wind velocity. That is, denoting the wind velocity by $\vec{V}$ and the turbulence in temperature from the mean state (mean profile) by ΔT

$$\Delta T(\vec{r}, t) = \Delta T(\vec{r} - \vec{V}t) \qquad (III.2.21)$$

If the wind velocity is assumed to be constant (or more slowly varying than any turbulence in temperature T), the temporal and spatial spectra are conformal mappings of each other of the form $k_h = \omega / U$ where U denotes the mean magnitude of the wind. This assumption fails when different wavenumber components of the temperature field are convected at different velocities (Wyngaard and Clifford, 1977).

Another important quantity in the theory of turbulence that provides information about the spectrum of temperature fields is the spectrum of the enstrophy. Enstrophy is defined as the mean square value of the vorticity, which is the microscopic measure of rotation of a fluid as defined by the curl of the wind field (Holton,

1972). The spatial derivative relationship between vorticity and wind yields the multiplication of the wind spectrum by k_h to get the vorticity spectrum, and a k_h^2 multiplication to obtain the enstrophy spectrum.

Given these different pieces of information, the general shape of temperature spectrum can be summarized as follows: The spectrum falls as the inverse third power of the horizontal wavenumber k_h (Shur, 1962) merging to a slowly decreasing range, conforming to $k_h^{-5/3}$ for larger wavenumbers (Lilly, 1984, Van Zandt, 1982). For small wavenumbers, large wavelengths, the spectrum is not well defined. This difficulty in defining the long wavelength spectrum arises from transient phenomena which take time to pass through the sensing locations used in the estimation of the spectrum (Boer and Sheperd, 1983). However, general trends can be established. The spectrum flattens somewhat as the wavelength increases, reaching a maximum, then decreases.

Figure III.2.3 (from Lilly and Petersen, 1983) presents the empirical horizontal energy spectrum from wavelengths of 5 km (5.10^{-2} Mm) to 20 Mm. This corresponds to planetary wavenumbers varying from 2 to 8000 (a planetary wavenumber corresponds to 0.156 $cycle/Mm$).

The kinetic spectrum and the isobaric temperature spectrum in the troposphere and the stratosphere follow the k_h^{-3} power law for planetary wavenumbers between 1.0 $cycle/Mm$ and 6.0 $cycle/Mm$ (Julian *et al.*, 1970, Julian and Cline, 1974, Chen and Wiin-Nielsen, 1978). The transition from the k_h^{-3} to the $k_h^{-5/3}$ regime takes place between wavenumbers of 6.0 $cycle/Mm$ and 35 $cycle/Mm$.

The lack of definition of the spectrum appears clearly at the low wavenumber where spectral gaps (often referred as humps) are present. This phenomenon yields variability in the low wavenumber part of the spectrum with respect to season and location. This is true because large scale phenomena are dominated by the topography, orography and atmospheric heat sources. Small scale components, on

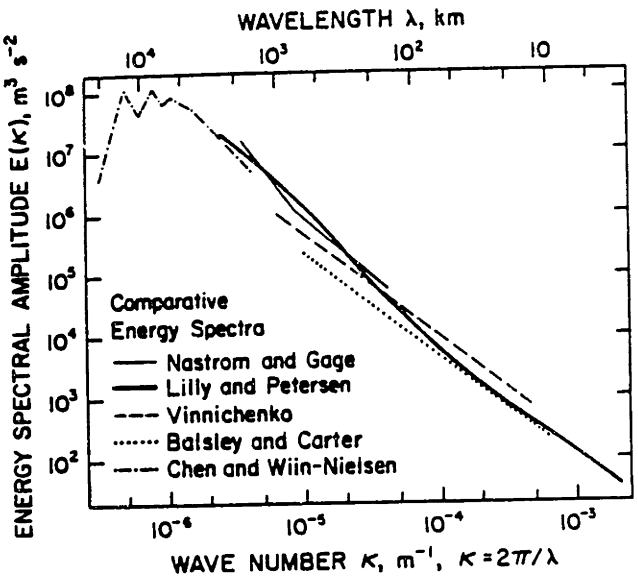


Figure III.2.3: Horizontal magnitude spectrum for energy as compiled by Lilly and Petersen (1982)

the other hand, obey turbulence theory.

Let us pause a moment before exploring the different ranges of the spectrum to consider the impact of the turbulent flow on the estimation of the spectrum and by induction, on its modeling. This analysis is similar to the predictability problems encountered in other facets of atmospheric science (Lilly, 1984).

For the purpose of illustration, let us model the flow in the mesoscale as purely turbulent with a $k_h^{-5/3}$ spectrum. The eddy turnover time, defined as the time it takes an eddy of a specific wavelength to fold on itself, can be approximated as proportional to (wavelength) $^{2/3}$ (Whittle, 1962). The eddy turnover time can also be interpreted from a statistical point of view as the amount of time required to obtain an independent measurement (VanMarcke,1984). It plays the role of the scale of fluctuation for spatial correlation (see section IV.3 for more details). For

an eddy of wavelength 50 km, the turnover time is about 7 hours, whereas for eddy of 3000 km, the turnover time increases to 4.5 days. This implies that, for a given duration over which measurements are performed, the isobaric correlation can be estimated with better accuracy (that is, with a larger number of degrees of freedom) for smaller isobaric displacement. This analysis will be more detailed in appendix IV.A.

The first range of interest in the energy spectra (kinetic and potential) is the k_h^{-3} range which is characterized by a clear separation between the vertical and horizontal directions. Two different models (through related to each other) have been advanced for the atmospheric flow.

The first is based on a two-dimensional modeling of the atmospheric flow (Kraichnan, 1967, Batchelor, 1969, Tennekes, 1978). In this model the kinetic energy is assumed to be constant and its spectral shape is taken to be time preserved. The enstrophy is constant with a constant downscale flux through its spectrum. Dimensional analysis yields a k_h^{-3} energy spectrum for wavenumbers larger than the forcing wavenumber corresponding to atmospheric baroclinic instabilities. For wavenumbers smaller than the forcing wavenumber, as the scale becomes larger, the spectrum varies as k_h^3. Lilly (1984) introduced a analytical function satisfying these two subranges. Namely

$$E(k_h) = C_o(k_m)\frac{k_h^3}{k_m^6 + k_h^6} \qquad (III.2.22)$$

Interpreted in terms of the poles and zeroes representation of linear system theory, this model corresponds to zero of order 3 at the origin and to six single-order poles at $k_m \exp(j\frac{k}{3}\pi)$ for $k = 0, 1, \ldots, 5$.

The second model that yields the same power law decay rate for the energy spectrum is a quasi-geostrophic analysis of the three-dimensional flow (Charney, 1971). In this analysis, the sum of the kinetic and the potential energy is taken to be constant. The adiabatic equation of geostrophic frictionless motion is solved,

and for the same subrange as the Kraichnan analysis, the kinetic and potential energy spectra representing wind velocity and temperature respectively decay as k_h^{-3}. (A more detailed analysis of the geostrophic flow is performed in section III.2.4). Furthermore, the energy is equally divided between the two components of the kinetic energy (azimutal and zonal) and the potential energy. The spectrum follows

$$E_{TT}(k_h) = \alpha \eta^{2/3} k_h^{-3} \qquad (III.2.23)$$

where η is the enstrophy cascade rate (Gage, 1979).

Herring (1980) performed a statistical analysis of the geostrophic turbulence and obtained a log-modified k_h^{-3} spectrum of the form

$$E_{TT}(k_h) = \frac{1}{k_h^3} (\ln k_h)^{-\frac{1}{3}} \qquad (III.2.24)$$

Herring further showed that as the scale increases, so does the tendency of the fluid to follow a two-dimensional behavior.

The second spectral subrange of interest for the energy spectra is called the inertial subrange (Kolmogorov, 1941). It is characterized by a large Reynolds number, i.e a low viscocity. In the inertial subrange, the spectrum is modeled as monotonically decreasing as a function of the wavenumber, following a (wavenumber)$^{-5/3}$ power law. That is

$$E_{TT}(k) = \alpha \epsilon^{2/3} k^{-5/3} \qquad (III.2.25)$$

where α is a scaling constant and ϵ a parameter describing the processes involved in the formation of the spectrum (see below). Strictly speaking, for the isobaric domain, this range can be separated into two subregions. The first of these regions is characterized by truly three-dimensional isotropic behavior and the wavenumber k is three-dimensional. As the wavelength increases, the vertical component of the atmospheric flow decorrelates and does not behave similarly to the isobaric component. This characteristic length varies as a function of pressure. Its value also varies

from author to author, from a few centimeters close to the ground to a few hundred meters at the tropopause. For wavelengths larger than this characteristic scale of the order of about a centimeter close to the ground and about one hectometer in the stratosphere, the isobaric spectrum remains characterized by the $k_h^{-5/3}$ behavior of the form (III.2.25) but with two directional wavenumber and a different constant α.

This power spectrum decay rate is the result of the assumption that the atmosphere is in equilibrium. That is, the energy produced in the atmosphere equals the energy dissipated. The rate of dissipation of the energy ϵ uniquely specifies the process (Batchelor, 1953). It equals the rate at which energy is transferred through the spectrum from large wavelengths into smaller wavelengths. The rate ϵ is analogous to the rate η of equation (III.2.23). This constant flux of energy explains the constant slope of the spectrum. As for the geostrophic flow subrange, the energy is two thirds kinetic and one third potential. Moreover, the wind and temperature spectra are identical.

To explain the transition to a larger decay rate at smaller wavenumbers, Lumley (1982) concentrated his analysis on the impact of the energy flux on the energy spectrum. He considered decaying isotropic turbulences at all scales and obtained a spectrum of the form:

$$E_{TT}(k_h, p) = C_o(\epsilon)(1 + (\frac{k_h}{k_b})^{-4/3})k_h^{-5/3} \qquad (III.2.26)$$

that is,

$$E_{TT}(k_h, p) = C_o(\epsilon)k_h^{-5/3} + C_o(\epsilon)k_b^3 k_h^{-3} \qquad (III.2.27)$$

In Lumsley's theory, the wavenumber k_b is related to the wavenumber corresponding to the excitation of baroclinic instabilities in the atmosphere.

The $k_h^{-5/3}$ spectrum does not yield a finite energy temperature field and moreover does not take into account the dissipation of the turbulent energy into

heat. This viscous dissipation is modeled by appending a Gaussian roll-off to the $k_h^{-5/3}$ spectrum (Tatarsky, 1941, Colavita, 1985). Namely,

$$E_{TT}(k_h) = \alpha\epsilon^{2/3}k_h^{-5/3}\exp(-k_h^2/k_m^2) \qquad (III.2.28)$$

where k_m is related to the inner scale of turbulence of the Kolgomorov spectrum. The standard deviation k_m is typically on the order of 10^9 *cycle/Mm*.

A possible spectrum shape which satisfies the models introduced for all the subranges is

$$E_{TT}(k_h) = \left(C_0(\epsilon)\frac{k_h^3 k_b^3}{k_l^6 + k_h^6} + C_1(\epsilon)\frac{k_h^3}{1 + k_h^{14/3}}\right)\exp(-k_h^2/k_m^2) \qquad (III.2.29)$$

Figure III.2.4 presents the proposed theoretical spectrum for wavenumber varying between 0.1 *cycle/Mm* and 100 *cycle/Mm* with the parameters k_b= 30 *cycle/Mm*, k_l= 1 *cycle/Mm*, k_m= 10^9 *cycle/mm*, C_0= 1 ,and C_1= 100.

III.2.3 Spatial behavior of atmospheric isobaric temperature fields

It is difficult to perform an analytical analysis of the isobaric covariance in the spatial domain. This is true because the isobaric covariance is the Fourier transform of a spectrum with different characteristic subranges. For this reason, most of the spatial analysis is empirical. A table of published spatial correlations for wind and temperature can be found in Schlatter (1979).

As for the analysis of spectra, the correlation can be examined in both the spatial and temporal domains. The transition from one domain to the other is based on the Taylor hypothesis (Taylor, 1935, Vinnichenko, 1970). Recalling equation (III.2.21),

$$\Delta T(\vec{r}, t) = \Delta T(\vec{r} - \vec{v}t) \qquad (III.2.30)$$

Assuming stationarity and homogeneity of the frozen turbulence field, the spatial covariance (the one of interest in this study) can be expressed as

$$< \Delta T(\vec{r}, t)\Delta T(\vec{r'}, t) >= K_{TT}(\vec{r} - \vec{v}t - \vec{r'} + \vec{v}t)$$

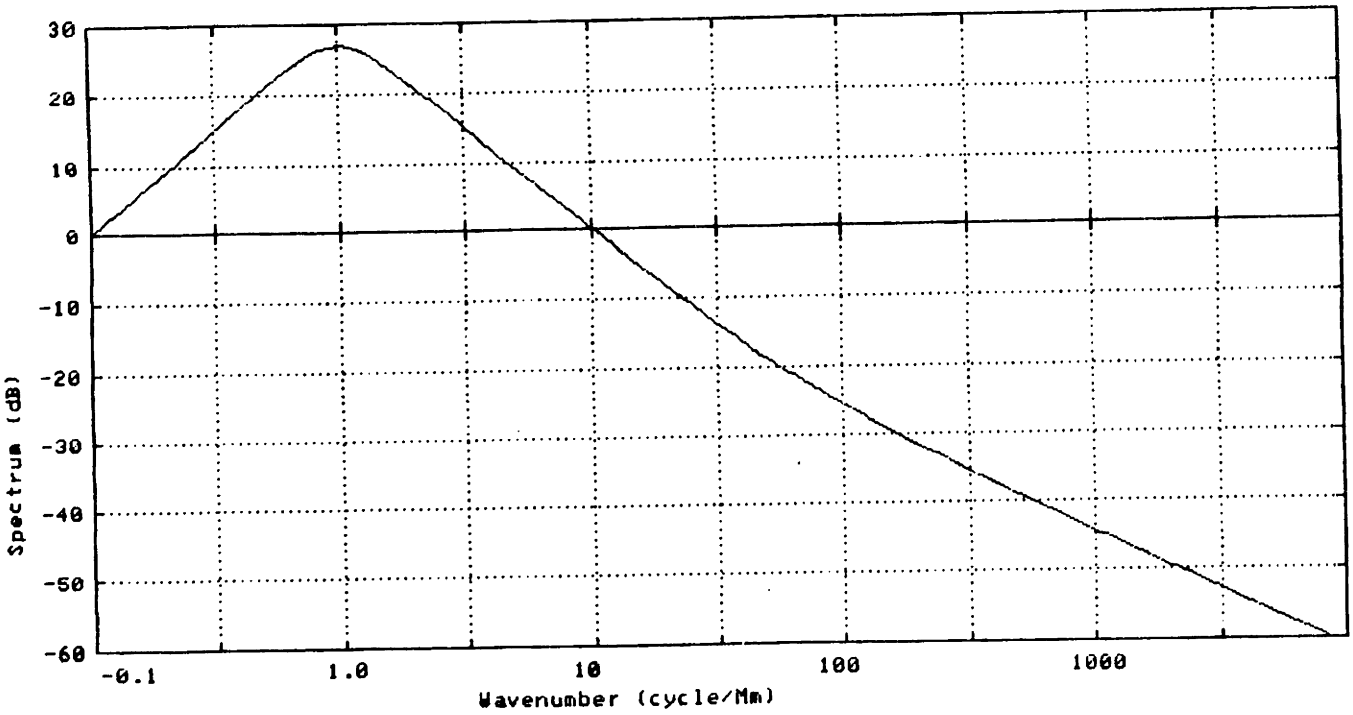


Figure III.2.4: Spectral Model (magnitude spectrum) based on the k_h^{-3} and $k_h^{-5/3}$ subranges

$$= K_{TT}(\vec{r'} - \vec{r}) \qquad (III.2.31)$$

Likewise the temporal covariance function is written as

$$< \Delta T(\vec{r}, t)\Delta T(\vec{r}, t') >= K_{TT}(\vec{r} - \vec{v}t - \vec{r} + \vec{v}t')$$

$$< \Delta T(\vec{r}, t)\Delta T(\vec{r}, t') >= K_{TT}(\vec{v}(t' - t)) \qquad (III.2.32)$$

Most of the analysis of the correlation of temperature field is performed in the spatial domain in objective analysis and data assimilation (Schlatter *et al.*, 1976, Bengtsson *et al.*, 1981) and in the time domain in forecasting (Lorenz, 1969).

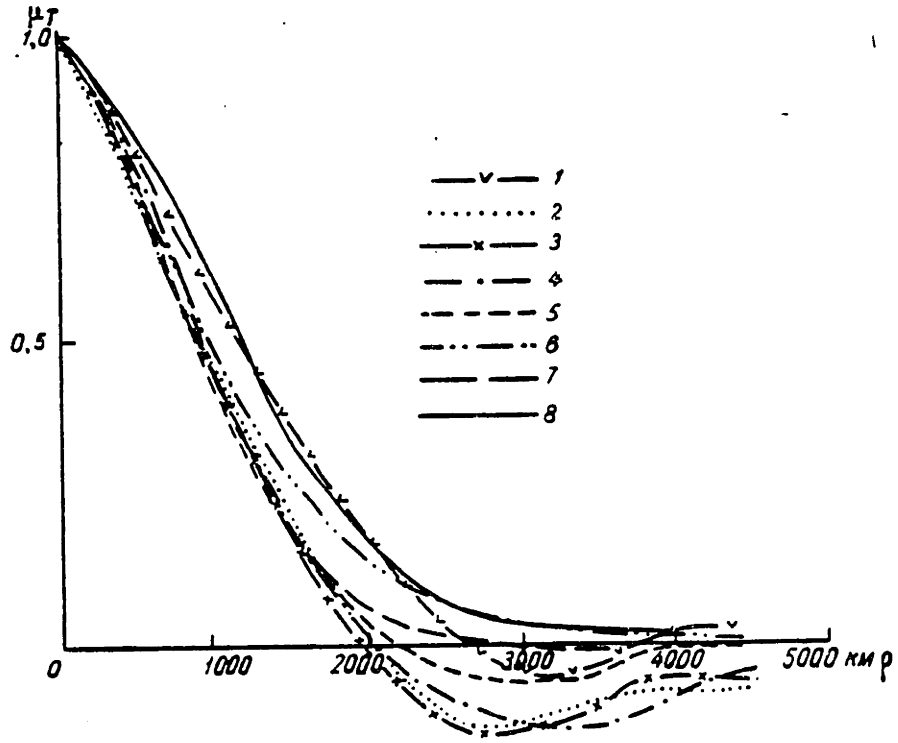


Рис. 2. Нормированные автокорреляционные функции температуры воздуха на различных уровнях. Зима.
1 — 1000 мб, *2* — 850 мб, *3* — 700 мб, *4* — 500 мб, *5* — 400 мб, *6* — 300 мб, *7* — 200 мб, *8* — 100 мб.

Figure III.2.5: Isobaric correlation for different pressure levels over Eastern Europe during the winter as a function of displacement. (from Boltenkov, 1966)

Figure III.2.5 (from Boltenkov, 1966) presents the correlation of isobaric temperature field as a function of distance for different pressure levels. This data was collected during the winter over Eastern Europe.

The general shape of these function is characteristic of decaying exponentials or of damped cosines. The decay length (defined here as the $\frac{1}{e}$ point) varies from 1.1 to 1.5 Mm. It should be noted that the term with the largest damping factor, or equivalently, the least amount of the negative correlation are at the highest pressure levels used in the analysis (300, 200, 100 $mbar$).

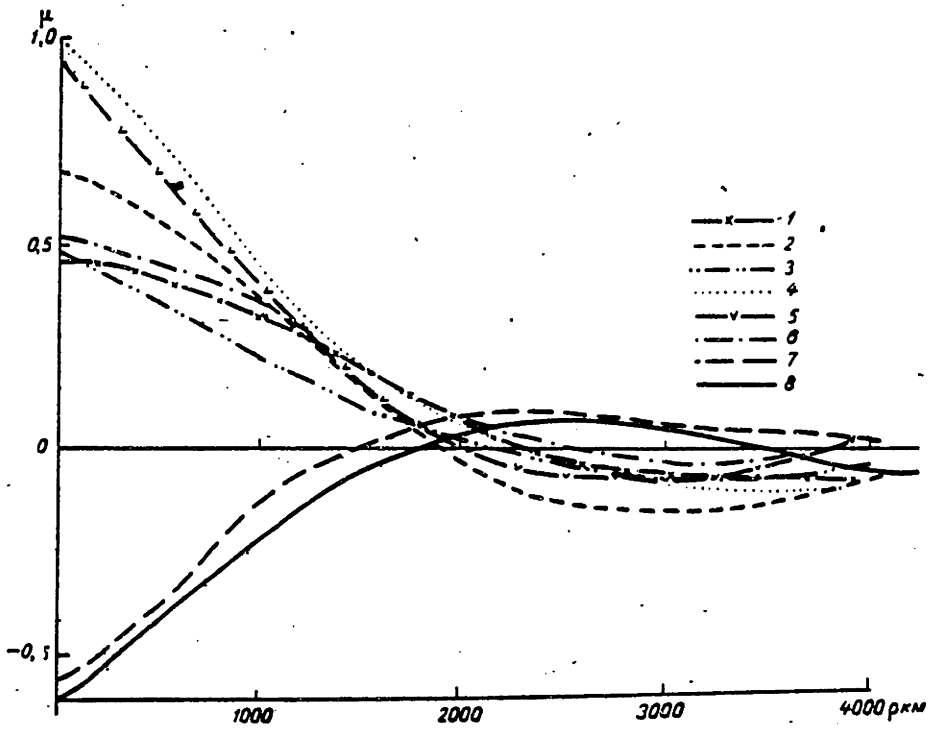


Рис. 5. Двухуровенные корреляционные функции температуры воздуха для зимнего сезона.

1 — μ_T (500, 1000, ρ), *2* — μ_T (500, 850, ρ), *3* — μ_T (500, 700, ρ), *4* — μ_T (500, 500, ρ), *5* — μ_T (500, 400, ρ), *6* — μ_T (500, 300, ρ), *7* — μ_T (500, 200, ρ), *8* — μ_T (500, 100, ρ).

Figure III.2.6: Two levels correlation of the 500 mbar pressure level with other pressure levels as a function of distance (from Boltenkov, 1966)

Figure III.2.6 (also from Boltenkov, 1966) presents the two-level correlation function for the 500 *mbar* with other pressure levels as a function of distance over the same data set.

Except for the value at the origin, the correlation follows the damped cosine behavior. The effect of the tropopause inversion can be seen for the two-levels correlations (500,200) and (500,100) as an anticorrelation in the temperature at zero displacement.

To the extent that the correlations all cross the zero axis at about the same

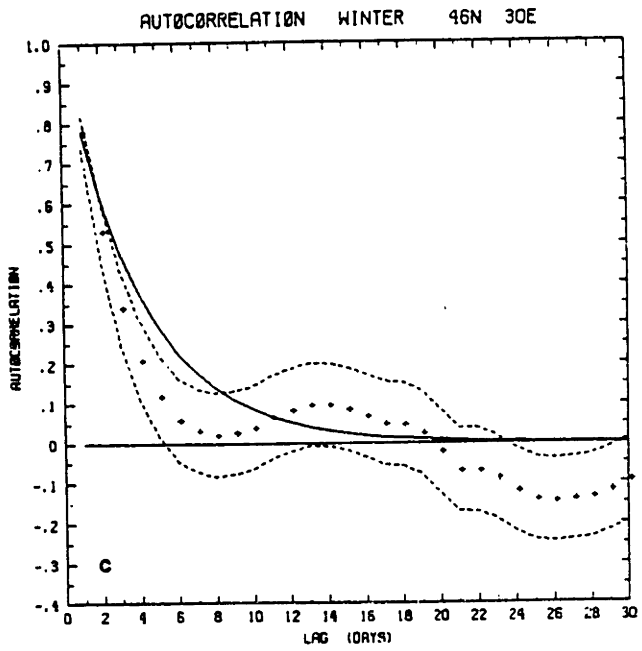


Figure III.2.7: Temporal correlation of 500 mbar height (from Gutzler and Mo, 1983).

displacement and present a minimum (or maximum for the levels placed across the tropopause) at the same displacement, Figure III.2.6 suggests a separability in the statistics.

Figure III.2.7 (from Gutzler and Mo, 1983) presents the autocorrelation function for winter of the 500 *mbar* height. The height's autocorrelation function is the same as the temperature's one if the earth atmosphere is modeled as a perfect gas (Bergmann, 1979). The measured correlation is marked by crosses and the 95 % confidence interval of the observation by dashed lines. The shape of the observed temporal correlation is the same than the previously discussed spatial correlation.

Several analytical functions have been advanced to model this correlation. Jones (1975) and Leith (1973) model the temperature as a first order autoregressive temporal Markov process which yields a decaying exponential. Second order

Markovian models have been advanced to account for the negative overshoot of the correlation. Thiebaux (1976) extended the principle of autoregressive moving average (ARMA) modeling used in time series analysis to the spatial domain by taking the displacement to be the distance between points. ARMA-derived autocorrelation functions are typically decaying exponentials or damped cosines. Extending the ARMA modeling from time to space series is, in fact, a reformulation of the Taylor approximation.

This modeling of spatial series as Markov processes is justified principally by its implications in the prediction of isobaric correlation (Thiebaux, 1975) and the design of efficient dynamic retrieval methods such as Kalman filtering (Ledsham, 1978, Ledsham and Staelin, 1981).

Other analytical functions include Gaussian functions, damped Bessel functions, and Legrendre Polynomials (Julian and Thiebaux, 1975, Rinne and Jarvenoja, 1985).

The interesting characteristic about the decaying exponential is that its Hankel transform is a 3^{rd} order Hankel spectrum. Namely, if one designates $R(s)$ the isobaric covariance and $E(k)$ its spectrum,

$$R(s) = \exp(-\alpha\sigma) \rightarrow E(k) = \frac{\alpha k}{(\alpha^2 + k^2)^{3/2}} \qquad (III.2.33)$$

which decays as $E(k) = \alpha k^{-2}$ for large wavenumbers. This spectrum has a decay rate midway between the spectrum of quasigeostrophic (enstrophy cascading) subrange. and the spectrum of the inertial (energy cascading) subrange.

III.2.4 Geostrophic Constraints

The geostrophic constraints on the isobaric temperature covariance kernel are derived from the geostrophic approximation to the laws of fluid mechanics and thermodynamics that govern the atmosphere. In this section, only a sketch of the

derivation of this approximation is presented. This presentation will be followed by a more complete (and novel) derivation of the impact of the geostrophic approximation onto the isobaric covariance kernel of atmospheric temperature fields. The interested reader is referred to books either on fluid mechanics or on physical meteorology (eg., Eskinazi, 1975) for a complete derivation and analysis of fluid mechanics equations which form the basis for the derivations of this section.

This analysis of the geostrophic approximation will start with the Navier-Stokes equation of fluid mechanics expressed as

$$\frac{\partial \vec{U}}{\partial t} + (\vec{U}.\nabla)\vec{U} + 2\vec{\Omega} \times \vec{U} = -\frac{1}{\rho}\nabla P + \vec{g} + \frac{\mu}{\rho}\nabla^2 \vec{U} \qquad (III.2.34)$$

where $\vec{U}$ is the velocity of a parcel or of the atmosphere in general, usually represented in cartesian coordinates by components in the meridional (north-south), zonal (east-west) and vertical directions, P the pressure, $\vec{g}$ the gravitational acceleration, ρ the fluid density, $\vec{\Omega}$ the angular velocity vector associated with the rotation of the earth, and μ the dynamic viscocity of the atmosphere (a measure of the friction in the motion of the geofluid). The viscocity varies from typically 1.8 $kg/ms.$ at the ground level to 1.4 $kg/m.s$ at an altitude of 20 km.

(III.2.34) can be recognized as the expression of Newton's inertial law for a differential element of volume. The term $\partial\vec{U}/\partial t$ is the change of velocity of the atmosphere at a specific location in space, $(\overline{U}.\nabla)\vec{U}$ is the convective part of the change of velocity. It represents the change in velocity as the atmosphere moves from one point in space to another. The term $2\overline{\Omega} \times \overline{U}$ is the Coriolis acceleration caused by the motion of the atmosphere as referred to the earth. The right hand side of (III.2.34) consists of the forces applied to the differential volume of the atmosphere. $\vec{g}$ is the gravitational acceleration and $1/\rho\nabla P$ is the gradient of pressure. This corresponds to the divergence (Eskinazi, 1975) of the surface stress vector (surface forces) applied on the side of the elementary volume of fluid. The last term $\frac{\mu}{\rho}\nabla^2\vec{U}$ is

the shearing stress which represents the friction introduced by density stratification and surface.

Close to the surface, the air flow is turbulent and frictional forces cause sharp changes in velocity. Above this boundary layer, whose thickness depends on the wind speed (of the order of about 1000 m, pressure of 900 *mbar*), fluid motion is free of viscous and turbulent shear. In the absence of these shearing stresses, the Navier-Stokes equation (III.2.34) becomes the Euler equation

$$\frac{\partial \vec{U}}{\partial t} + (\vec{U}.\nabla)\overline{U} + 2\vec{\Omega} \times \vec{U} = -\frac{1}{\rho}\nabla P + \vec{g} \qquad (III.2.35)$$

The latter equation describes fluid mechanics dominated by the motion of the earth which is therefore *geostrophic*. Neglecting the acceleration term in (III.2.35) yields a steady air flow and thus the simplified geostrophic equation of fluid mechanics,

$$2\vec{\Omega} \times \vec{U} = -\frac{1}{\rho}\nabla P + \vec{g} \qquad (III.2.36)$$

This steady state analysis of geostrophic fluid mechanics is an approximation whose validity is restricted especially at the tropics and to a lesser extent at mid-latitudes (Schlatter *et al.*, 1976).

Applying (III.2.36) to the horizontal components of the wind yields, after vector multiplication,

$$fV = \frac{1}{\rho}\frac{\partial P}{\partial y} \qquad fU = -\frac{1}{\rho}\frac{\partial P}{\partial x} \qquad g = -\frac{1}{\rho}\frac{\partial P}{\partial z} \qquad (III.2.37)$$

where $f = 2|\Omega| \sin\psi$ is the Coriolis parameter which depends on the latitude ψ. The introduction of the atmospheric temperature is performed with the introduction of the perfect gas law. Thus

$$P = \rho RT \qquad (III.2.38)$$

Plugging (III.2.38) into (III.2.37) yields

$$fV = \frac{RT}{P}\frac{\partial P}{\partial x}, \qquad fU = -\frac{RT}{P}\frac{\partial P}{\partial y} \qquad (III.2.39)$$

Taking the differential element of (III.2.37) yields

$$f\partial V = \frac{R\partial T}{P}\frac{\partial P}{\partial x}, \qquad f\partial U = -\frac{R\partial T}{P}\frac{\partial P}{\partial y} \tag{III.2.40}$$

and finally after regrouping the variables

$$\begin{aligned}\left(\frac{\partial u}{\partial p}\right) &= \frac{R}{fp}\frac{\partial T}{\partial y}\\ \left(\frac{\partial v}{\partial p}\right) &= \frac{R}{fp}\frac{\partial T}{\partial x}\end{aligned} \tag{III.2.41}$$

At the equator, the coriolis parameter tends to zero and the left hand side of (III.2.41) becomes infinite. As stated previously, the geostrophic approximation is a poor one at the equator. To compensate for this discrepancy between the value predicted by (III.2.41) and observations, the right hand side of (III.2.41) is multiplied by a "coefficient of geostrophy" (Bergmann, 1979) whose variations with latitude are empirical. Therefore,

$$\begin{aligned}\left(\frac{\partial u}{\partial p}\right) &= \frac{GR}{fp}\frac{\partial T}{\partial y}\\ \left(\frac{\partial v}{\partial p}\right) &= \frac{GR}{fp}\frac{\partial T}{\partial x}\end{aligned} \tag{III.2.42}$$

These two relationships are often referred to as the wind temperature relationships (Bergmann, 1979), (Thiebaux, 1975).

Let us determine the impact of these geostrophic relationships between the meridional and zonal components of the wind and temperature on the isobaric covariance kernel of the temperature field $K^i_{TT}(\sigma,p)$ by analyzing the relationships between wind and temperature covariance kernels.

Applying (III.2.42) to two points i and j in the atmosphere, referred to by the triplet (x,y,p) where x and y are isobaric displacements in the meridional and the zonal direction and p the pressure yields

$$\begin{aligned}\frac{\partial^2}{\partial \ln p_i \partial \ln p_j}(u_i u_j) &= \frac{G_i G_j R^2}{f_i f_j}\frac{\partial^2}{\partial y_i \partial y_j}(T_i T_j)\\ \frac{\partial^2}{\partial \ln p_i \partial \ln p_j}(v_i v_j) &= \frac{G_i G_j R^2}{f_i f_j}\frac{\partial^2}{\partial x_i \partial x_j}(T_i T_j)\end{aligned} \tag{III.2.43}$$

Taking the expected values of equations (III.2.42) and (III.2.43), after noting that partial derivation and expectation are linear operators and therefore commute with each other, yields the following set of relationships between the first and second moments of the temperature and horizontal wind fields,

$$\left(\frac{\partial \overline{u}}{\partial \ln p}\right) = \frac{GR}{fp}\frac{\partial \overline{T}}{\partial y}$$
$$\left(\frac{\partial \overline{v}}{\partial \ln p}\right) = \frac{GR}{fp}\frac{\partial \overline{T}}{\partial x} \qquad (III.2.44)$$

and

$$\frac{\partial^2}{\partial \ln p_i \partial \ln p_j}(\overline{u_i u_j}) = \frac{G_i G_j R^2}{f_i f_j}\frac{\partial^2}{\partial y_i \partial y_j}(\overline{T_i T_j})$$
$$\frac{\partial^2}{\partial \ln p_i \partial \ln p_j}(\overline{v_i v_j}) = \frac{G_i G_j R^2}{f_i f_j}\frac{\partial^2}{\partial x_i \partial x_j}(\overline{T_i T_j}) \qquad (III.2.45)$$

where the barred quantities denote the expected values over all possible realizations of a specific climatology.

Subtracting (III.2.44) from (III.2.45) and setting identical pressures for the two points i and j yields

$$\frac{\partial^2}{\partial(\ln p)^2}K^i_{vv}(x_i - x_j, y_i - y_j; p) = \frac{G_i G_j R^2}{f_i f_j}\frac{\partial^2}{\partial x_i \partial x_j}K^i_{TT}(x_i - x_j, y_i - y_j; p) \qquad (III.2.46)$$

As stated in section III.2.2 on correlation isopleths, the isobaric temperature kernel can be modeled as isotropic (after conformal mapping for elliptical correlation). (III.2.46) can then be rewritten introducing the isotropic displacement between points i and j, σ_{ij}. Assuming that the appropriate scalings have been performed on the zonal and the meridional displacements, the isotropic displacement is

$$\sigma_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \qquad (III.2.47)$$

The right hand side of (III.2.46) is a scaled version of the second derivative with respect to the zonal direction of the temperature isobaric covariance kernel. This derivative can be computed using the chain rule. That is,

$$\frac{\partial^2}{\partial x_i \partial x_j}K^i_{TT}(\sigma_{ij}; p) = \frac{\partial \sigma_{ij}}{\partial x_i}\frac{\partial}{\partial \sigma_{ij}}\left(\frac{\partial \sigma_{ij}}{\partial x_j}\frac{\partial}{\partial \sigma_{ij}}K^i_{TT}(\sigma_{ij}; p)\right) \qquad (III.2.48)$$

Applying the standard rules for the computation of derivatives yields

$$\frac{\partial^2}{\partial x_i \partial x_j} K^i_{TT}(\sigma_{ij};p) = \frac{\partial \sigma_{ij}}{\partial x_i}\left[\frac{\partial}{\partial \sigma_{ij}}\left(\frac{\partial \sigma_{ij}}{\partial x_j}\right)\frac{\partial K^i_{TT}(\sigma_{ij};p)}{\partial \sigma_{ij}}\right] + \frac{\partial \sigma_{ij}}{\partial x_i}\frac{\partial \sigma_{ij}}{\partial x_j}\frac{\partial^2 K^i_{TT}(\sigma_{ij};p)}{\partial \sigma^2_{ij}} \qquad (III.2.49)$$

and

$$\frac{\partial^2}{\partial x_i \partial x_j} K^i_{TT}(\sigma_{ij};p) = \frac{\partial^2 \sigma_{ij}}{\partial x_i \partial x_j}\frac{\partial K^i_{TT}(\sigma_{ij};p)}{\partial \sigma_{ij}} + \frac{\partial \sigma_{ij}}{\partial x_i}\frac{\partial \sigma_{ij}}{\partial x_j}\frac{\partial^2 K^i_{TT}(\sigma_{ij};p)}{\partial \sigma^2_{ij}} \qquad (III.2.50)$$

The latter relationship will be satisfied without any problem except for small displacements, in the vicinity of $\sigma_{ij} = 0$. To analyze the behavior of the isobaric covariance kernel in that vicinity, one must compute the partial derivatives of the isobaric displacement with respect to the rectangular coordinates. Namely,

$$\begin{aligned} \frac{\partial \sigma_{ij}}{\partial x_i} &= (x_i - x_j)\sigma_{ij}^{-1} \\ \frac{\partial \sigma_{ij}}{\partial x_j} &= -(x_i - x_j)\sigma_{ij}^{-1} \qquad (III.2.51) \\ \frac{\partial^2 \sigma_{ij}}{\partial x_i \partial x_j} &= ((x_i - x_j)^2 - \sigma_{ij}^2)\sigma_{ij}^{-3} \end{aligned}$$

Plugging (III.2.51) into (III.2.50) yields, dropping the ij index in the isobaric displacement

$$\frac{\partial^2}{\partial x_i \partial x_j} K^i_{TT}(\sigma;p) = -\frac{1}{\sigma}\frac{\partial K^i_{TT}(\sigma;p)}{\partial \sigma} + \left(\frac{(x_i - x_j)^2}{\sigma^2}\right)\left[\frac{\partial^2 K^i_{TT}(\sigma;p)}{\partial \sigma^2} + \frac{1}{\sigma}\frac{\partial K^i_{TT}(\sigma;p)}{\partial \sigma}\right] \qquad (III.2.52)$$

The wind field has finite energy if the derivative of the isobaric temperature covariance kernel is finite. The temperature covariance kernel must thus satisfy

$$\lim_{\sigma \to 0} \frac{1}{\sigma}\frac{\partial K^i_{TT}(\sigma;p)}{\partial \sigma} \neq \infty \qquad (III.2.53)$$

and

$$\lim_{\sigma \to 0}\left(\frac{1}{\sigma}\frac{\partial K^i_{TT}(\sigma;p)}{\partial \sigma} - \frac{\partial^2 K^i_{TT}(\sigma;p)}{\partial \sigma^2}\right) = 0 \qquad (III.2.54)$$

These two conditions, often referred to as the geostrophic constraints in the objective analysis literature, must be met in order for the computation of wind energy from the isobaric covariance kernel of temperature fields to be valid (Julian and Thiebaux, 1975). The second constraint can be rewritten in a more compact form, noting that if λ is the limit of (III.2.54)

$$\begin{aligned} \frac{\lambda}{\sigma} &= \frac{1}{\sigma^2}\frac{\partial K^i_{TT}(\sigma;p)}{\partial\sigma} - \frac{1}{\sigma}\frac{\partial^2 K^i_{TT}(\sigma;p)}{\partial\sigma^2} \\ \frac{\lambda}{\sigma} &= \frac{\partial}{\partial\sigma}\left(-\frac{1}{\sigma}\frac{\partial K^i_{TT}(\sigma;p)}{\partial\sigma}\right) \end{aligned} \qquad (III.2.55)$$

The geostrophic constraints expressed in the spatial domain become

$$\lim_{\sigma\to 0}\frac{1}{\sigma}\frac{\partial K^i_{TT}(\sigma;p)}{\partial\sigma} \neq \infty \qquad (III.2.56)$$

$$\lim_{\sigma\to 0}\sigma\frac{\partial}{\partial\sigma}\left(\frac{1}{\sigma}\frac{\partial K^i_{TT}(\sigma;p)}{\partial\sigma}\right) = 0 \qquad (III.2.57)$$

In order to gain more insight on the physical significance of these geostrophic constraints for the planar geometry and to further simplify their expression, let us express them in the horizontal wavenumber domain. The isobaric covariance kernel can be expressed as the inverse Hankel transform of its Hankel spectrum (Mook, 1983). Within a scaling factor,

$$K^i_{TT}(\sigma;p) = \int_0^\infty S^i_{TT}(k;p)J_0(k\sigma)kdk = \int_0^\infty E^i_{TT}(k;p)J_0(k\sigma)dk \qquad (III.2.58)$$

Taking the derivative of the covariance kernel with respect to the displacement under the integration sign yields the set of equations

$$\frac{1}{\sigma}\frac{\partial K^i_{TT}(\sigma;p)}{\partial\sigma} = \int_0^\infty S^i_{TT}(k;p)\frac{J_1(k\sigma)}{\sigma}k^2dk \qquad (III.2.59)$$

and

$$\begin{aligned} \sigma\frac{\partial}{\partial\sigma}\left(\frac{1}{\sigma}\frac{\partial K^i_{TT}(\sigma;p)}{\partial\sigma}\right) &= \int_0^\infty S^i_{TT}(k;p)\left(J_0(k\sigma) - 2\frac{J_1(k\sigma)}{k\sigma}\right)k^3dk \\ &= \int_0^\infty E^i_{TT}(k;p)\left(J_0(k\sigma) - 2\frac{J_1(k\sigma)}{k\sigma}\right)k^3dk \end{aligned} \qquad (III.2.60)$$

The geostrophic constraints can be expressed in terms of the spectrum by taking the limits of (III.2.59) and (III.2.60) as σ approaches zero. The Taylor expansions of the zero and first order Bessel functions are (Gradshteyn and Ryzhik, 1980)

$$J_0(k\sigma) = \sum_{i=0}^{\infty} \frac{(-\sigma^2 k^2/4)^i}{i!i!} \qquad (III.2.61)$$

and

$$J_1(k\sigma) = \frac{1}{2} k\sigma \sum_{i=0}^{\infty} \frac{(-k^2\sigma^2/4)^i}{i!(i+1)!} \qquad (III.2.62)$$

Thus,

$$\lim_{\sigma \to 0} \frac{J_1(k\sigma)}{\sigma} = \frac{1}{2} k \qquad (III.2.63)$$

and

$$\begin{aligned} J_0(k\sigma) - 2\frac{J_1(k\sigma)}{k\sigma} &= \sum_{i=0}^{\infty} \frac{(-\sigma^2 k^2)^i}{i!i!} - \sum_{i=0}^{\infty} \frac{(-\sigma^2 k^2)^i}{i!(i-1)!} \\ &= \sum_{i=1}^{\infty} \frac{(-\sigma^2 k^2)^i}{(i-1)!(i+1)!} \end{aligned} \qquad (III.2.64)$$

This implies that

$$\lim_{\sigma \to 0} \left(J_0(k\sigma) - 2\frac{J_1(k\sigma)}{k\sigma} \right) = 0 \qquad (III.2.65)$$

This last equality shows that if one considers equation (III.2.60) to be an integral equation with its kernel being the difference between the two Bessel functions, and provided that the spectrum $S_{TT}(k;p)$ is integrable, then the second integral equation is redundant with the first constraint equation.

The geostrophic constraint (now unique) can finally be expressed as

$$\lim_{\sigma \to 0} \frac{1}{\sigma} \frac{\partial K^i_{TT}(\sigma;p)}{\partial \sigma} = -\frac{1}{2} \int_0^{\infty} S^i_{TT}(k;p) k^3 dk \neq \infty \qquad (III.2.66)$$

In order to insure the convergence of (III.2.66), the integrand must decay toward zero faster than the inverse power of the wavenumber for large arguments of the wavenumber. Thus, the power Hankel spectrum of the isobaric temperature fields should decay at a rate faster than the inverse fourth power of the wavenumber, for

wavenumbers larger than a specific value. That is the magnitude spectrum must decay at a rate faster than the inverse third power of the wavenumber. Similarly, for small wavenumbers (large spatial wavelengths), the spectrum must follow a power law of order larger than the inverse fourth power of the wavenumber.

The above analysis demonstrates the redundancy of the geostrophic constraints when expressed in the wavenumber domain. Let us now demonstrate this redundancy property in the spatial domain. Consider a power law expansion of the isobaric covariance kernel close to the origin. In the neighborhood of 0,

$$K^i_{TT}(\sigma;p) = a_0 + a_1\sigma^\alpha \qquad (III.2.67)$$

where a_0, a_1 and α are functions of pressure. The first geostrophic constraint (III.2.56) imposes that

$$\lim_{\sigma\to 0} \frac{1}{\sigma}\frac{\partial K^i_{TT}(\sigma;p)}{\partial\sigma} = \alpha_1\sigma^{\alpha-2} \neq \infty \qquad (III.2.68)$$

Thus the exponent α has to be greater than or equal to two. Using the same power law expansion, the second geostrophic constraint becomes

$$\lim_{\sigma\to 0} \sigma\frac{\partial}{\partial\sigma}\left(\frac{1}{\sigma}\frac{\partial K^i_{TT}(\sigma;p)}{\partial\sigma}\right) = \alpha(\alpha-2)\sigma^{\alpha-2} = 0 \qquad (III.2.69)$$

The second condition will be met by any value of the exponent α that satisfies the first condition. Equation (III.2.69) shows therefore that the two geostrophic constraints are indeed redundant.

The analysis of the geostrophic constraint in the wavenumber domain provides the important conclusion that a k_h^{-3} or a $k_h^{-5/3}$ power law for isobaric temperature field spectrum and the geostrophic condition are incompatible with each other, unless some accommodation is provided. This accommodation must preserve the spectral behavior of temperature fields at wavelengths of interest and at the same time prevent the derivative of the isobaric covariance kernel from becoming infinite.

Components of the spectrum do not, past a certain wavenumber, contribute to the covariance at small displacements. Since the purpose of the intended modeling is limited to distances larger than 50 km (for satellites with good resolution such as AMSU), the high frequency spectrum (amplitude and decay rate) is not important. Low pass filtering the temperature field will then guarantee the compatibility of the spectral decay rate and the geostrophic estimates of wind fields.

Let $K^i_{TT}(\sigma)$ denote the true (or the best possible model of) isobaric covariance kernel and $L(\sigma)$ the point spread function of the low-pass filter. The point spread function of the filtered temperature field is given, as for one dimensional signals, by

$$K^i_{T_l T_l}(\sigma) = K^i_{TT}(\sigma) * L(\sigma) * L(\sigma) \qquad (III.2.70)$$

The asterisk denote the convolution.

The ideal low pass filter with cutoff frequency k_o and unit area has a point spread function of

$$\frac{1}{\pi k_o^2} \int_0^{k_o} k J_0(k\sigma) dk = \frac{1}{\pi k_o \sigma} J_1(\sigma k_o) \qquad (III.2.71)$$

Although it eliminates all convergence problems, this point spread function is not adequate for analytical convolution (Petiau, 1955). Furthermore, it can tail in cases of perfect observations which are known to be ill-behaved in the presence of modeling errors (Van Trees, 1968).

An alternate low-pass filter can be considered which agrees with the modeling developed in turbulence theory to account for the roll-off of spectra at high wavenumbers (Kolmogorov, 1941, Tatarsky, 1971) where an exponential roll off of the spectrum of the form $\exp(-\alpha k^2)$ is assumed for k greater than some value.

One can accomplish this exponential roll-off by filtering the temperature field through a series of filters whose frequency response is a Gaussian function whose variance is larger than any wavenumber of interest. The point spread of such

filters are narrow Gaussian impulses. Since the analysis of the geostrophic condition(s) have been traditionally performed in the spatial domain, let us compute the derivative of the resulting isobaric covariance kernel in that domain. The results of derivation of convolution products extend straight-forwardly from one dimension to two dimensions and by rotation of the Jacobian to two-dimensional isotropic convolution. Recall (Ibragimov and Rozanov, 1978) that if the function f is summable and if the function g is continuously differentiable and admits a bounded derivative, then

$$(f(x) * g(x))' = f(x) * g(x)' \qquad (III.2.72)$$

Thus

$$\frac{\partial}{\partial\sigma} K^i_{T_l T_l}(\sigma;p) = K^i_{TT}(\sigma;p) * \frac{\partial}{\partial\sigma} L(\sigma) * L(-\sigma) \qquad (III.2.73)$$

The geostrophic constraint is now expressed in terms of the point spread function of the low pass filter $L(\sigma)$. Taking the low pass filter to be a Gaussian function. That is,

$$L(\sigma) = \exp(-a\sigma^2) \qquad (III.2.74)$$

and recognizing that $\frac{1}{\sigma}\frac{\partial}{\partial\sigma} = 2\frac{\partial}{\partial(\sigma)^2}$ yields

$$\frac{1}{\sigma}\frac{\partial}{\partial\sigma} K^i_{T_l T_l}(\sigma;p) = -2aK^i_{TT}(\sigma;p) * L(\sigma) * L(\sigma) \qquad (III.2.75)$$

which is finite as σ tends towards zero.

Let us note that in the above derivation no assumptions were made on the specific form of the true spectrum $K^i_{TT}(\sigma;p)$. This means that the geostrophic constraint can be satisfied for any polynomial spectrum as long as a low pass filter eliminates the high wavenumber components.

III.2.5 Positive Definiteness of Isobaric Covariance Kernel

The positive definiteness of the isobaric covariance kernel is necessary to avoid pathological situations that yield random variables with negative covariance,

design equations for retrieval operator with none or an infinite number of optimal solutions, or negative power spectrum representations. Negative power spectra have been shown to degrade the quality of assimilation methods based on a description of the covariance of temperature fields (Gandin, 1963, Rutherford, 1972).

The isobaric covariance kernel is positive definite if for any arbitrary set of weights w_k and locations (x_k, y_k), the variable $\overline{T}$ defined as

$$\overline{T} = \sum_{k=0}^{N-1} w_k T(x_k, y_k; p) \qquad (III.2.76)$$

has a positive variance. This implies that

$$\sum_{k=0}^{N-1} \sum_{l=0}^{N-1} w_k K^i_{TT}(x_k - x_l, y_k - y_l; p) w_l \geq 0 \qquad (III.2.77)$$

Bert *et al.* (1984) showed than the weights can be constrained to be signed integers and still guarantee the positivity of equation (III.2.77) for all weights w_i. This property is practical in instances where only an empirical verification of the positive definiteness is possible because its restricts the set of values to be tested.

A lower bound of the isobaric covariance kernel for two-dimensional isotropic temperature fields that is different from that of one-dimensional fields can be found by considering considering the variance of the sum of temperatures at three points (VanMarcke, 1983). These three points are taken to be on a circle placed at 120 degrees from each other. Using the isotropy of the covariance kernel, the inequality (III.2.77) becomes

$$3K^i_{TT}(0; p) + 6K^i_{TT}(\sigma; p) \geq 0 \qquad (III.2.78)$$

or

$$K^i_{TT}(\sigma; p) \geq -\frac{1}{2} K^i_{TT}(0; p) \qquad (III.2.79)$$

.

This bound is tighter than the one encountered in one dimensional stochastic processes (Wosencraft and Jacobs, 1965) where the covariance kernel can be as small as the opposite of the signal variance. This tighter bound can thus be used to test the isotropy of a two dimensional stochastic process.

As for one dimensional signals, the positive definiteness of the covariance kernel also implies that the Fourier transform of the temperature field covariance kernel be positive. A stronger statement about the positive definiteness can be made by invoking the Bochner theorem (Bochner, 1955, Parsen, 1961). Let Ξ denote the sum

$$\Xi = \sum_i \sum_j w_i w_j K^i_{TT}(x_i - x_j, y_i - y_j; p) \qquad (III.2.80)$$

Expressing the isobaric covariance kernel as the two dimensional Fourier transform of its spectrum yields, after inverting the order of summation,

$$\Xi = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sum_{i,j} w_i w_j \exp(-jk_x(x_i - x_j) - jk_y(y_i - y_j)) S^i_{TT}(k_x, k_y) dk_x dk_y \qquad (III.2.81)$$

after ordering of the exponents

$$\Xi = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Big(\sum_i w_i \exp(-j(k_x x_i + k_y y_i))\Big) \Big(\sum_i w_i \exp(+j(k_x x_i + k_y y_i))\Big) S^i_{TT}(k_x, k_y; p) dk_x dk_y \qquad (III.2.82)$$

Introducing the Fourier transform of the weighting function $W(k_x, k_y)$, one can write

$$\begin{aligned} \Xi &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} W(k_x, k_y) W^*(k_x, k_y) S^i_{TT}(k_x, k_y; p) dk_x dk_y \\ &= \int_{-\infty}^{\infty} |W(k_x, k_y)|^2 S^i_{TT}(k_x, k_y; p) dk_x dk_y \end{aligned} \qquad (III.2.83)$$

Because the weighting function can be chosen at will, the magnitude of the frequency response $|W(k_x, k_y; p)|^2$ is an arbitrarily positive function. This shows the

intimate relationship between positive definiteness of the covariance kernel and positive spectrum (Bochner, 1955).

One can extend Bochner's result to two-dimensional isotropic processes by taking the spectrum $S^i_{TT}(k_x, k_y; p)$ to be isotropic. It also provides a tighter lower bound for the normalized covariance.

Making the change of variables from (x_i, y_i) to (k_i, ϕ_i) and from (k_x, k_y) to (k, θ), equation (III.2.83) can be rewritten as

$$\Xi = \int_0^\infty \int_0^{2\pi} |\sum_j W_j \exp(jkr_j \cos(\theta - \phi_j))|^2 S^i_{TT}(k; p) k dk \qquad (III.2.84)$$

Integrating the positive weighting function with respect to θ yields the alternate form

$$\Xi = \int_0^\infty \sum_{ij} w_i w_j J_0(kr_{ij}) S^i_{TT}(k; p) k dk \qquad (III.2.85)$$

This result shows that a necessary and sufficient condition for a two dimensional isotropic kernel to be positive definite is for its Hankel transform to be positive.

The lower bound on the normalized covariance is found by noting that in order for the covariance to be as small as possible, the spectrum has to be a single narrow impulse. The corresponding autocorrelation function is a zeroth order Bessel function whose minimum value is approximately -0.403. This implies that

$$K^i_{TT}(0; p) \geq K^i_{TT}(\sigma; p) \geq -0.403 K^i_{TT}(0; p) \qquad (III.2.86)$$

II.2.6 Moment constraints on the isobaric covariance kernel

Moment constraints are defined in terms of the Fourier transform relationship between spectrum and covariance and assure the definiteness of the spectrum at the wavenumber origin. The kl^{th} moment of the isobaric two-dimensional covariance kernel is defined as (VanMarcke, 1979)

$$\varsigma_{kl} = \int_{-\infty}^\infty \int_{-\infty}^\infty \sigma_x^k \sigma_y^l K^i_{TT}(\sigma ax, \sigma_y; p) d\sigma_x d\sigma_y \qquad (III.2.87)$$

This definiteness is analogous to the definiteness of moments in probability and statistics. These moments are important when studying the spectrum near the origin.

Let us consider a Taylor expansion of the spectrum in a neighborhood of the origin.

$$S^i_{TT}(k_x, k_y) = a_o + a_{1x}k_x + a_{1y}k_y + a_{2xx}k_x^2 + a_{2xy}k_xk_y + a_{2yy}k_yk_y + \ldots \quad (III.2.88)$$

The elements of the expansion can be expressed as a function of the moments of the covariance kernel. Namely

$$\begin{aligned} a_0 &= \frac{\varsigma_{00}}{4\pi^2} \\ a_{2xx} &= -\frac{1}{4\pi^2}\varsigma_{02} \\ a_{2xy} &= \frac{\varsigma_{11}}{4\pi^2} \\ a_{2yy} &= -\frac{1}{4\pi^2}\varsigma_{20} \end{aligned} \quad (III.2.89)$$

For isotropic covariances, both a_{1x} and a_{1y} vanish at the origin due to symmetry. This implies that the spectrum has a local extremum at the origin. Likewise a_{2xy} goes to zero and the expansion becomes

$$S^p_{TT}(k_x, k_y; p) = \frac{1}{4\pi^2}\int_0^\infty \sigma K^p_{TT}(\sigma)d\sigma - \frac{1}{4\pi^2}\int_0^\infty \sigma^3 K^p_{TT}(\sigma)d\sigma(k_x^2 + k_y^2) + \ldots \quad (III.2.90)$$

For the spectrum to be valid, the covariance kernel must satisfy

$$\int_0^\infty \sigma K^p_{TT}(\sigma)d\sigma > 0 \qquad \text{finite} \quad (III.2.91)$$

and

$$\int_0^\infty \sigma^3 K^p_{TT}(\sigma)d\sigma \ \text{ finite} \quad (III.2.92)$$

Let us note that the second condition is similar to the geostrophic condition except that the integration being considered takes place in the spatial domain. The positiveness of the first moment of the covariance kernel is guarantee for positive definite

kernels since it corresponds to the DC value of the spectrum. However, from a real analysis point of view, for the Hankel transform to be defined, the integral defining the first moment must be finite.

III.3 Physical and mathematical constraints on the pressure covariance kernel

The analysis of the temperature structure in the vertical direction is less extensive (in views of the published literature) than in the horizontal direction. One of the reasons for this disproportion is that most objective analysis integration schemes are performed at constant pressure levels (Bergmann, 1979).

Analyses of phenomena involved in the evolution of temperature fields based on the equations of fluid dynamics and thermodynamics have been performed (Penk, 1965, Palmen and Newton, 1969, Charney, 1973). Unfortunately, these analyses did not provide covariances or spectra of temperatures.

One can derive conclusions on the vertical covariance by analyzing simplified descriptions of the physical phenomena involved in the evolution of temperature fields. Moreover, one can solve the problem of discretization in the vertical direction. This discretization is required to calculate the brightness temperature through the radiative transfer equation. It is justified by the high degree of correlation of the temperature field over short displacements in the vertical direction and by the gravitational stratification of the atmosphere.

III.3.1 Vertical Structure of Temperature Fields

In the vertical direction (this is the major difference with the isobaric direction) the medium, whose kinetic temperature we examine, changes characteristics. The presence of a boundary condition at high pressure levels (earth), the interaction with the ground for those locations close to it, and the absorption of solar radia-

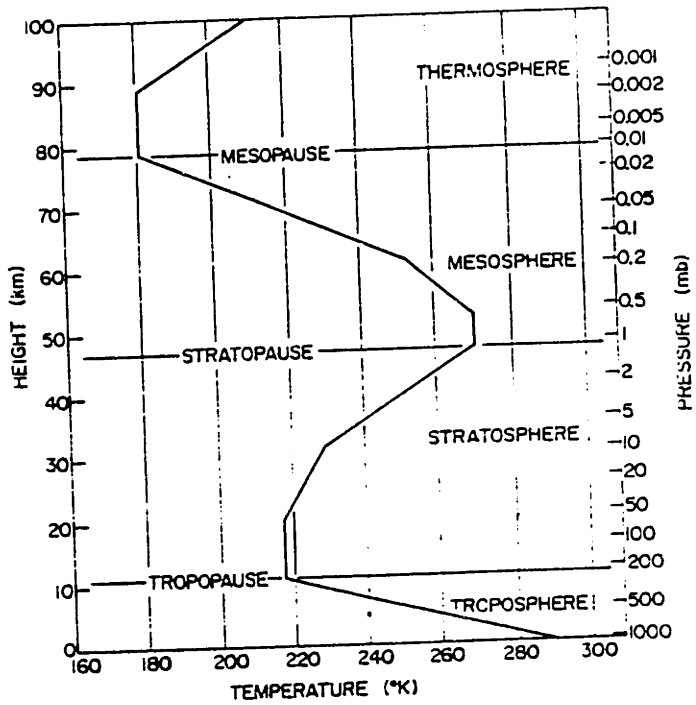


Figure III.3.1: Temperature Distribution for Standard Atmosphere (from Wallace and Hobbs, 1977)

tion at higher altitudes, create several "regions" where the statistics of temperature profiles will be different.

We will restrict ourselves to the lower part of the atmosphere (ground up to 50 *km*) because of the irrelevance of the above layers to the remote sensing problem. This region is referred to as the homosphere and is characterized by an almost uniform composition of the air. It contains all the water of the atmosphere (Qeney, 1974).

Figure III.3.1 (from Wallace and Hobbs, 1977) presents the US Standard Atmosphere in the mentioned region. This atmosphere is a piecewise linear fit

of numerous temperature profile characteristic of the northern mid-latitudes. Although the plot pertains to the mean and the desired analysis to the covariance, the different regions illustrated by Figure III.3.1 will have a direct impact on the covariance.

The lowermost layer of the atmosphere is the troposphere characterized by a fall of temperature with high (lapse rate) at about 6 $^o/km$ (Eskinazi, 1975). The upper limit of the troposphere varies from 7 to 15 km and varies as a function of latitude, time of year and surface type.

The mean temperature is often assumed to follow an adiabatic or constant potential temperature law. The potential temperature θ is defined as a reference temperature at a standard pressure p_o (generally 1000 $mbar$). The temperature obeys Poisson's equation (Wallace and Hobbs, 1977)

$$\theta = T(\frac{p}{p_o})^{\frac{R}{C_p}} \qquad (III.3.1)$$

where R is the perfect gas constant ($2875 K^{-1}kg^{-1}$ for dry air) and C_p the specific heat at constant pressure (1004 $K^{-1}kg^{-1}$). If the pressure distribution obeys the so called hyspometric equation and varies as the inverse exponential of height, the resulting temperature is an affine function of height, as presented on the plot of the US standard atmosphere. In general the temperature at a specific height is found by integrating the lapse rate $\ell(z)$ from the ground up

$$T(z) = T(0) + \int_0^z \ell(h)dh \qquad (III.3.2)$$

The deviation of the temperature from its mean profile (assumed adiabatic in the remaining of this analysis) is controlled by the buoyancy force. When a parcel of air is displaced from equilibrium at height h to an height h', two opposite forces, the gravity and the hydrostatic flotation, act on the parcel and result in the downward buoyancy force equal to

$$f = g(\rho - \rho_{ad}) \qquad (III.3.3)$$

where ρ_{ad} is the adiabatic density and ρ the density of the parcel. Using the law of perfect gas (ρT =constant at a given pressure), one relates the downward buoyancy force to the variation of the adiabatic lapse rate ℓ_{ad} and of the lapse rate of the profile ℓ between h and h'. Namely

$$f = \frac{g}{T_{ad}(h)} \int_h^{h'} \ell(z) - \ell_{ad}(z) dz \qquad (III.3.4)$$

At this point, one must note that the adiabatic lapse rate depends on the water vapor and liquid water content of the atmosphere.

Equations III.3.2 and III.3.4 demonstrate that the lapse rate and its value relative to the adiabatic lapse rate controls both the mean profile and the deviations, through the buoyancy force, of that profile from the adiabatic equilibrium.

Consider then the impact of vertical wind on the lapse rate and the temperature profile. The atmosphere is assumed isentropic, that is

$$p\rho^{-\gamma} = \text{constant} \qquad (III.3.5)$$

where γ is the ratio of air's specific heat at constant pressure over the specific heat at constant volume (γ typically values 1.4). Writing conservation of mass though the vertical motion yields

$$\frac{\partial \rho}{\partial t} + \nabla(\rho W) = 0 \qquad (III.3.6)$$

where W is the vertical wind, in general a function of time and height. Plugging the expression for the buoyancy force in (III.3.6) and using the law of perfect gas yields (after manipulations)

$$\frac{dT}{dz}(h) = T_{ad}(h)\left(\frac{1}{\rho_{ad}}\frac{d\rho_{ad}(h)}{dz} + \frac{g}{\gamma R T(h)}\right) \qquad (III.3.7)$$

Consider now the deviation $\Delta T(h) = T(h) - T_{ad}(h)$ of the temperature profile around its mean. Assuming that the deviation is small compared to the

mean profile so that a Taylor expansion of the differential equation is possible, one can write

$$\frac{d\Delta T}{dz}(h) \approx -\frac{g}{\gamma R T_{ad}(h)}\Delta T(h) \qquad (III.3.8)$$

This is a linear differential equation characterized by a decay constant. Using typical values for T_{ad}, that decay constant is equal to about 10 km (approximately the height of the troposphere) which corresponds to the vertical correlation length if the temperature is now considered to be a stochastic variable.

This simplified analysis shows that the correlation of temperature between points in the troposphere is large. Other physical phenomena (horizontal wind, change of phase,...) will reduce the correlation length. However, the correlation will remain large for a few kilometers.

During summer days, the troposphere can be become unstable and the cooling of the temperature can become superadiabatic and even change sign in the layer 1 km above the ground. These variations in gradient result in large variations in the temperature and thus large variances. The most unstable gradients are close to the ground (Treskoi, 1962). The possible change in sign in the lapse rate yield a correlation lower between the ground and 850 $mbar$ than between 850 and 700 $mbar$.

The top of the troposphere, the tropopause, is a very stable layer which hampers any ascending motion of air. The tropopause has thus a strong impact on the temperature profiles and their second moment. The tropopause is characterized by a abrupt change in the lapse rate, resulting in a region with a null or slightly positive lapse rate. Above the tropopause is the stratosphere. The tropopause is the location where synoptic scale disturbances are most pronounced (Palmen and Newton, 1969). These disturbances' amplitudes decrease as the vertical distance from the tropopause increases. When doing so, they compensate upper tropospheric and lower stratospheric temperature fields. That is, they tend to reduce the differences

between the temperature in these two layers. The deviations of the temperatures from their mean values will thus go in opposite directions. Hence, from a second moment point of view, the impact of the tropopause is a negative correlation when the points considered are across the tropopause inversion.

The height of the tropopause rises in late summer and early fall and lowers at late winter and early spring. Hence the location of the negative anticorrelation will vary with the time of year. The sharp inversion at the tropopause as well as the variation of its height from climate to climate is essential to our understanding of the impact of statistics on statistics on the retrieval process (check Chapter V and VI) and the impact of misspecifying that statistics (Chapter VII).

The stratosphere is characterized by an increase of temperature with height. This positive lapse rate makes the stratosphere mechanically stable independently of its water vapor content or possible changes of phase (Eskinazi, 1975). The structure of the atmospheric temperature in the stratosphere is controlled by upwelling baroclinic eddies (Charney, 1973, Holton, 1972) which results in strong mixing and a strong correlation that extends over a few kilometers in the vertical direction before rapidly decaying toward zero (Tennekes and Tumley, 1972).

Figure III.3.2 presents the correlation of temperature fields as a function of height normalized to different pressure levels distributed in the homosphere between the ground and 10 *mbar*. The temperature is strongly correlated only over a few kilometers. A strong anticorrelation occurs around 200 *mbar* which is the tropopause inversion. The specific set of temperature profiles used in Figure III.3.2 consists of one hundred profiles measured during successive summer at Peoria (Ill.).

Figure III.3.3 shows the variation of the variance of the same set of temperature profiles as a function of height. The larger variances below the tropopause are related to the previously discussed variations of lapse rate.

The form of the correlation varies as the reference pressure changes. These

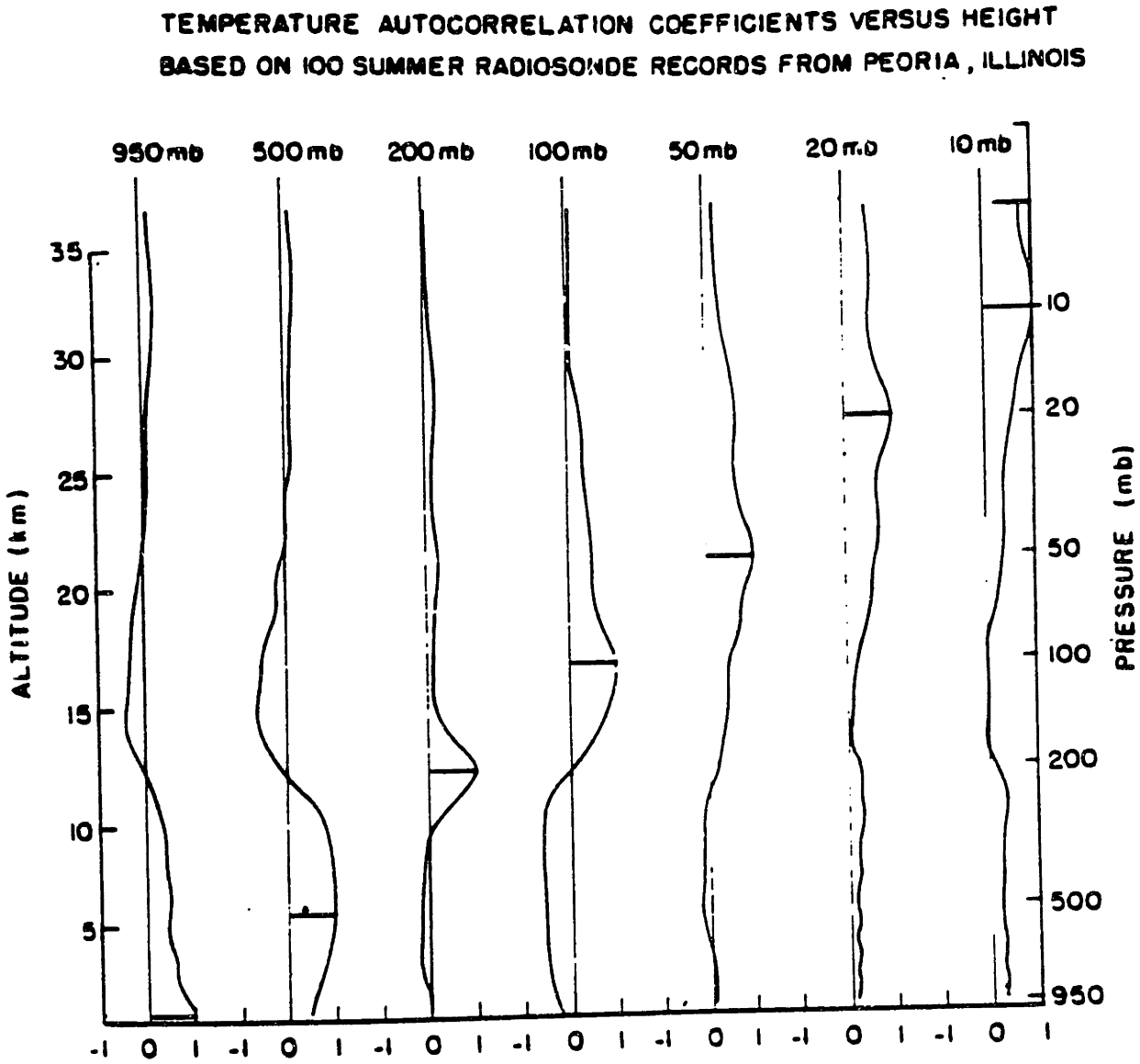


Figure III.3.2: Correlation of atmospheric temperature as a function of height for different reference pressure levels

changes indicate that the atmospheric temperature is not stationary in the vertical direction, as could be expected from the above analysis. The correlation is not a simple function of the difference in height between points. This non-stationarity combined with the heteroscedasticity of the profile makes the derivation of an analytical and tunable model for the temperature nearly impossible.

The best method to develop an adequate representation for the vertical covariance kernel is to use the strong correlation from pressure level to pressure level. This strong correlation allows the discretization of the temperature profiles and of the resulting vertical covariance kernel.

III.3.2 Discretization of Continuous Temperature Profiles

The principal issues in representing the atmosphere as a series of slabs are the number and locations of the sampled levels.

As the number of levels used in the discretization increases so does the number of degrees of freedom associated with the profile. This increase does not pose a problem in terms of defining a valid profile but it can have far reaching consequences on the compilation of statistics, precision of the computation of the radiative transfer and stability in the inversion of the radiative transfer equation. When compiling the covariance kernel (now a matrix) associated with a specific climatology, one must ensure that the number of outer products used is larger than the number of pressure levels used in the representation. This minimum number of profiles is necessary to have a positive definite covariance matrix.

If positive definiteness is not guaranteed, they are linear combinations of components which have no energy (zero variance) and are thus deterministically related. These linear combinations correspond to the eigenvectors of the sample covariance matrix associated with zero eigenvalues. Positive semidefinite matrices are not full rank and correspond to singular probability densities.

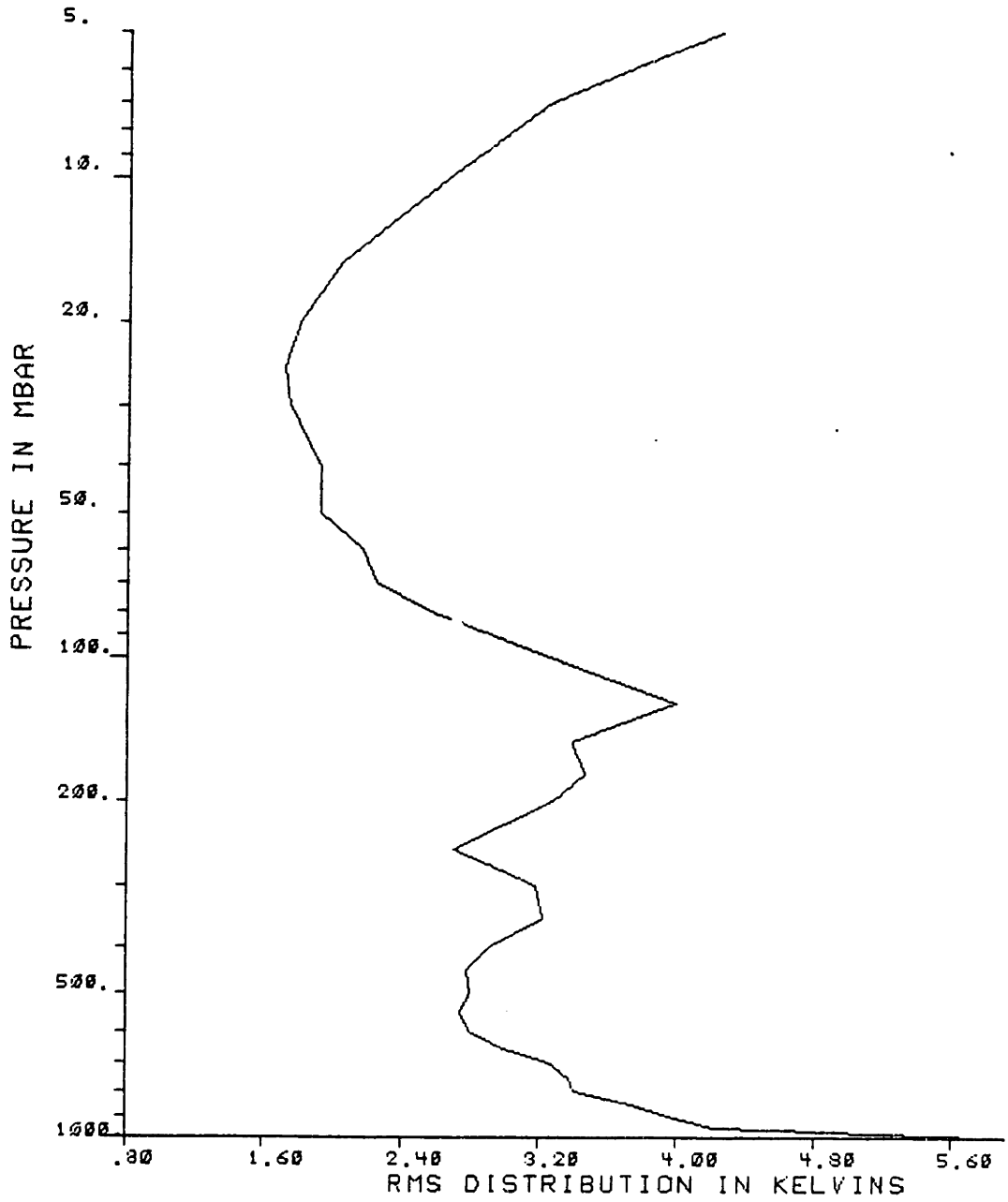


Figure III.3.3: Variance of atmospheric temperature as a function of height for the same set of temperature as Figure III.3.2

Having to use many profiles when compiling the statistics pauses a problem in terms of representing the tropopause and other sharp variations in the structure of temperature fields which vary with time and space.

From an estimation point of view, oversampling the temperature profile is bad because it casts the observation equation into a more and more unspecified framework. The number of columns in the observation matrix increases whereas the number of rows (channels) remains constant. This underspecification results in a more difficult inverse problem (Hunt, 1972). This is true because the estimation must rely more on the statistics and less on the observations and is thus more sensitive to errors in the collection of such statistics -More on this subject in Chapter VII-

Undersampling the profile is bad for obvious reasons: It assumes too few degrees of freedoms which translates into aliasing and errors in the computation of the brightness temperatures. Fine structures close to the ground and the tropopause inversions cannot be resolved. One must have enough levels to allow the implementation of detection and estimation schemes based on differences or ratios of temperatures at different pressure levels. Such schemes detect the type of precipitation in the troposphere (Wagner, 1965) or compute other geophysical parameters such as the wind through the geostrophic conditions (Herring, 1980) approximating derivative by first differences.

To represent the characteristic features of temperature profiles with fine resolution and still avoid oversampling, one must distribute the sampling locations adequately. The discretization grid will have three different patterns in the troposphere, tropopause, and stratosphere.

Last but not least in these considerations is the fact that the final user of temperature estimates is NOAA which wants them at specific pressure levels (1000, 850, 700, 500, 400, 300, 250, 200, 100, 70, and 50 *mbar*). Those levels must be part

of the representation.

As mentioned, the non-stationarity of the temperature as a function of height does not permit the definition of a power spectrum which would determine, if applicable, a Nyquist sampling rate (Papoulis, 1965). One may however estimate such a sampling rate by fitting a stationary analytical covariance to the sample covariance (Figure III.3.2). With a little imagination, one can be convinced that the sample correlations look like "mexican hats" (see for instance the correlation normalized to the 500, 200, and 100 *mbar* pressure levels).

"Mexican hat" shaped covariances are encountered in the propagation of electromagnetic waves in random media (Chernoff, 1960) and turbulent atmospheres (Tatarskii, 1975). They are also observed when studying the variations of the air refracting index (a function of temperature) from its mean values (Colavita, 1985). Moreover, one can obtain a "mexican hat" shaped covariance by applying the geostrophic condition to observed wind statistics (Bergmann, 1979). Wind covariance as a function of height has typically a Lorentzian shape reportedly independent of height. Namely

$$K_{\substack{uu \\ vv}}(\ln p, \ln p') = \frac{K}{1 + K_p^2(\ln p - \ln p')^2} \qquad (III.3.9)$$

where $K_{\substack{uu \\ vv}}$ refers to the vertical covariance of zonal and/or meridional wind. K_p is a decay constant which controls the extent of the vertical correlation and K is the wind's variance. Through the geostrophic conditions, one obtains the vertical covariance of temperature of the form

$$K_{TT}(\ln p, \ln p') = K \frac{1 - 3K_p^2(\ln p - \ln p')^2}{(1 + K_p^2(\ln p - \ln p'))^3} \qquad (III.3.10)$$

This model is stationary and admits a Fourier transform.

Figure III.3.4 presents the variations of the normalized autocorrelation function with normalized pressure difference $K_p(\ln p - \ln p')$. Its minimum value of minus one quarter is obtained when the difference in log pressure is equal to K_p^{-1}.

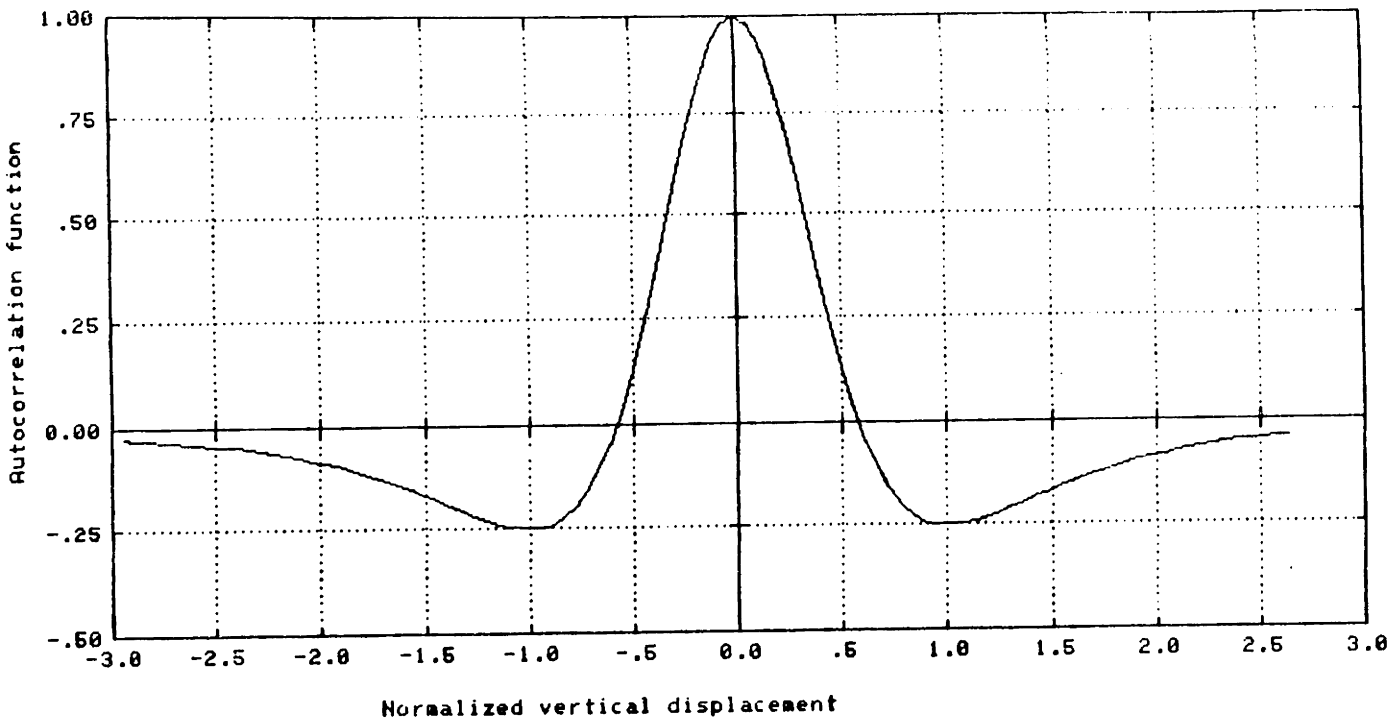


Figure III.3.4: Model for stationary temperature vertical covariance kernel

This function is a better fit to the real correlation in the troposphere and lower stratosphere than in the higher stratosphere. The constant K_p typically equals 2.3 (Bergmann, 1979).

The spectrum associated with such a covariance kernel is

$$S(K_z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} K \frac{1 - 3K_p\tau^2}{(1 + K_p^2\tau)^3} \exp(jK_z\tau) d\tau \qquad (III.3.11)$$

The vertical wavenumber is noted with a capital letter to signify that it is associated with the spectrum of a continuous variable (Oppenheim and Shaffer, 1973). After

tedious manipulations,

$$S(K_z) = \frac{K}{4}\frac{K_z^2}{K_p^3}\exp\left(-\frac{K_z}{K_p}\right) \qquad (III.3.12)$$

Figure III.3.5 plots the variations of the spectrum versus the normalized wavenumber K_z/K_p. Note the absence of DC power. The DC power is carried by the mean profile. There is not DC power in the variation of the profiles from that mean vector because the modeled covariance matrix does not allow for constant shift of the profiles at all pressure levels. This is a welcome result in view of the previous description because of the compensation of the deviations created by baroclinic disturbances at the tropopause. Recall that around the tropopause inversion, the deviation of temperature from the mean profile are in opposite directions at opposite sides of the tropopause. The proposed spectrum and sampling theory (Papoulis, 1965) allow us now to analyze the impact of discretization.

As discussed, the sampling grid should not be uniform either in the pressure or log-pressure (height) representation of profiles. However, the impact of aliasing may still be analyzed by introducing an equivalent sampling distance (Bhat, 1974). Let D denote the distance over which the N samples are taken. The equivalent sampling distance is defined as

$$\delta = \frac{D}{N} \qquad (III.3.13)$$

The distance D is computed as the difference between the natural logarithms of the pressure at the ends of the interval considered expressed in *mbar*. Hence for the layers between the ground and 5 *mbar*, the equivalent sampling distance is $\delta = (\ln 1000 - \ln 5)/N = 5.29/N$

For purely band-limited processes ($S(K_z) = 0$ for $K_z \geq K_{zo}$) being digitized, the sampling locations are irrelevant to the ability to reconstruct the analog signal from its samples as long as the equivalent sampling period satisfies the Nyquist criterion (Yen, 1956). This is because band-limited processes are analytical and can be extrapolated from clustered or distributed samples (Papoulis, 1975,

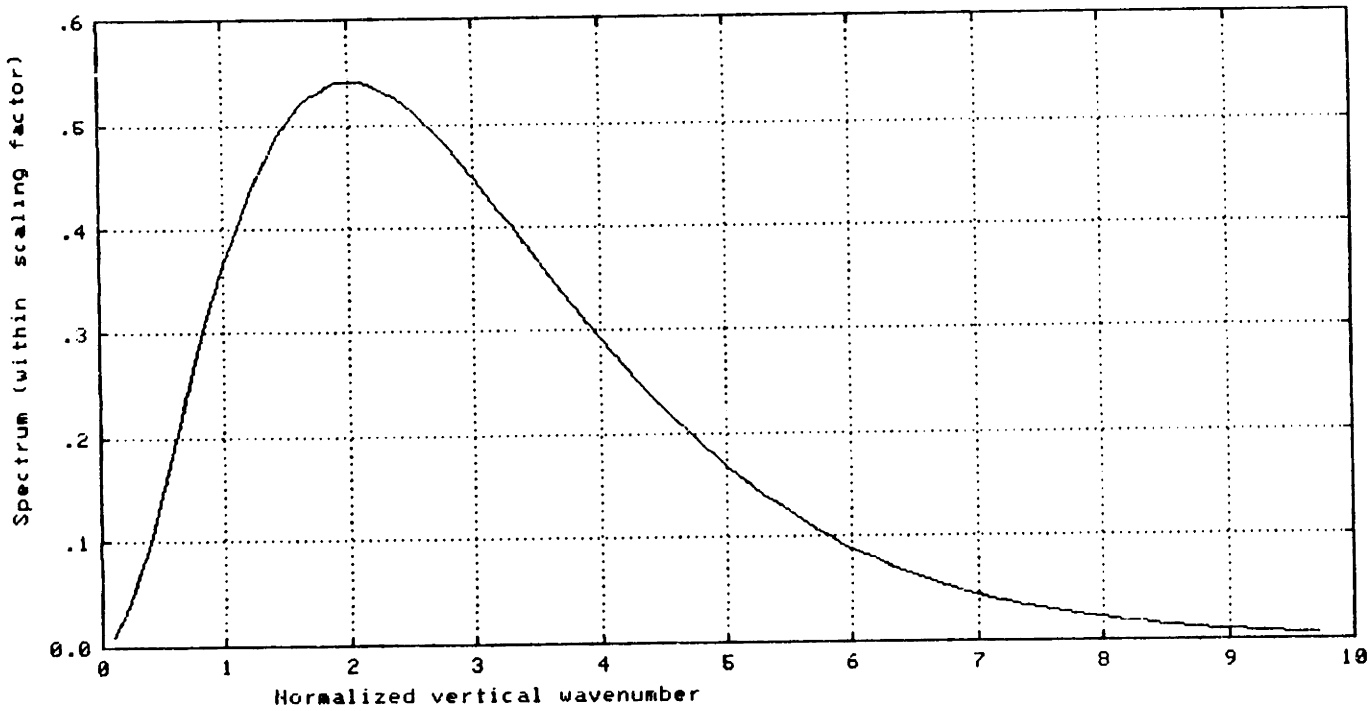


Figure III.3.5: Spectrum associated with the modeled covariance kernel as a function of normalized wavenumber K_z/K_p

Sabri, 1978, Jain and Rangantath, 1981). The main problems associated with the reconstruction from an uneven sampling raster are: (1) the computation of the interpolation function which is no longer a simple cardinal sinus, and (2) the sensitivity of the reconstructed analog signal to errors in the discrete time series.

This independence of the reconstructed signal of the sampling location is no more true for signals which are not truly band-limited but only exhibit a decaying spectra at high wavenumbers. However, the impact of undersampling can be analyzed using this equivalent sampling distance as long as the sampling pattern does not depart dramatically from a regular one. This condition will be respected

in the final representation.

Consider the impact of sampling a stochastic process with the spectrum given by equation III.3.12 with a sampling period δ in the vertical direction. The spectrum of the resulting sequence $T(n)$ is

$$\begin{aligned} S_d(k_z) &= \sum_{n=-\infty}^{\infty} T(n\delta)exp(-jk_z n) \\ &= \frac{1}{\delta}\sum_{r=-\infty}^{\infty} S(\frac{k_z}{\delta} + \frac{2\pi r}{\delta}) \end{aligned} \qquad (III.3.14)$$

where δ is given by III.3.13.

The energy of the unaliased signal is given by

$$E_u = \frac{1}{\delta}\int_{-\pi}^{\pi} S(\frac{k_z}{\delta})dk_z \qquad (III.3.15)$$

whereas the energy in the aliased signal is

$$E_a = \frac{1}{\delta}\int_{-\pi}^{\pi}\sum_{r\neq 0} S(\frac{k_z}{\delta} + \frac{2\pi r}{\delta}) \qquad (III.3.16)$$

Because of the exponential decay, the aliased energy can be approximated by the first two terms of the series if δ is small enough. Hence

$$E_a = \frac{1}{\delta}\int_{-\pi}^{\pi} S(\frac{k_z}{\delta} + \frac{2\pi}{\delta}) + S(\frac{k_z}{\delta} - \frac{2\pi}{\delta}) \qquad (III.3.17)$$

The energy in the original signal is

$$E = \frac{1}{\delta}\int_{-\infty}^{\infty} S(\frac{k_z}{\delta})dk_z \qquad (III.3.18)$$

The criteria used to determine the required sampling distance (or equivalently the number of pressure levels) is the ratio of aliased energy to total energy which must be set less than a small percentage η. That is

$$C = \frac{\int_{-\pi}^{\pi} S(\frac{k_z}{\delta})dk_z}{\int_{-\infty}^{\infty} S(\frac{k_z}{\delta})dk_z} \leq \eta \qquad (III.3.19)$$

The ratio C is found by integrating the spectrum over the appropriate ranges and is equal to

$$C = \exp(\frac{\pi}{\delta K_p})(1 + \frac{\pi}{\delta K_p} + \frac{\pi^2}{2\delta^2 K_p^2}) \qquad (III.3.20)$$

Plugging the equivalent sampling distance associated with the 1000 to 5 *mbar* range and setting K_p to 2.3 yields

$$C = \exp(-0.2582N)(1 + 0.2582 + 0.0333N^2) \qquad (III.3.21)$$

Representing the continuous profile with the 14 mandatory pressure levels of the NMC (Ledsham, 1978, Baumann, 1980, Nathan, 1983) preserves 70 % of the signal, which is not adequate. The 25 levels used by Smith and Woolf (1975) preserves 95 % of the energy, a much larger percentage. Better, 99 % of the energy is unaliased if the atmosphere is discretized into 33 slabs between the ground and 5 *mbar*. These levels will be determined in section III.4.1.

Remember that this calculation was performed in the context of signal representation and not related to the radiative transfer equation which typically reduces bandwidth (Gautier and Revah, 1975) and hence the required number of samples necessary to represent the temperature profile.

III.3.3 Mathematical Constraints on the Vertical Covariance Kernel

Mathematical constraints are less troublesome for the vertical covariance kernel than for the isobaric one because no Fourier relationship is involved. The only constraint is the positive definiteness of the kernel which requires that any weighted line integral of the temperature profile have a positive variance (Van Trees, 1963). Defining

$$A(T) = \int_0^\infty W(h)T(h)dh \qquad (III.3.22)$$

$A(T)$ has a positive variance if for any weighting function $W(h)$

$$\int_0^\infty \int_0^\infty W(h) K_{TT}(h,h') W(h') dh dh' \geq 0 \qquad (III.3.23)$$

After discretization, this condition can be expressed in terms of the eigenvalues of the resulting covariance matrix. They must all be positive.

III.4 Proposed model

III.4.1 Model for the pressure covariance kernel

As discussed, because of the non-stationarity of the vertical covariance kernel, one cannot derive an analytical expression to model it and must thus use finite sums of outer products of temperature profiles with themselves. Nameiy,

$$\begin{aligned} K_{TT} &= < \overline{T}\overline{T}' > \\ &= \frac{1}{N}\sum_{k=1}^{N} \overline{T}_k \overline{T}_k^t - \frac{1}{N}\sum_{k=1}^{N} \overline{T}_k \frac{1}{N}\sum_{k=1}^{N} \overline{T}_k^t \end{aligned} \qquad (III.4.1)$$

The pressure levels of the vectors used in this analysis are the 33 levels used in the water vapor retrieval experiments performed by the Massachusetts Institute of Technology Microwave Remote Sensing Group (Rosenkranz *et al.*, 1982). These levels are 1000, 950, 900, 850, 750, 700, 650, 600, 550, 500, 450, 400, 350, 300, 250, 200, 175, 150, 125, 100, 80, 70, 60, 50, 40, 30, 25, 20, 15, 10, 7, and 5 mbar. Figure III.4.1 presents a scatter plot of those pressure levels as a function of pressure and of log pressure. The sampling raster is regular in the pressure domain in the troposphere and in the log pressure domain in the stratosphere.

III.4.2 Model for the isobaric covariance kernel

The model for the isobaric covariance is a damped cosine function convolved with a Gaussian impulse. The damped cosine provides a second order covariance

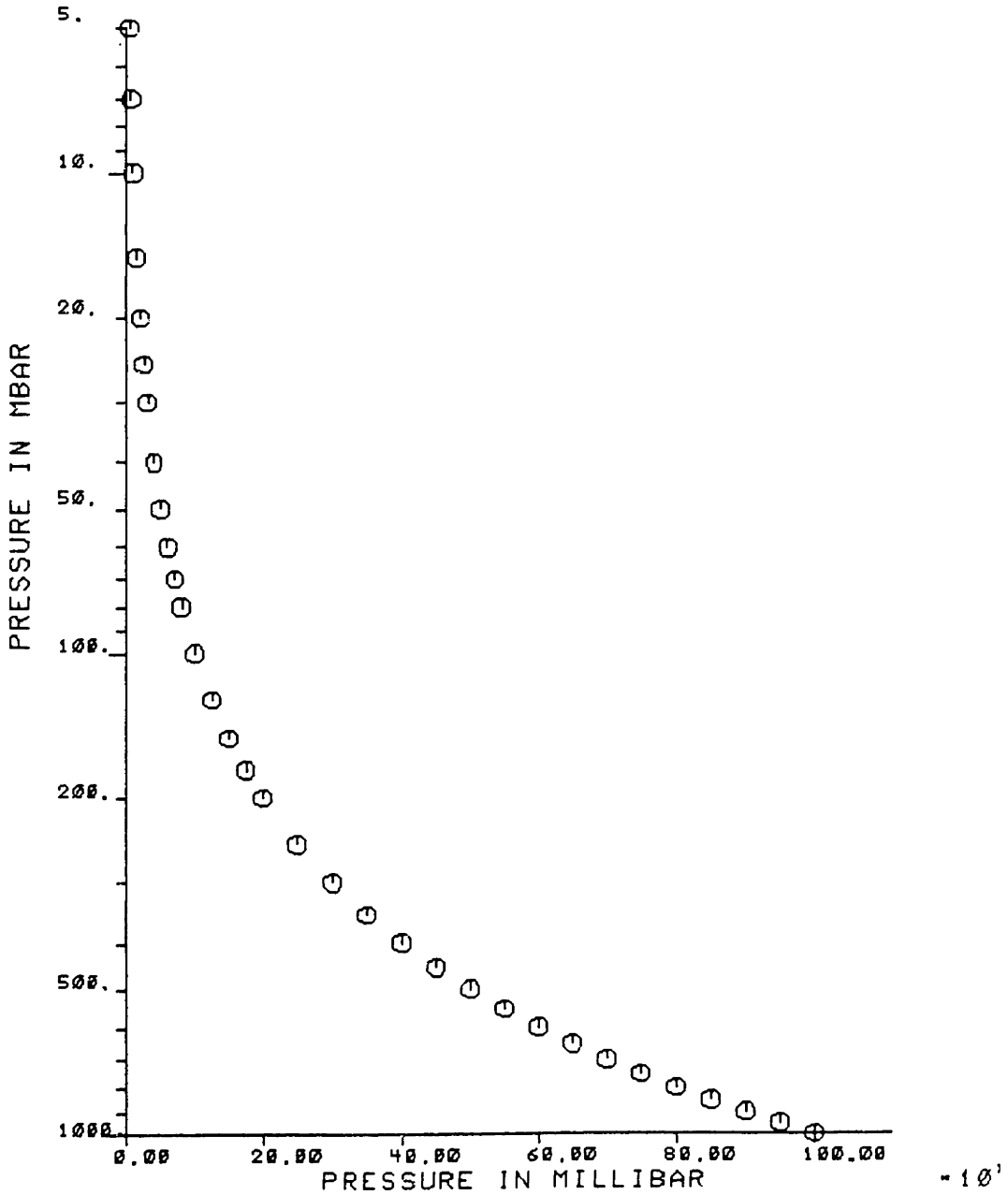


Figure III.4.1: Sampling locations of discretized temperature profiles

and a third order spectrum. The Gaussian impulse does not alter the shape of the covariance but satisfies the geostrophic conditions.

Let $K_{TT}(k;p)$ denote the isobaric covariance as a function of the isobaric

displacement σ at a given pressure level p. The displacement is the Euclidian distance measured along a great circle on the earth, weighted in order to map ellipsoidal isopleths into circular isopleths. Denoting σ_M the displacement along the major axis of the correlation isopleths and σ_m along the minor axis, the displacement used in the computation of the covariance is

$$\sigma = \sqrt{\sigma_m^2 + a^2\sigma_M^2} \qquad (III.4.2)$$

where a is the ratio of minor to major axis which will referred to thereafter as the contraction factor. Because the principal axes of the ellipsoidal isopleths do not have to be zonally and meridionally oriented, one must consider the direction cosine $\cos v$, $\sin v$ of the major axis. Namely,

$$\begin{aligned} \sigma_M^2 &= \sigma_x^2 \cos^2 v + \sigma_y^2 \sin^2 v \\ \sigma_m^2 &= \sigma_x^2 \sin^2 v + \sigma_y^2 \cos^2 v \end{aligned} \qquad (III.4.3)$$

Both a and v are functions of pressure.

Without writing the convolution with the impulse function, the covariance is

$$K_{TT}(\sigma; p) = K_{TT}(0; p) \exp(-\alpha\sigma) \cos(\omega\sigma) \qquad (III.4.4)$$

where $K_{TT}(0; p)$ is the variance of the temperature at the pressure p which will be given by the model for the pressure covariance kernel, α the decay or damping constant in Mm^{-1}, and ω the oscillatory constant in Mm^{-1}. Both these constants are functions of the pressure.

The decay constant controls the rate of decay of the covariance towards zero at the origin whereas the oscillatory constant characterizes the amount (if any) of negative overshoot of the covariance.

Let us introduce the complex frequency $\underline{\xi} = \alpha + j\omega$. The isobaric covariance can be written as a function of the complex frequency

$$\begin{aligned} K_{TT}(\sigma; p) &= K_{TT}(0; p) \Re \exp(-\underline{\xi}\sigma)] \\ &= \frac{K_{TT}(0; p)}{2} [\exp(-\underline{\xi}\sigma) + \exp(-\underline{\xi}^*\sigma)] \end{aligned} \qquad (III.4.5)$$

where $\Re$ denotes the real part.

The corresponding isobaric spectrum is (Gradshteyn and Ryzhik, 1980)

$$S_{TT}(k_h;p) = K_{TT}(0;p)\Re\Big(\frac{\underline{\xi}}{(\underline{\xi}^2 + k_h^2)^{3/2}}\Big) \qquad (III.4.6)$$

The spectrum presents a peak at the wavenumber $k_h = \alpha$ whose height depends on the ratio ω/α. The larger the amount of negative isobaric correlation, the larger the ratio ω/α, the highest the peak at resonance.

The observed variability of empirical isobaric temperature spectra at low wavenumbers can be interpreted by seasonal and geographical variations in the oscillation term ω.

Let us consider a large scale structure such as a weather front passing through an array of observing stations. Stations located in opposite sides of the front will present a negative correlation. When fitting the damped cosine to the observed correlation, this will yield (all things considered) a large oscillatory term, that is a peak in the wavenumber domain. As the large scale structure passes in and out of the observing array, the oscillatory term ω and the peak will vary. This results in variability in the spectrum.

Figure III.4.2 presents the mean (Hankel) spectrum and the envelop of the spectra obtained by varying α and ω. The data used in the fit were measured during January 1979 over western Europe and are the concatenation of 450, 500, and 550 *mbar* temperature fields. The parameters of the model (α, ω) were fitted separately to four weeks in January.

The proposed model for the isobaric spectra presents two limitations as far as specific subranges are concerned. The first one is the overestimation of the spectrum for wavenumber larger than 10 $cycle/Mm$ and underestimation of the spectrum for wavenumber smaller than 10 $cycle/Mm$. This limitation comes from the fact that two subranges of spectra are represented by the same analytical

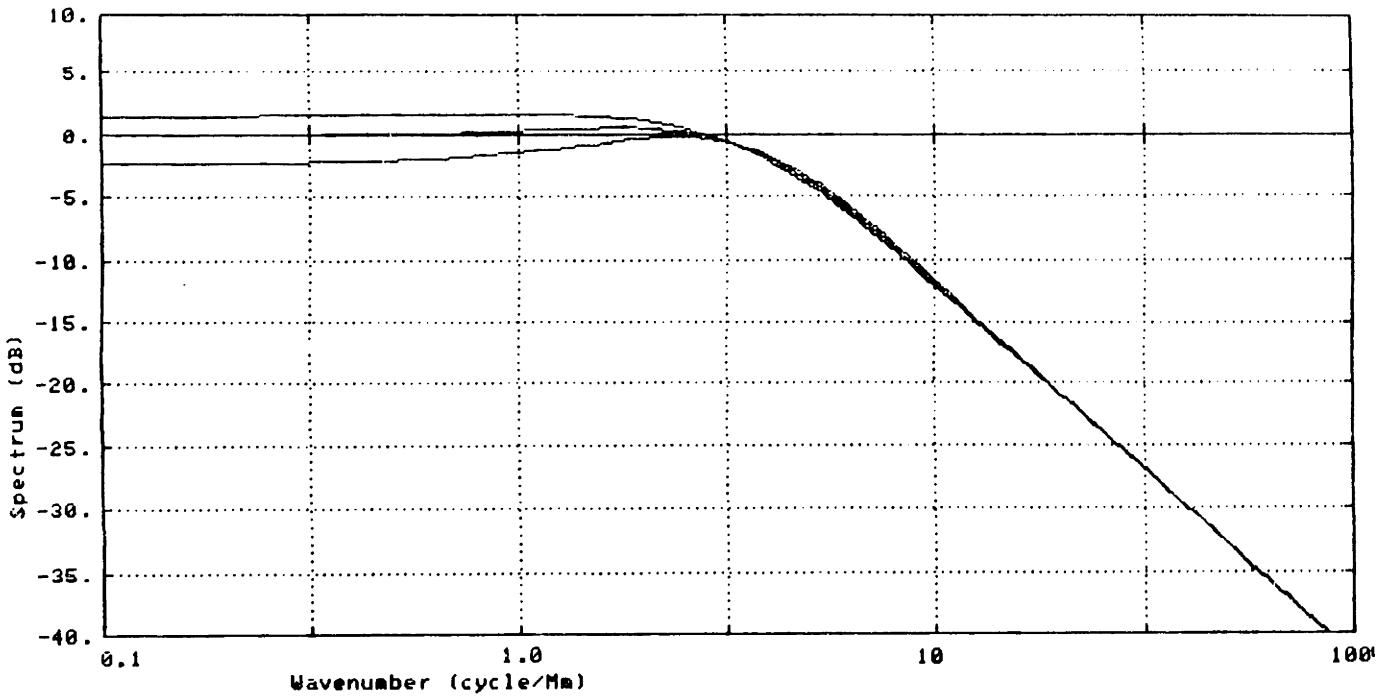


Figure III.4.2: Impact of the change in ω on low wavenumber component of the modeled spectrum

expression which is a compromise between both regimes. The second limitation is the low wavenumber end where the spectrum tends to a non zero DC value except when α and ω are comparable. Because low wavenumber spectra are ill-defined, there is not proof that the modeled representation is correct (or incorrect for that matter).

Figure III.4.3 presents the integrated magnitude spectrum based on the model for isobaric covariance fitted to February data collected over the United States. The 10 pressure levels used in the integration are the 10 sigma levels of the University of Wisconsin limited area forecasting model (Gauntlett *et al.*, 1978).

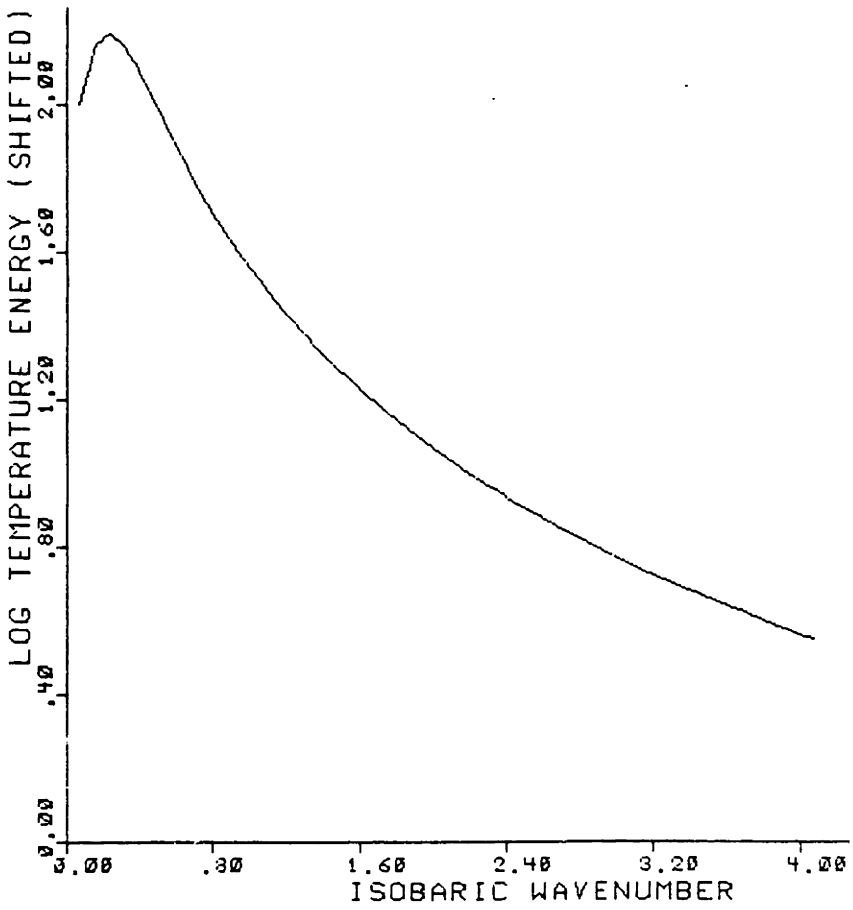


Figure III.4.3: Integrated temperature magnitude spectrum based on the model fitted to February data over the United States

The spectrum is plotted in a semi logarithmic fashion.

Figure III.4.4 illustrates, for the purpose of comparison, the difference in temperature energy between different baselines computed using the Wisconsin grid applied to the Continental United States with a 132 km resolution. (Rosenkranz and Staelin, 1985).

The agreement in shape between model and experimental results is excellent in the wavenumber scale of interest which are characteristic of remote-sensing and

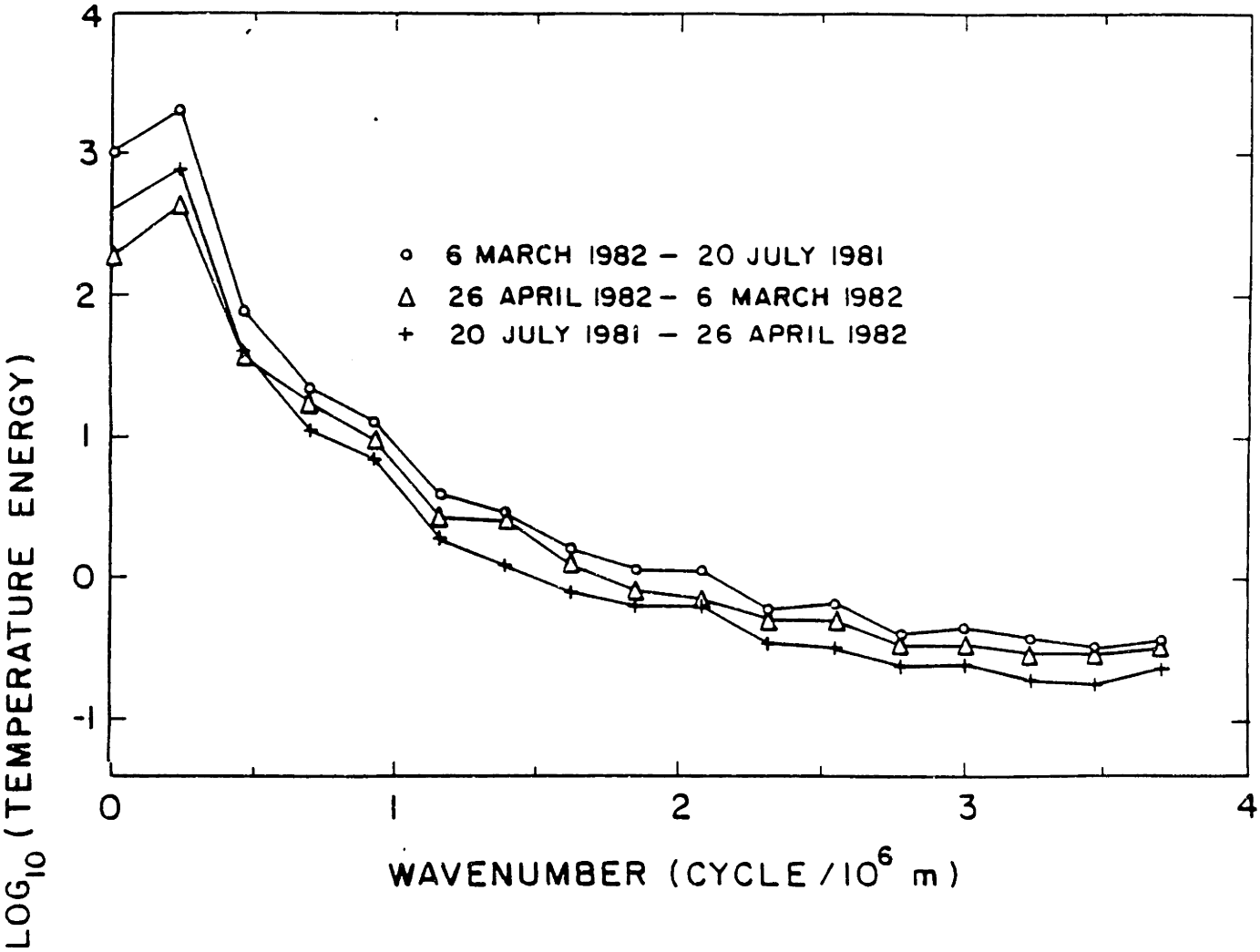


Figure III.4.4: Integrated temperature magnitude spectrum based on limited area numerical weather prediction (from Rosenkranz and Staelin, 1985)

data assimilation.

The model for the isobaric covariance must now satisfy the geostrophic constraint and the positive definiteness. One must now consider in the isobaric covariance modeling is the size of the Gaussian impulse required to satisfy the geostrophic conditions. Its size is found by solving the relationship between wind

variance and temperature isobaric covariance kernel.

Recalling (III.2.46)

$$\frac{\partial^2}{\partial \ln^2 p} K^i_{uu}(x_i - x_j, y_i - y_j; p) = \frac{G_i G_j R^2}{f_i f_j} \frac{\partial^2}{\partial x_i \partial x_j} K^i_{TT}(x_i - x_j, y_i - y_j; p) \quad (III.4.7)$$

The limit of the derivative as the displacement goes to zero is

$$\begin{aligned} \lambda &= \lim_{\sigma \to 0} \frac{1}{\sigma} \frac{\partial K^i_{TT}(\sigma; p)}{\partial \sigma} \\ &= \frac{1}{2} \int_0^\infty S^i_{TT}(k; p) k^3 dk \end{aligned} \quad (III.4.8)$$

Taking the Hankel spectrum as the 3^{rd} order spectrum derived from the damped cosine covariance multiplied by a Gaussian function yields

$$\frac{\partial^2}{\partial \ln^2 p} K^i_{uu}(x_i - x_j, y_i - y_j; p) = \frac{G_i G_j R^2}{f_i f_j} \frac{1}{2} \int_0^\infty \Re\left(\frac{\underline{\xi}}{(\underline{\xi}^2 + k^2)^{3/2}}\right) \exp\left(-\frac{k^2}{k_m^2}\right) k^3 dk \quad (III.4.9)$$

This integral cannot be computed exactly in an analytical fashion and must therefore be approximated. Recalling that k_m the cutoff wavenumber of the exponential roll-off is much larger than α the real part of the complex pole $\underline{\xi}$ (see Equation III.2.28), one can compute the integral on the right hand side of III.4.9 by separating the integral into two parts.

Let k_ν be an intermediate wavenumber between α and k_m. For k less than k_ν $\exp(-k^2/{k_m}^2) \approx 1$ and for k greater than k_ν $\Re[\underline{\xi}/(\underline{\xi}^2 + k^2)^{3/2}] \approx \Re[\underline{\xi}] k^{-3}$. Thus the integral becomes

$$\int_o^{k_\nu} \Re\left(\frac{\underline{\xi}}{(k_2 + \underline{\xi}^2)^{3/2}}\right) k^3 dk + \int_{k_\nu}^\infty \Re(\underline{\xi}) \exp\left(-\frac{k^2}{k_m^2}\right) dk \quad (III.4.10)$$

The latter integral can be easily computed.

The finiteness of the moments of the isobaric covariance kernel are guaranteed by the presence of the decaying exponential. The first moment is given by

$$\int_0^\infty \exp(-\alpha\sigma) \cos(\omega\sigma) d\sigma = \frac{\alpha^2 - \omega^2}{(\alpha^2 - \omega^2)^2 + 4\omega^2\alpha^2} \quad (III.4.11)$$

It is positive for any complex pole $\underline{\xi}$ satisfying α greater than ω.

Testing the positive definiteness of the covariance kernel is performed by analyzing the positivity of the spectrum. Equation (III.4.11) gives the spectrum corresponding to the model. After expanding the complex pole $\underline{\xi}$ the spectrum can be written as

$$S_{TT}(k_h;p) = K^o_{TT}(0;p)\Re\Big(\frac{\alpha+j\omega}{\sqrt{\alpha^2-\omega^2+k_h^2+2j\alpha\omega}^3}\Big) \qquad (III.4.12)$$

where the square root of the denominator must satisfy the inequality (Watson, 1944)

$$|\alpha+\sqrt{\alpha^2-\omega^2+k_h^2+2j\alpha\omega}| > |k| \qquad (III.4.13)$$

Let us express the spectrum and the condition pertaining to the square root using phasers. Let write the complex pole $\underline{\xi}$ as

$$\underline{\xi} = \alpha + j\omega = Re^{j\phi}, \qquad \phi = \tan^{-1}\frac{\omega}{\alpha} \qquad (III.4.14)$$

The denominator of (III.2.12) can be written as

$$D = \alpha^2-\omega^2+k_h^2+2j\alpha\omega = re^{j\theta} \qquad (III.4.15)$$

where

$$\theta = \tan^{-1}\frac{2\alpha\omega}{\alpha^2-\omega^2+k_h^2} = \tan^{-1}\frac{\frac{2\omega}{\alpha}}{1-\frac{\omega^2}{\alpha^2}+\frac{k_h^2}{\alpha^2}} \qquad (III.4.16)$$

The spectrum is then

$$S = \pm Rr^{-\frac{3}{2}}\cos(\phi-3\frac{\theta}{2}) \qquad (III.4.17)$$

where the sign is determined by the inequality

$$|Re^{j\phi}\pm r^{1/2}e^{j\frac{\theta}{2}}| > |k_h| \qquad (III.4.18)$$

or

$$|1\pm r^{1/2}R^{-1}e^{j\frac{\theta}{2}}| > \frac{|k_h|}{|R|} \qquad (III.4.19)$$

At the origin ($k_h = 0$), the phase of the denominator θ equals 2ϕ and tends to zero in a monotonous fashion as the horizontal wavenumber k_h tends to infinity. Thus

$$2\phi < \phi - \frac{3\theta}{2} < \phi \qquad (III.4.20)$$

and

$$-\phi < \frac{\theta}{2} < 0 \qquad (III.4.21)$$

The spectrum will not change sign and therefore become negative because of the cosine if the angles 2ϕ and ϕ belong to the same quadrant, that is if

$$|\phi| < \frac{\pi}{4} \qquad (III.4.22)$$

The inequality (III.4.22) will be satisfied without change of sign if $\frac{\theta}{2} - \phi$ remains in the first or the fourth quadrant. That is, if

$$|\phi| < \frac{\pi}{2} \qquad (III.4.23)$$

Combining these two conditions for positivity, the spectrum will be positive if the angle ϕ is less than $\frac{\pi}{2}$ in absolute value. That is $-\alpha < \omega < \alpha$. This condition defines the set of valid values for the complex pole $\underline{\xi}$ as a quadrant whose vertex is the origin and symmetric with respect to the real axis.

The normalized correlation $\rho(\sigma)$ can be written as a function of the dimensionless displacement σ by introducing the ratio $\mu = \frac{\alpha}{\omega}$.

$$\rho(\sigma) = \exp(-\mu\sigma)\cos(\sigma) \qquad (III.4.24)$$

whose minimum value is

$$\rho_{min} = -\frac{\exp(-\mu \tan^{-1}\mu)}{\sqrt{1+\mu^2}} exp(-\mu\pi) \qquad (III.4.25)$$

Setting μ equals to one yields a minimum value for the normalized correlation ρ_{min} of -0.067. This value is much larger than the minimum possible value of -0.403 because the spectrum is quite different from an impulse.

III.4.3 Model for the Three-Dimensional Covariance Kernel

The three-dimensional model is based on the observation that isobaric temperature fields are stationary up to 3 Mm. The isobaric temperature can thus be modeled as a white noise filtered by a space invariant filter (Thiebaux, 1976, VanMarcke, 1983). A second important observation is that each pressure level can be considered as a specific component of a random vector expressed in a pressure representation.

One can therefore attempt to model three dimensional temperature fields as filtered stochastic vector fields. The dimension of these vectors may be infinite in the pressure direction if one does not wish to introduce any vertical discretization.

Let $w(x,y,p)$ denote a zero mean stochastic vector field with l (l finite) components in the pressure direction where x and y are cartesian coordinates on a polar stereographic grid and p the pressure. This vector is white in both the isobaric and the pressure direction. In other words, denoting the Dirac impulse by δ

$$K_{ww}(x-x',y-y',p,p') = \overline{w(x,y,p)w(x',y',p')} = \delta(x-x')\delta(y-y')\delta(p-p') \qquad (III.4.26)$$

where the overbar denotes the probability average.

Let K^p_{TT} denote the desired pressure covariance matrix for the temperature field. Since this matrix is positive semi-definite, its square root matrix R_{TT} can be computed using a Cholesky type decomposition (Gelb, 1974, Musicus, 1981). The square root matrix is defined as a positive semi-definite matrix which satisfies the equation

$$R_{TT}R^t_{TT} = K_{TT} \qquad (III.4.27)$$

where t denotes the matrix transposition. In general, this decomposition is not unique since one can apply any unitary matrix to the right of the square root matrix

and still satisfy (III.4.27). The decomposition is made unique by constraining the square root matrix to a canonical form, such as upper or lower triangular.

In all cases, the exact square root of the pressure covariance matrix is irrelevant for our purpose. Let the stochastic vector $w(x,y,p)$ be projected onto $\mathbf{R}^l$ through the matrix R_{TT}. The resulting vector $\psi(x,y,p)$ satisfies the following covariance kernel of

$$K_{\psi\psi}(x-x',y-y',p,p') = \overline{\psi(x,y,p)\psi(x',y',p')} = \delta(x-x')\delta(y-y')K^p_{TT} \quad (III.4.28)$$

Note that the pressure covariance kernel does not have to be stationary in the pressure direction.

Let us filter the resulting semi-white noise through a bank of planar two-dimensional Linear Space Invariant filters whose point spread function $h(x,y;p)$ are solely a function of pressure. These filters are assumed to have the same functional form and are dependent on a vector of parameters $\underline{\xi}(p)$ whose value varies with the pressure. In order to force a two-dimensional aspect to the isobaric modeling, these filters are zero height or at most of height smaller than any scale of variation of the parameter vector $\underline{\xi}(p)$.

The filtered temperature $T(x,y,p)$ has for a covariance kernel

$$K_{TT}(x-x',y-y',p,p') = \overline{T(x,y,p)T(x',y',p')}$$

$$= K^p_{TT}\int\int h(\tau,\mu;p)h(x-x'-\tau,y-y'-\mu;p')d\tau d\mu \qquad (III.4.29)$$

For the resulting field to be isotropic in the isobaric direction, the point spread function $h(\tau,\mu;p)$ must itself be isotropic after proper conformal mapping. Denoting by (a,v) the contraction mapping and direction cosine, the point spread function satisfies

$$h(\tau,\mu;p) = h(\sqrt{\tau^2(a^2\cos^2 v+\sin^2 v)+\mu^2(\cos^2 v+a^2\sin^2 v)};p) \qquad (III.4.30)$$

The corresponding spectrum is

$$S_{TT}(k,p,p') = K^p_{TT} H(k;p) H^*(k,p') \qquad (III.4.31)$$

In general, the Fourier transform of a cross-correlation between two different variables (such as the temperature at two different pressure level) is hermitian because of the real nature of the cross correlation. That is its real part is an even function of the wavenumber and its imaginary part is an odd function. In the case of isotropic (and ellipsoidal anisotropic) symmetry, the Fourier transform has to be symmetric with respect to the frequency (or wavenumber). Thus, the Fourier transform must be real and the Hankel transforms $H(k;p) = H(k;\underline{\xi}(p))$ are real for any value of k, p, and p'.

Under the isotropy assumption, the covariance kernel can be expressed as the Hankel transform of the resulting spectrum. Namely,

$$K_{TT}(\sigma,p,p') = K^p_{TT} \int_0^\infty H(k;p) H^*(k;p') J_0(k\sigma) k dk \qquad (III.4.32)$$

When the two pressure levels are identical, the resulting spectrum is the one defined by (III.4.6). The parameter vector is the complex pole of the filter. Thus, without writing the Gaussian function,

$$H(k;p)^2 = \Re\Big(\frac{\underline{\xi}(p)}{(\underline{\xi}(p)^2 + k^2)^{\frac{3}{2}}}\Big) \qquad (III.4.33)$$

or equivalently

$$H(k;p) = \Big(\Re\big(\frac{\underline{\xi}(p)}{(\underline{\xi}(p)^2 + k^2)^{\frac{3}{2}}}\big)\Big)^{\frac{1}{2}} \qquad (III.4.34)$$

These isobaric filtering filters can be regarded as having a complex pole at the location $\underline{\xi}(p)$ and of order 3/4. This order makes their point spread function irrational. Thus, one can not model the temperature field as been an Auto Regressive Moving Average (ARMA) type process. This shows that one cannot describe isobaric temperature fields as Markov processes of finite order as suggested by Thiebaux (1975)

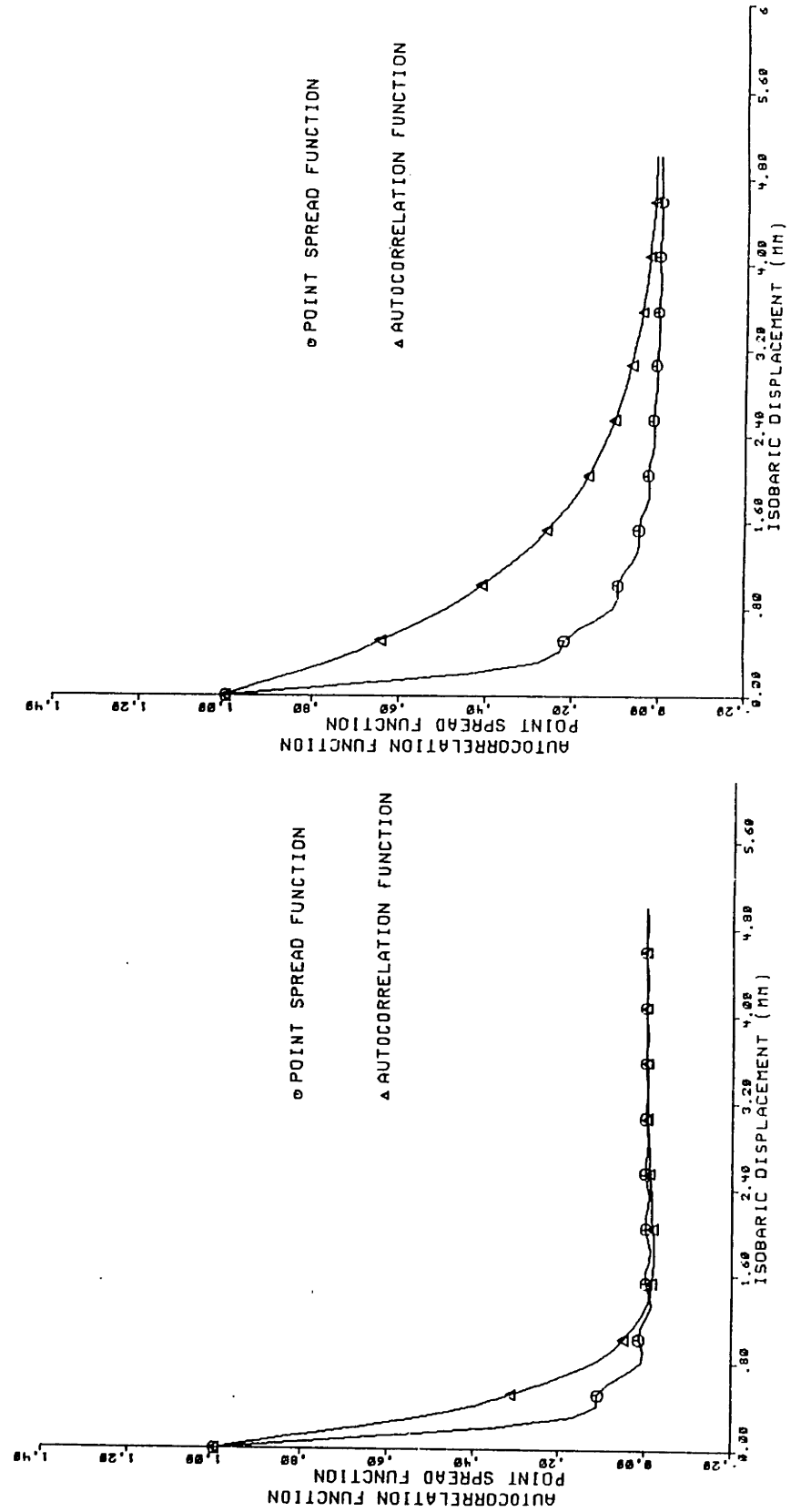


Figure III.4.5: Point spread function of two isobaric filters and resulting covariance function.

An alternate representation for the covariance kernel can be found for (III.4.35) by manipulating it as an integral equation rather than as a functional Hankel transformation. Factorizing k_h the wavenumber response of the isobaric filters yields for the three-dimensional covariance kernel

$$K_{TT}(\sigma,p,p') = K^p_{TT}\int_0^\infty \sqrt{\Re(\frac{k_h^2/\underline{\xi}(p)^2}{(1+k_h^2/\underline{\xi}(p)^2)^{\frac{3}{2}}})}\sqrt{\Re(\frac{k_h^2/\underline{\xi}(p')^2}{(1+k_h^2/\underline{\xi}(p')^2)^{\frac{3}{2}}})}J_0(k_h\sigma)\frac{dk_h}{k_h} \qquad (III.4.36)$$

The computation of the above integral has to be performed numerically, since no analytical expression can be found for it. Appendix III.A presents the backsmear algorithm that was used to perform this operation and analyze its performances.

It should be noted at this point that if future uses of the model prove the modeling of the isobaric covariance as a second order (in the spatial domain) covariance as inappropriate, one derive a more elaborate model by taking better representations of the isobaric spectrum and using the convolutional combination described above.

The filtering operation described by equation III.4.29 can also be described by a differential equation in the spatial domain. Using Taylor's assumption, this differential equation can include derivatives with respect to time and to displacement. The forcing term of this equation is the field $\psi(x,y,p)$ which has a white spectrum. To a certain degree, this modeling parallels the scale dependent stochastic climate model introduced by North and Cahalan (1981). In their modeling, the energy balance climate model (North, 1975) is driven by a white stochastic field whose physical origin lies in eddy transport fluctuations, storms, and other phenomena of short time scale.

The positive definiteness of the model can be proven using the composition of Schur (1911) on the closure of positive definite positive kernels under point wise multiplication. This theorem states that the product element per element of two

positive definite covariance kernels is also a positive covariance kernel.

Consider then a series of N temperature profile T_n at different locations $\vec{r}_n (n = 1, \ldots, N)$. One can catenate these vectors in a large temperature vector $\tilde{T}^t = [T_1^t, T_2^t, \ldots, T_N^t]$. The proposed model will be positive definite if the covariance matrix associated with $\tilde{T}$ is positive definite for any set of locations. Denoting by σ_{nm} between the n^{th} and m^{th} locations, the covariance matrix corresponding to all pressure levels can be written as

$$\begin{pmatrix} K_{TT}(\sigma_{11}) & \ldots & K_{TT}(\sigma_{1N}) \\ \vdots & & \vdots \\ K_{TT}(\sigma_{N1}) & \ldots & K_{TT}(\sigma_{NN}) \end{pmatrix} \qquad (III.4.37)$$

where each individual submatrix K_{TT} is given by the model. This matrix can be written as the product of three matrices

$$K_{TT}^p \bigotimes J \bigcirc \begin{pmatrix} K_{TT}^i(\sigma_{11}) & \ldots & K_{TT}^i(\sigma_{1N}) \\ \vdots & & \vdots \\ K_{TT}^i(\sigma_{N1}) & \ldots & K_{TT}^i(\sigma_{NN}) \end{pmatrix} \qquad (III.4.38)$$

where $\bigotimes$ denotes the kronecker product of matrices (Bellman, 1960) and $\bigcirc$ the point wise multiplication of matrices. The first term of the product is the vertical covariance kernel. The second term is a $N \times N$ amtrix whose entries are all equal to one. The last matrix is a collectien of damped cosine like functions resulting from equation III.4.34. The second matrix is positive semidefinite (all eigenvalues equal to zero, except for one equal to N). Since the vertical covariance kernel is positive definite, the first kronecker product is positive definite (Brewer, 1978).

The third matrix is positive definite if its spectrum associated with the covariance kernel is positive. That is,

$$\begin{aligned} S_{TT}^i(k_h; p, p') &= \int_0^\infty K_{TT}^i(\sigma; p, p') J_0(k_h \sigma) \sigma d\sigma \\ &= \left(\Re\left(\frac{k_h^2 / \underline{\xi}(p)^2}{(1 + k_h^2 / \underline{\xi}(p)^2)^{\frac{3}{2}}} \right) \right)^{\frac{1}{2}} \left(\Re\left(\frac{k_h^2 / \underline{\xi}(p')^2}{(1 + k_h^2 / \underline{\xi}(p')^2)^{\frac{3}{2}}} \right) \right)^{\frac{1}{2}} \end{aligned} \qquad (III.4.39)$$

In section III.4.2, it was shown that each ones of the frequency responses is positive provided that both $\underline{\xi}(p)$ and $\underline{\xi}(p')$ have an imaginary part less in magnitude that the real part. If these conditions hold, then the spectrum $S^{i}_{TT}(k_h; p, p')$ is positive and by virtue of the Schur's theorem, the three-dimensional kernel is positive definite.

III.4.4 A Computationally Efficient Approximation to the Model

An approximation to the proposed three dimensional model is required for applications where the computation of the correlation between a large ensemble of points in the atmosphere is necessary. This is the case in the design of retrieval operators which operate on a large of points both in the spatial and in the sensing frequency domain. The computation of the correlation between two brightness temperatures require the evaluation of the correlation between all the pressure levels used in the numerical computation of the radiative transfer equation.

Such time saving approximations are not necessary, for example, in objective analysis where fewer points (typically a dozen) are taken as inputs to the assimilation process.

Let us introduce the "generic" isobaric filter wavenumber response $H_c(x)$ defined as

$$H_c(x) = \Re\left(\frac{x^2}{(1+x^2)^{\frac{3}{2}}}\right) \qquad (III.4.40)$$

The three-dimensional covariance can now be expressed using the generic filter response. Let α and β denote the complex poles for the pressure p and p' respectively. $K_{TT}(\sigma, p, p')$ becomes

$$K_{TT}(\sigma, p, p') = K^{p}_{TT} \int_0^{\infty} H_c(k/\alpha) H_c(k/\beta) J_0(k\sigma) \frac{dk}{k} \qquad (III.4.41)$$

The symmetry in the roles of α and β is apparent in the above expression and should be preserved by any approximation.

Let γ denote the ratio of the two poles $\frac{\alpha}{\beta}$. Let us make the change of variable $\kappa = k\sqrt{\alpha\beta}$ (III.4.41) becomes

$$K_{TT}(\sigma, p, p') = K_{TT}^{p} \oint_0^{\infty} H_c(\kappa/\gamma) H_c(\kappa\gamma) J_0(\sigma\sqrt{\alpha\beta}\kappa) \frac{d\kappa}{\kappa} \qquad (III.4.42)$$

The integral is a line integral with a integration path being a straight line in the complex plane which forms a angle with the real axis of $\tan^{-1}\sqrt{\alpha\beta}$. The above equation can be analyzed as defining the function $K(\gamma, \sqrt{\alpha\beta}\sigma)$. Because of the symmetry in the roles of α and β

$$K(\gamma, \sqrt{\alpha\beta}\sigma) = K(1/\gamma, \sqrt{\alpha\beta}\sigma) \qquad (III.4.43)$$

The desired approximation can now be found by considering a Taylor expansion of the definiteness of K around γ equal to one.

$$K(\gamma, \sqrt{\alpha\beta}\sigma) = K(1, \sqrt{\alpha\beta}\sigma) + \frac{\partial}{\partial\gamma} K(1, \sqrt{\alpha\beta}\sigma)(\gamma - 1) + \frac{\partial^2}{\partial\gamma^2} K(1, \sqrt{\alpha\beta}\sigma)(\gamma - 1)^2 + O((\gamma - 1)^3) \qquad (III.4.44)$$

Let us consider the behavior of the Taylor expansion as the inverse of γ is taken. As $\gamma \to 1/\gamma$, $\gamma - 1 \to 1 - \gamma$. The symmetry of K with respect to γ implies that the first order term (and all odd number terms) of the Taylor expansion equals zero. Therefore,

$$K(\gamma, \sqrt{\alpha\beta}\sigma) = K(1, \sqrt{\alpha\beta}\sigma) + \frac{\partial^2}{\partial\gamma^2} K(1, \sqrt{\alpha\beta}\sigma)(\gamma - 1)^2 + O((\gamma - 1)^4) \qquad (III.4.45)$$

The function $K(\gamma, \sqrt{\alpha\beta}\sigma)$ can be therefore approximated by its value at the origin and the covariance kernel becomes

$$K_{TT}(\sigma, p, p') = K_{TT}^{p} \Re[\exp(-\sqrt{\underline{\xi}(p)\underline{\xi}(p')}\sigma)] \qquad (III.4.46)$$

Figure III.4.6 presents a graphical analysis of the approximation based on the concept of the Bode plot. Rather than considering the wavenumber response as having

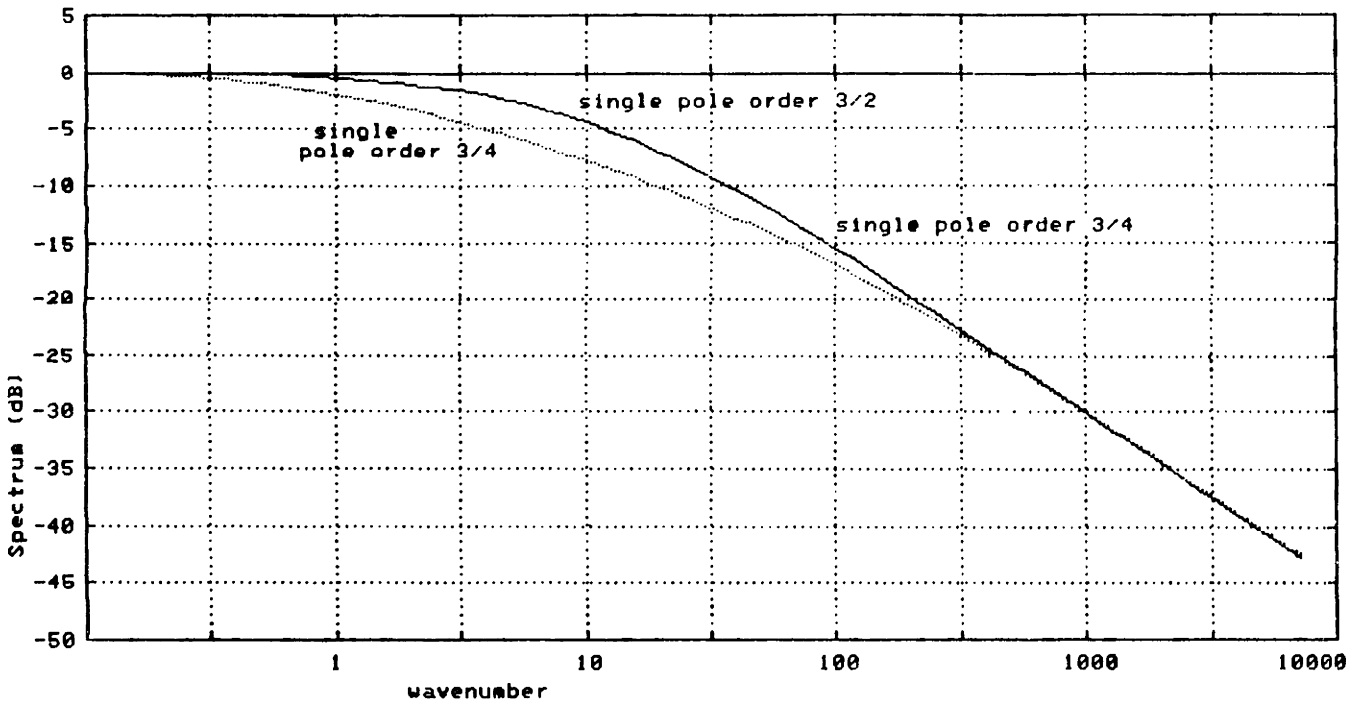


Figure III.4.6: Bode plot representation of the horizontal cross correlation term of the three dimensional model and proposed approximation.

two poles of order 3/4 at $\underline{\xi}(p)$ and $\underline{\xi}(p')$, the approximation takes the geometric average of these two poles, the mid point on a Bode plot, and then assigns a single pole of twice the order, 3/2.

This approximation will be the most accurate when the two poles are near one another. Likewise, as the displacement σ increases, the effective distance between the two poles (for a fixed wavenumber k) increases, thus worsening the approximation. The approximation has the effect of dimishing the horizontal correlation length by not allowing a more constructive integration of the wavenumber spectrum

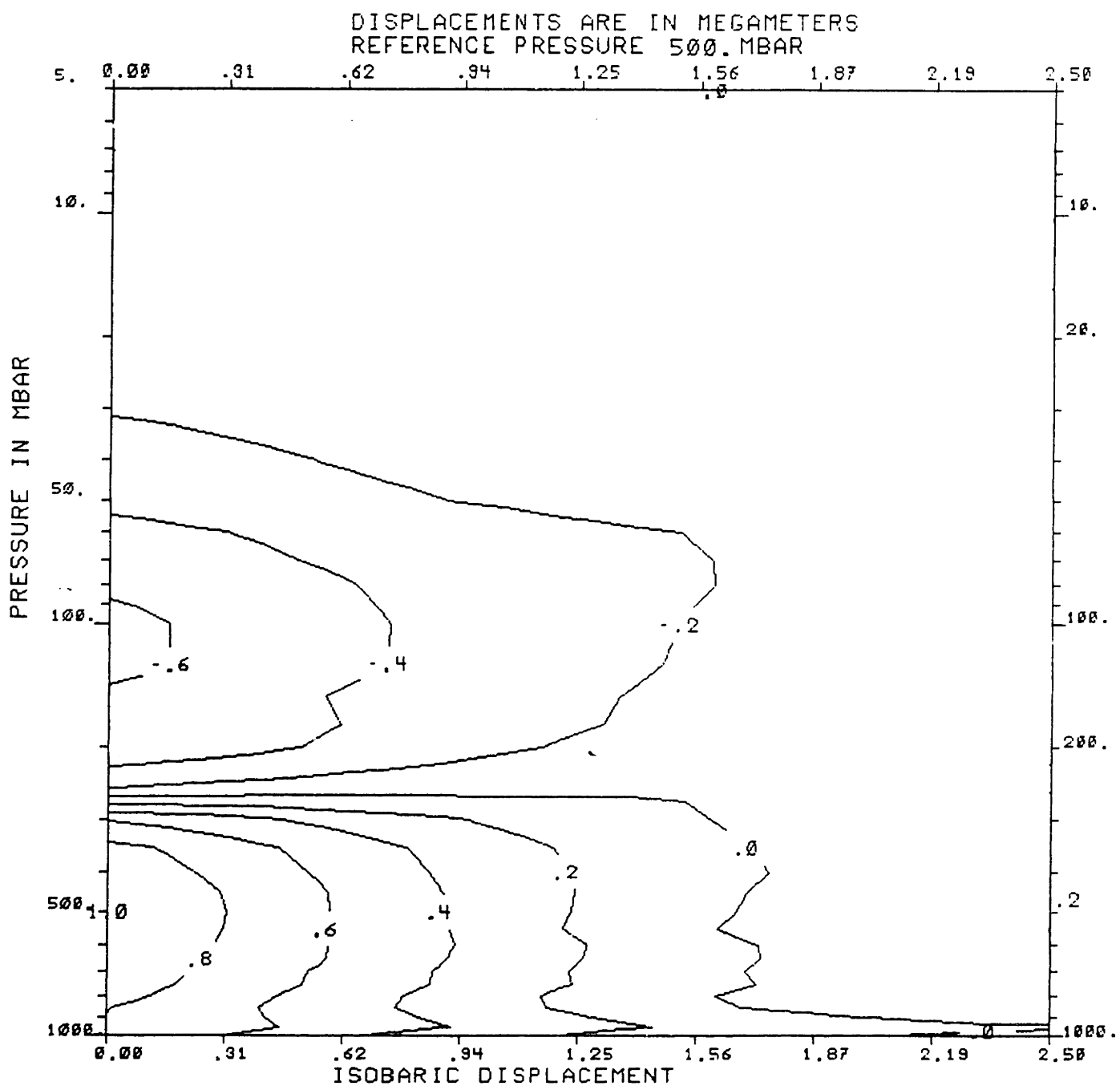


Figure III.4.7: Correlation isopleth for the developed three dimensional model. The atmospheric model is fitted to data measured over the Continental United States during the month of February 1981.

at large displacements.

This latter effect is illustrated by Figures III.4.7 and III.4.8. The model and its approximation have been fitted to rawinsonde data collected during the month of February 1981 over the Continental United States. The correlation is

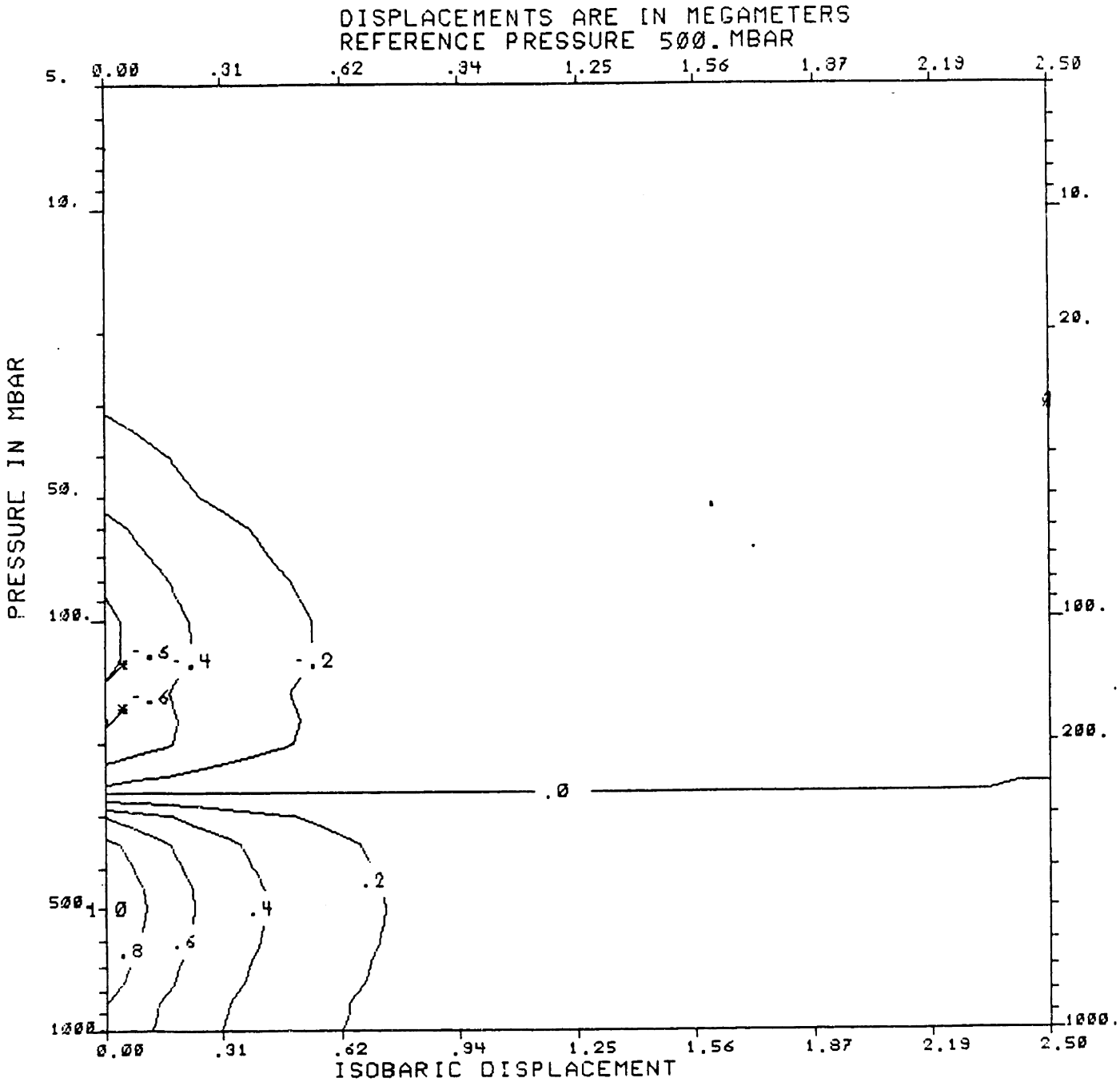


Figure III.4.8: Correlation isopleth for the approximation to the three dimensional model. The atmospheric model is fitted to data measured over the Continental United States during the month of February 1981.

normalized to one at zero displacement and a pressure level of 500 mbar. Whereas in the model, the isopleth $\rho = 0.4$ extends horizontally up to 0.95 Mm, it stops at a displacement of 0.4 Mm in the approximation. The approximation also tends to round the correlation isopleth in the pressure direction. This makes the three

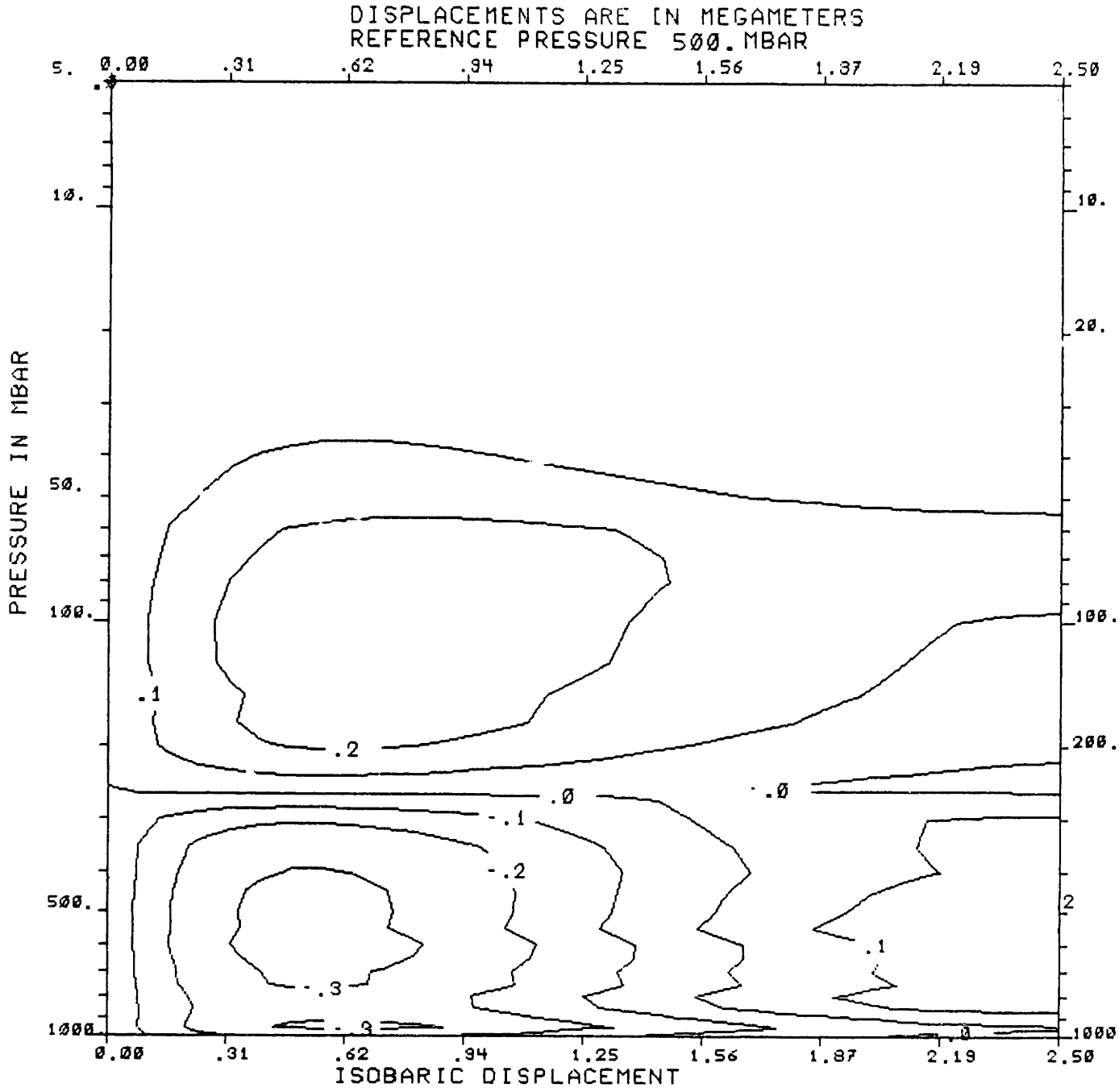


Figure III.4.9: Error in the correlation isopleths (referenced at 500 mbar) between proposed model and approximation. The atmospheric model is fitted to data measured over the Continental United States during the month of February 1981.

dimensional covariance kernel more separable and closer to the model proposed by Bergmann (1976) in multivariate objective analysis.

Figure III.4.9 presents the error in the correlation when replacing the three-dimensional model with its approximation.

Despite its truncating behavior at large displacements, the above approximation presents the advantage of offering an analytical expression for the computation of the three dimensional covariance kernel of temperature field which is mathematically tractable and which can be calculated at great speed on any digital computer.

Invoking Schur's theorem, the approximation to the model is positive definite if the kernel

$$K = \exp[-\sqrt{\underline{\xi}(p)\underline{\xi}(p')}\sigma] \qquad (III.4.47)$$

is positive definite. This is the case if the complex pole $\sqrt{\underline{\xi}(p)\underline{\xi}(p')}$ lays within the cone described in section III.4.2. Simple arithmetic with phaser representation shows that if both poles have larger real part than imaginary part, so will the square root of their product. This guarantee the positive definiteness of the approximation.

III.5 Conclusions

In this Chapter, a model for the covariance of temperature fields as a function of vertical and horizontal displacement was developed. It is the convolutional combinations of two models, one for the vertical direction and one for the horizontal directions, both of which were derived from mathematical and physical considerations. The model's validation is the subject of the next Chapter. A simple expression for the covariance which can be used in the design equation of multidimensional retrieval operators was derived.

Appendix III.A: Numerical Computation of the Hankel Transform

In this section, the numerical algorithm used to compute the Hankel transform of equation (III.4.35) is presented.

The backsmear algorithm exploits the efficiency of the fast Fourier transform to perform the Hankel transform. It is not the most efficient algorithm to date (Mook, 1983) but it is adequate whenever the computation of the Hankel transform is required at a few points or on an irregular grid. This presentation follows the presentation of Mook (1983).

The basis for the backsmear algorithm is the analysis of the Hankel transform as a double integral (see Equation III.2.15). Namely,

$$\begin{aligned} F(\sigma) &= \int_0^\infty f(k_h) J_0(\sigma k_h) k_h dk_h \\ &= \frac{1}{2\pi} \int_0^\infty \int_0^{2\pi} f(k_h) \exp(j k_h \sigma \cos\theta) d\theta k_h dk_h \end{aligned} \qquad (III.A.1)$$

Invoking Fubini's theorem, one can invert the order of integration. This yields

$$F(\sigma) = \frac{1}{2\pi} \int_0^{2\pi} \int_0^\infty f(k_h) k_h \exp(j(\sigma \cos\theta) k_h) dk_h d\theta \qquad (III.A.2)$$

Denoting by $\tilde{f}$ the one-sided Fourier transform of the product $f(k_h)k_h$, the Hankel transform becomes

$$F(\sigma) = \frac{1}{2\pi} \int_0^\pi \tilde{f}(\sigma \cos\theta) d\theta \qquad (III.A.3)$$

Using the symmetry of cosine function, this integral becomes

$$F(\sigma) = \frac{1}{\pi} \int_0^\pi \tilde{f}(\sigma \cos\theta) d\theta = \frac{1}{\pi} \int_0^{\frac{\pi}{2}} \Big(\tilde{f}(\sigma \cos\theta) + \tilde{f}(-\sigma \cos\theta) \Big) d\theta \qquad (III.A.4)$$

Introducing $\overline{f}$ the cosine transform of the original product $k_h f(k_h)$

$$F(\sigma) = \frac{1}{\pi} \int_0^{\frac{\pi}{2}} \overline{f}(k_h \cos\theta) d\theta \qquad (III.A.5)$$

The cosine transform $\overline{f}$ can be efficiently using the FFT. The integration (III.A.5) is computed numerically in the variable θ after a linear interpolator generates $\overline{f}$ at the quadrature points required by the integrator.

The algorithm takes for input $f[n]$ for $n = 0, 1, \ldots, N-1$ where the function to be transformed in truncated at some maximum range k_{hmax}. The Fourier transform of $nf[n]$ $\tilde{f}[i]$ $i = 0, 1, \ldots, N-1$ is then computed. This Fourier transform is reflectively additionned to itself to yield $\overline{f}[i]$ $i = 0, 1, \ldots, N-1$ which spans the frequency range 0 to 2π. The cosine transform $\overline{f}(\sigma)$ is linearly interpolated. Denoting by U_{-1} the Heaviside step, the interpolation can be written as

$$\overline{f}(x) = \sum_{i=0}^{N-2} \Big(U_{-1}(x-i) - U_{-1}(x-i-1)\Big)\Big(\overline{f}[i+1] - \overline{f}[i])(x-i) + \overline{f}[i]\Big) \quad (III.A.6)$$

or equivalently

$$\begin{aligned} \overline{f}(x) = &\sum_{i=0}^{N-2} ((U_{-1}(x-i) - U_{-1}(x-i-1))(\overline{f}[i+1] - \overline{f}[i])x + \\ &\sum_{i=0}^{N-2} (U_{-1}(x-i) - U_{-1}(x-i-1))((i+1)\overline{f}[i] - i\overline{f}[i+1]) \end{aligned} \quad (III.A.7)$$

The Hankel transform is found by integrating $\overline{f}(\sigma\cos\theta)$ with respect to θ.

$$\begin{aligned} F(\sigma) = &\sum_{i=0}^{N-2} (\overline{f}[i+1] - \overline{f}[i]) \int_{\cos^{-1}\frac{i}{\sigma}}^{\cos^{-1}\frac{i+1}{\sigma}} \sigma\cos\theta d\theta \\ &+ \sum_{i=0}^{N-2} (\overline{f}[i](i+1) - \overline{f}[i+1]i) \int_{\cos^{-1}\frac{i}{\sigma}}^{\cos^{-1}\frac{i+1}{\sigma}} d\theta \end{aligned} \quad (III.A.8)$$

The domain of definiteness of cos^{-1} is such that the summation on i has to be carried out for $i = 0, 1, \ldots, j-1$ where $j = \sigma k_{hmax}$. Thus

$$\begin{aligned} F(jk_{hmax}) = &\sum_{i=0}^{j-1} (\overline{f}[i+1] - \overline{f}[i]) j \Big(\sin(\cos^{-1}\frac{i}{j}) - \sin(\cos^{-1}\frac{i+1}{j})\Big) \\ &+ \sum_{i=0}^{j-1} (\overline{f}[i](i+1) - i\overline{f}[i+1]) \big(\cos^{-1}\frac{i+1}{j} - \cos^{-1}\frac{i}{j}\big) \end{aligned} \quad (III.A.9)$$

which becomes after manipulation

$$F(jk_{hmax}) = \sum_{i=0}^{j-1} (\overline{f}[i+1] - \overline{f}[i]) \left(\sqrt{j^2 - i^2} - \sqrt{j^2 - (i+1)^2}\right)$$
$$+ \sum_{i=0}^{j-1} (\overline{f}[i](i+1) - i\overline{f}[i+1]) \cos^{-1}\left(\frac{i^2 + i + \sqrt{j^2 - i^2}\sqrt{j^2 - (i+1)^2}}{j^2}\right) \qquad (III.A.10)$$

As stated, the backsmear algorithm is appropriate when the Hankel transform is required at a few points which is the case for example when evaluating the isobaric correlation between points of the Microwave Sounding Unit sampling grid.

This method of computation of the Hankel transform is most accurate when the transform $\overline{f}[i]$ is slowly varying. In the case of the proposed model, the sequence $f[n]$ to be sampled is the product

$$K_{TT}^{p}\left(\Re\left(\frac{\underline{\xi}(p)}{(\underline{\xi}(p)^2 + \Delta k.n^2)^{\frac{3}{2}}}\right)\right)^{\frac{1}{2}} \left(\Re\left(\frac{\underline{\xi}(p')}{(\underline{\xi}(p')^2 + \Delta k.n^2)^{\frac{3}{2}}}\right)\right)^{\frac{1}{2}} \qquad (III.A.11)$$

This function has a known analytical expression only when the two complex poles $\underline{\xi}(p)$ and $\underline{\xi}(p')$ are coincident.

Two cases of isobaric spectral models analytically transformed and processed using the backsmear algorithm are presented in Figures III.A.1. and III.A.2. In the first case, the spectra corresponding to a purely decaying exponential is inverted.

$$f(k_h) = \frac{\alpha}{(\alpha^2 + k_h^2)^{\frac{3}{2}}} \rightarrow F(\sigma) = \exp(-\alpha\sigma) \qquad (III.A.12)$$

1024 samples of the transform are generated between $k_h = 0$ and 6 $cycle/Mm$. This yields a resolution of about 0.03 Mm^{-1}. The decay constant α is chosen in the first case equal to 0.89 Mm^{-1} ($500mbar$, continental United States during August). In the second case, the decaying exponential is replaced by a weakly damped cosine wave (large peak at the resonance of the spectrum).

$$f(k_h) = \Re\left(\frac{\underline{\xi}}{(\underline{\xi}^2 + k_h^2)^{\frac{3}{2}}}\right) \rightarrow F(\sigma) = \exp(-\Re\underline{\xi}\sigma)\cos(\Im\underline{\xi}\sigma) \qquad (III.A.13)$$

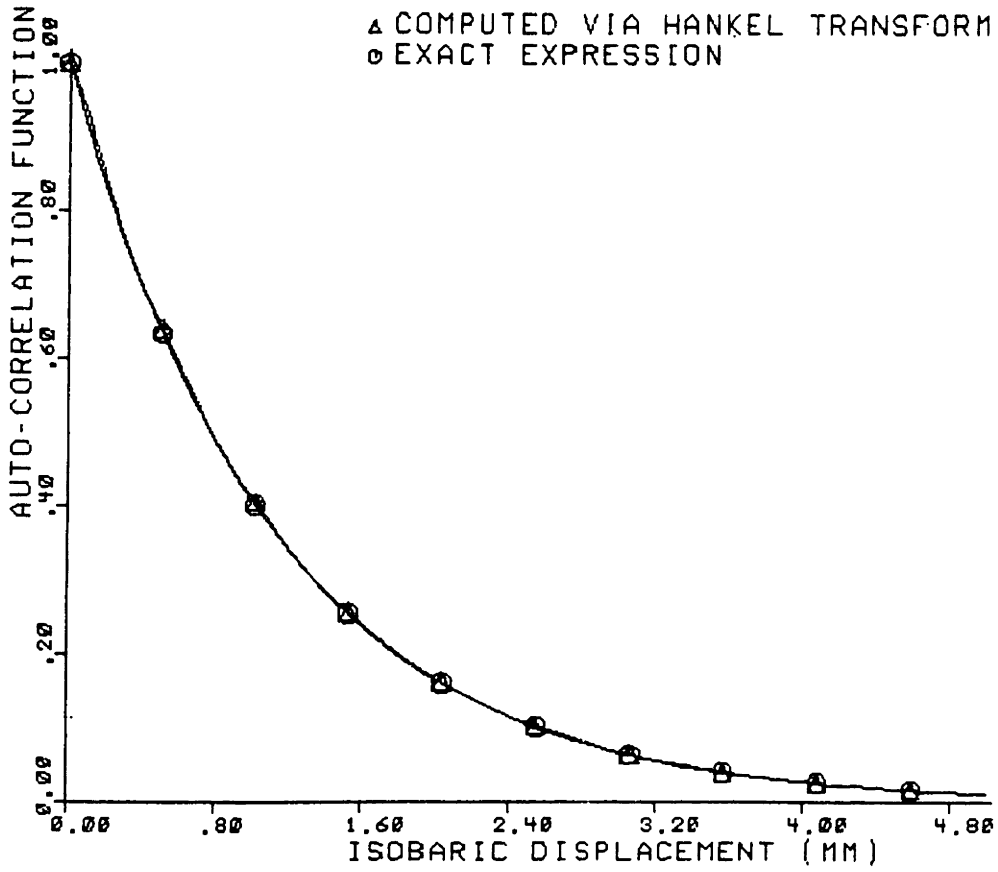


Figure III.A.1: Superposition of numerically generated Hankel transform of a two third-order real pole spectra and corresponding analytical transforms

where $\Im$ denotes the imaginary part. The sampling locations in the wavenumber domain are the same than in the previous cases. The complex pole $\underline{\xi}$ is chosen equal to $1.865 + j1.10 Mm^{-1}$ (500 *mbar*, Usa February).

In both cases, the output of the numerical algorithm matches the analytical solution up to 5 Mm. The largest error (always in the form of an overestimation of the auto-correlation function occurs at small displacement. Although not visible on these plots, the numerically computed transform tends to oscillate at large displacements because of the $\cos\theta$ interpolation of the calculations. Figure III.A.3

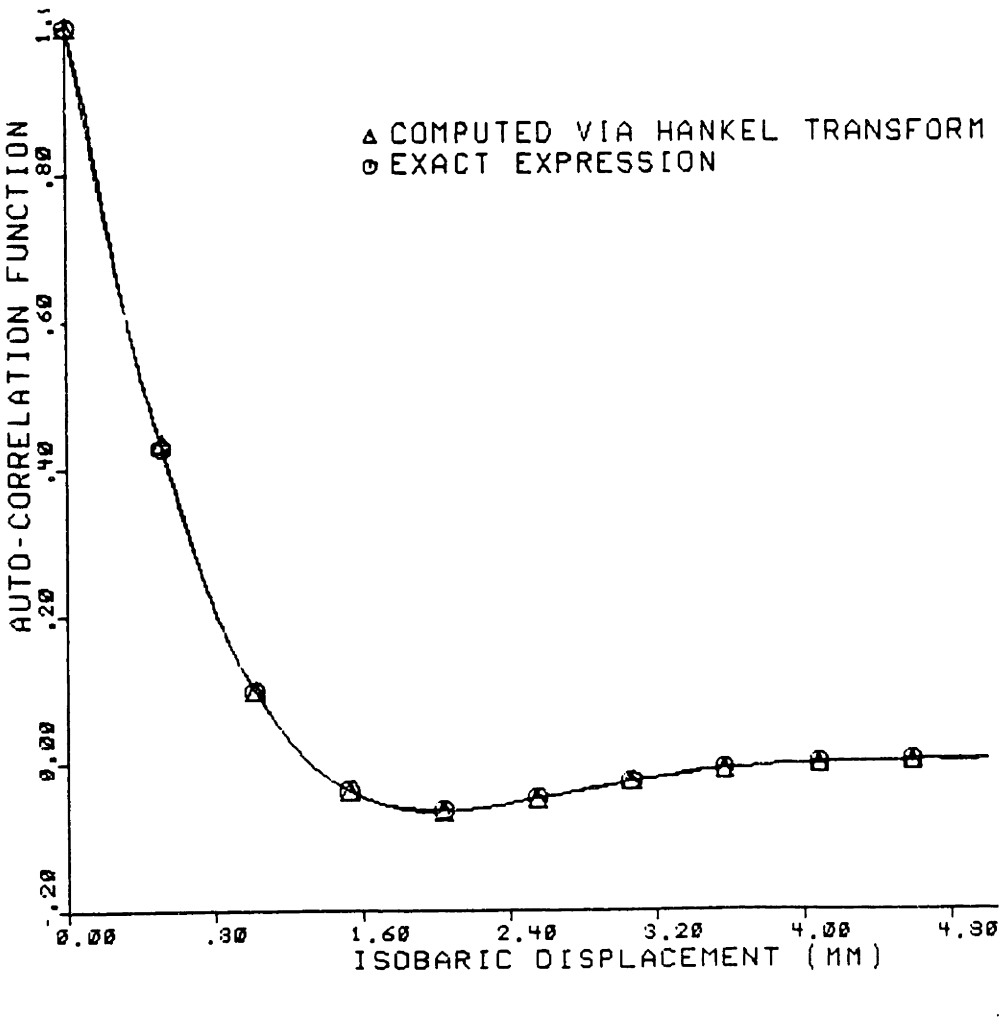


Figure III.A.2: Superposition of numerically generated Hankel transform of a two third-order complex pole spectra and corresponding analytical transforms

presents the errors associated with the first case discussed. It is always less than 4 %. This type of error is typical of the range of isobaric poles tested.

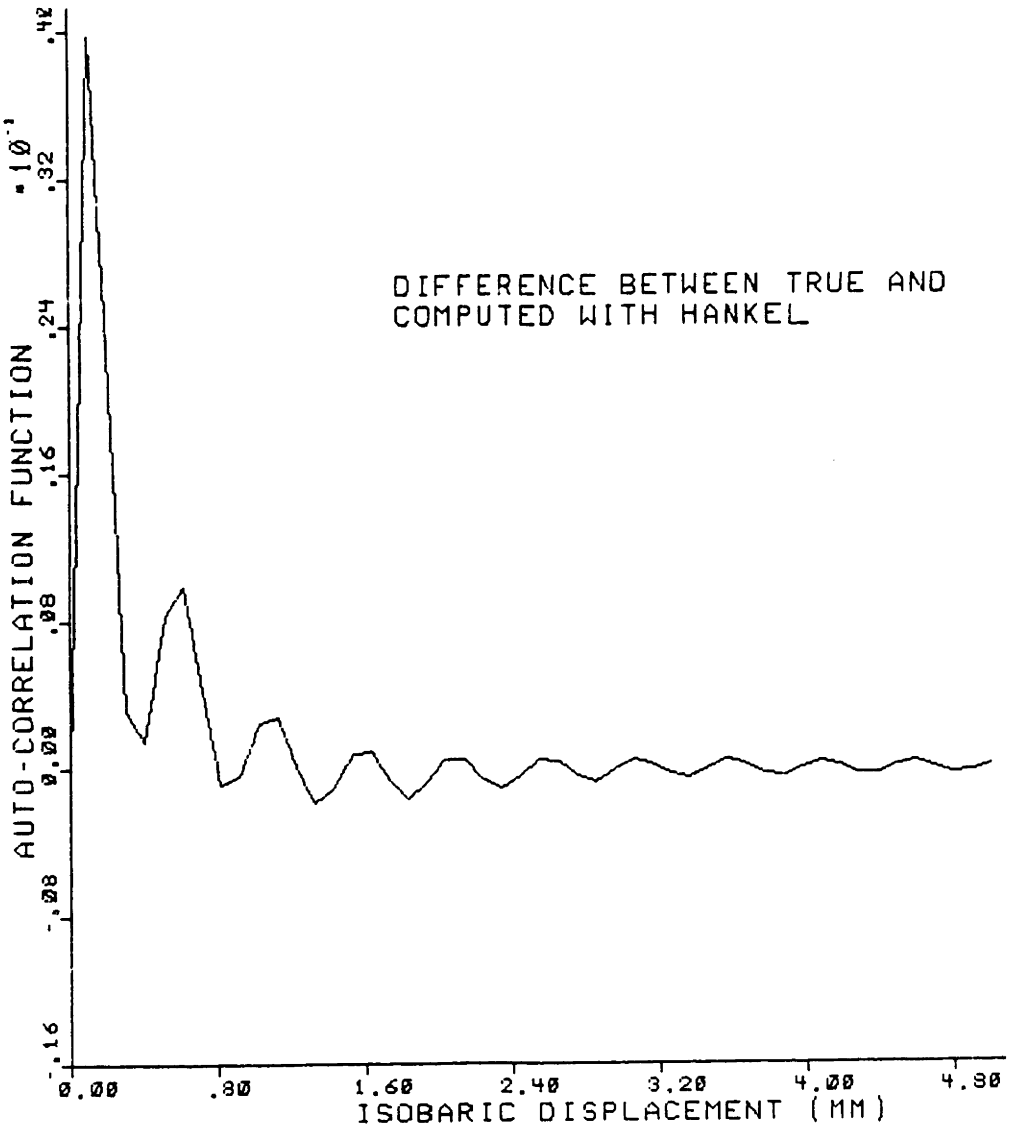


Figure III.A.3: Error between numerically generated Hankel transform of a third order complex pole spectra and corresponding analytical transforms

CHAPTER IV
THE THREE-DIMENSIONAL MACROSTRUCTURE OF ATMOSPHERIC TEMPERATURE FIELDS: IDENTIFICATION ANALYSIS AND APPLICATIONS

"Men who use ideas without measuring them are walking on wind." Flannery O'Connor, The barber

IV.1 Purpose

The fitting of the model to observed (substitute real, true) temperature and its reporting in this dissertation serves three purposes. The first is to assess the validity of the model and its shortcomings by analyzing the variations of its parameters *vis à vis* known results of atmospheric science. The second one is to provide the ground work of chapters V, VI and VII. The last one is to illustrate the major difficulties involved in collecting and processing statistical information about atmospheric processes without introducing corruption (substitute kluddge, fudge).

IV.2 The databases

The data used in this study consist of temperature fields for January 1979, February 1981, July 1979 and August 1981. These dates were chosen to coincide with the Microwave Sounding Unit (MSU) data owned by the Massachusetts Institute of Technology remote sensing group. These data are observational data of the National Meteorological Center (ADP reports) and not global analyses such as the FGGE-IVa global data set.

The reasons for that choice were the following: The NMC analysis provides data at 12 coarsely separated pressure levels (1000, 850, 700, 500, 400, 300, 250, 200, 150, 100, 70, and 50 *mbar*) and at the two standard synoptic times (McPherson

et al., 1979). Note here that these pressure levels provide the basis of comparison (the ground truth) in the retrieval experiments described in Chapters VI and VII. This fixed vertical sampling does not allow for a careful examination of changes in the fitted parameters modeling the isobaric covariance of temperature fields. This poses a problem in parts of the atmosphere that presents large changes in these parameters, such as close to the ground and across the tropopause inversion. Another reason for disregarding the NMC analysis field for the fitting of the model to observations it that the spectral and spatial content of this dataset has been altered by the Global Assimilation program of the NMC and it would therefore bias the results. The method of assimilation of the high pressure levels (70 and 50 *mbar*) is not clear and thus yield unreliable data. Lastly and more importantly, as discussed in chapter III, 12 levels are not a fine enough discrete representation of temperature profiles.

The disadvantages in using direct observation reports are numerous and at moments overwhelming. Among them, one can point out the intricate programming required to successfully decode the reports and the cleaning of the resulting temperature profiles (the NMC format for observational data handbook issued by the National Weather Service, N.O.A.A., U.S. department of commerce, is a 50 page booklet).

NMC observation data are taken at the two standard synoptic times (00Z and 1200Z) and by different types of observing instruments. Figure IV.3.1 presents the locations of the reports contained in the four months of data used in this study.

Only the rawinsonde data (RAOB) were considered in this study. These reports (if complete) provide temperature profiles at 20 mandatory pressure levels (1000, 850, 700, 500, 400, 300, 250, 200, 150, 100, 70, 50, 30, 10, 7, 5, 3, 2, and 1 mb) and at the significant pressure levels. The latter are defined as the levels at which the slope of the variation of temperature with height changes (Lenhard, 1970,

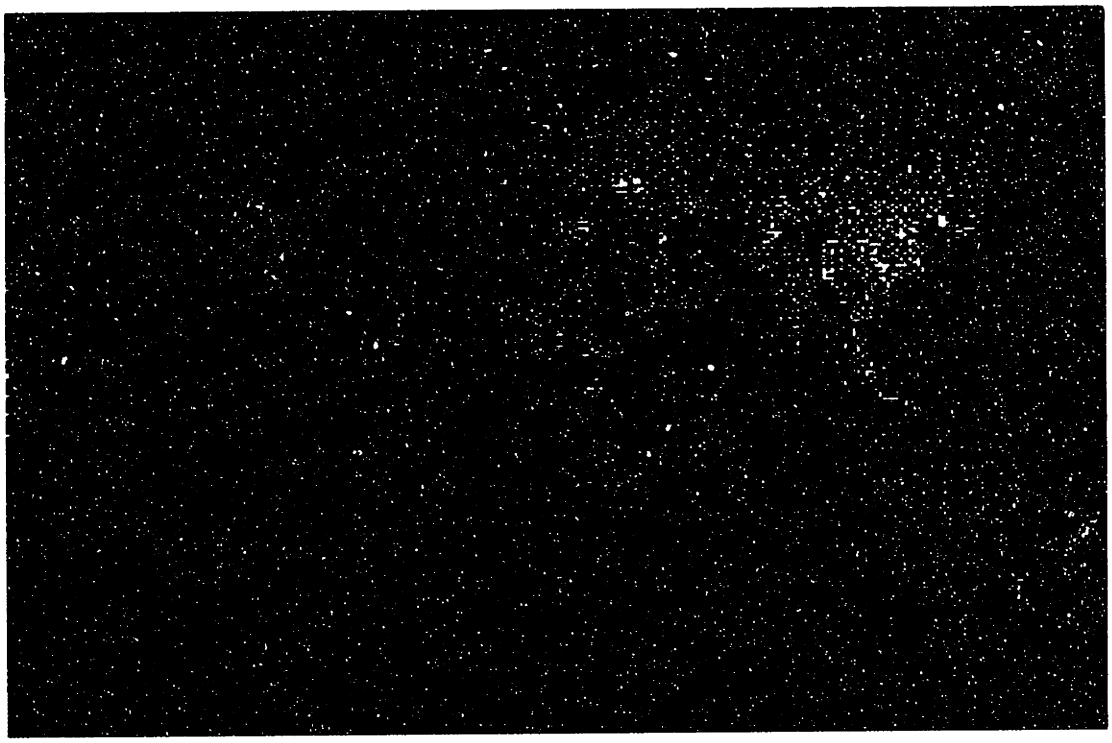


Figure IV.3.1: Location of the ADP observation data for January and July 1979 and February and August 1981.

1973). These readings allow one, by linear interpolation in the height domain, to estimate the temperature at any pressure level between the lowest and the highest recorded temperatures. Reports with less than 5 mandatory pressure levels were disregarded. For each reporting location, the minimum number of reporting days within a month was set at 10 in order to eliminate stations with bad reporting records. The profiles were interpolated and sampled at 33 pressure levels, namely

1000, 950, 900, 850, 750, 700, 650, 600, 550, 500, 450, 400, 350, 300, 250, 200, 175, 150, 125, 100, 80, 70, 60, 50, 40, 30, 25, 20, 15, 10, 7, and 5 mbar. In addition, the means and variances of the temperature for these 33 pressure levels were computed on a monthly basis at each reporting station. Any reading whose deviation from the mean was larger than two and a half times the corresponding standard deviation was rejected. This cleaning process was repeated twice to eliminate abnormal readings whenever possible. The reports were divided into different regions. In both all cases, the maximum number of stations considered in the fits was limited to 50. This number of stations provides 1275 baselines distributed in length from typically .05 to 3 Mm.

CONUSA Continental United States; 48 stations located in longitudes between 135 West to 75 West and latitudes between 34 North and 54 North. This region has been analyzed in great detail by Thiebaux (1975, 1976), Kao (1970), and Schlatter, (1975).

EUROPE Western Europe; 50 Stations located to longitudes between 10 West and 30 East and latitudes between 35 North and 55 North. These two regions do not present very homogeneous temperature fields, but have good quality readings over a large range of isobaric displacement. Boltenkov (1966) compared the behavior of the correlation function (strictly speaking of the structure function) of temperature fields at different pressure levels as a function of displacement. These two regions present the advantage of having numerous reporting stations. This yields an analysis field of good quality and good ground truth for comparison with remote sensing experiments.

Figure IV.3.2 presents the locations of the reporting stations of the CONUSA, EUROPE datasets. Figures IV.3.3 and IV.3.4 present the baselines corresponding to these two fitting regions. These baselines are presented in a Mercator projection. That is, the distance between the two stations is projected on the zonal

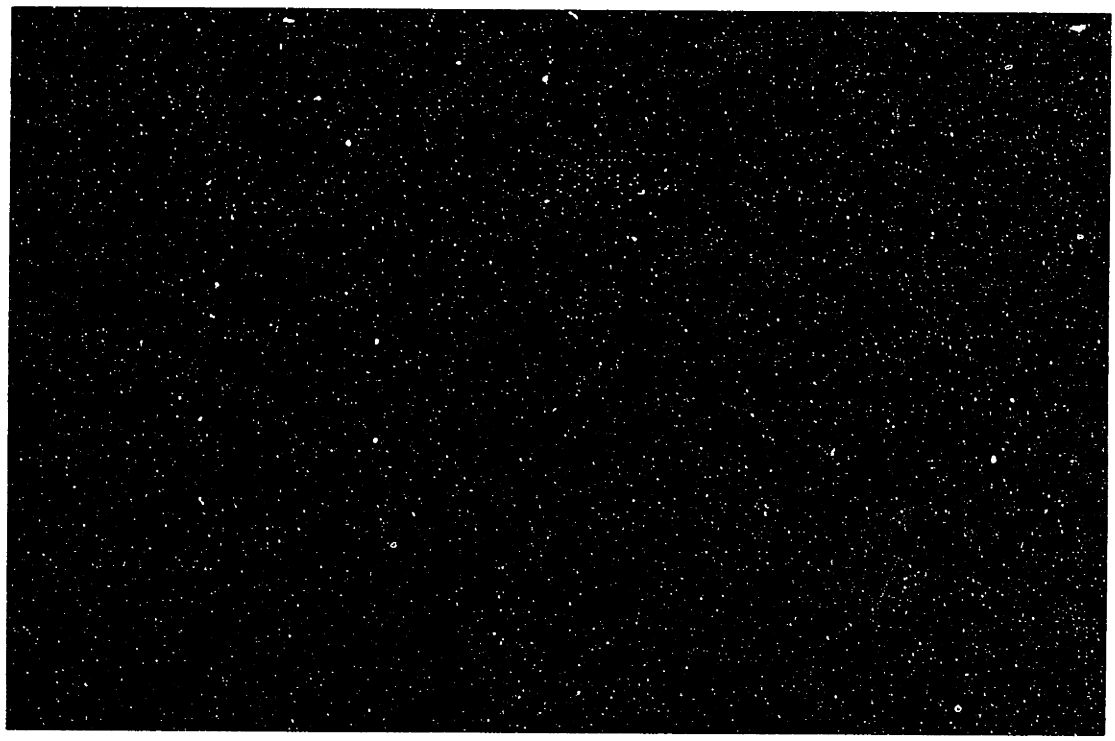


Figure IV.3.2: Location of the reporting stations corresponding to the CONUSA and EUROPE fitting regions.

and the equatorial axes according to the angle formed by the differences in latitude and longitude of the two stations. Since for each pair of recording stations, two symmetrical baselines can be formed with respect to the origin, only the baselines with a positive zonal component are plotted.

IV.3 Parameter Identification

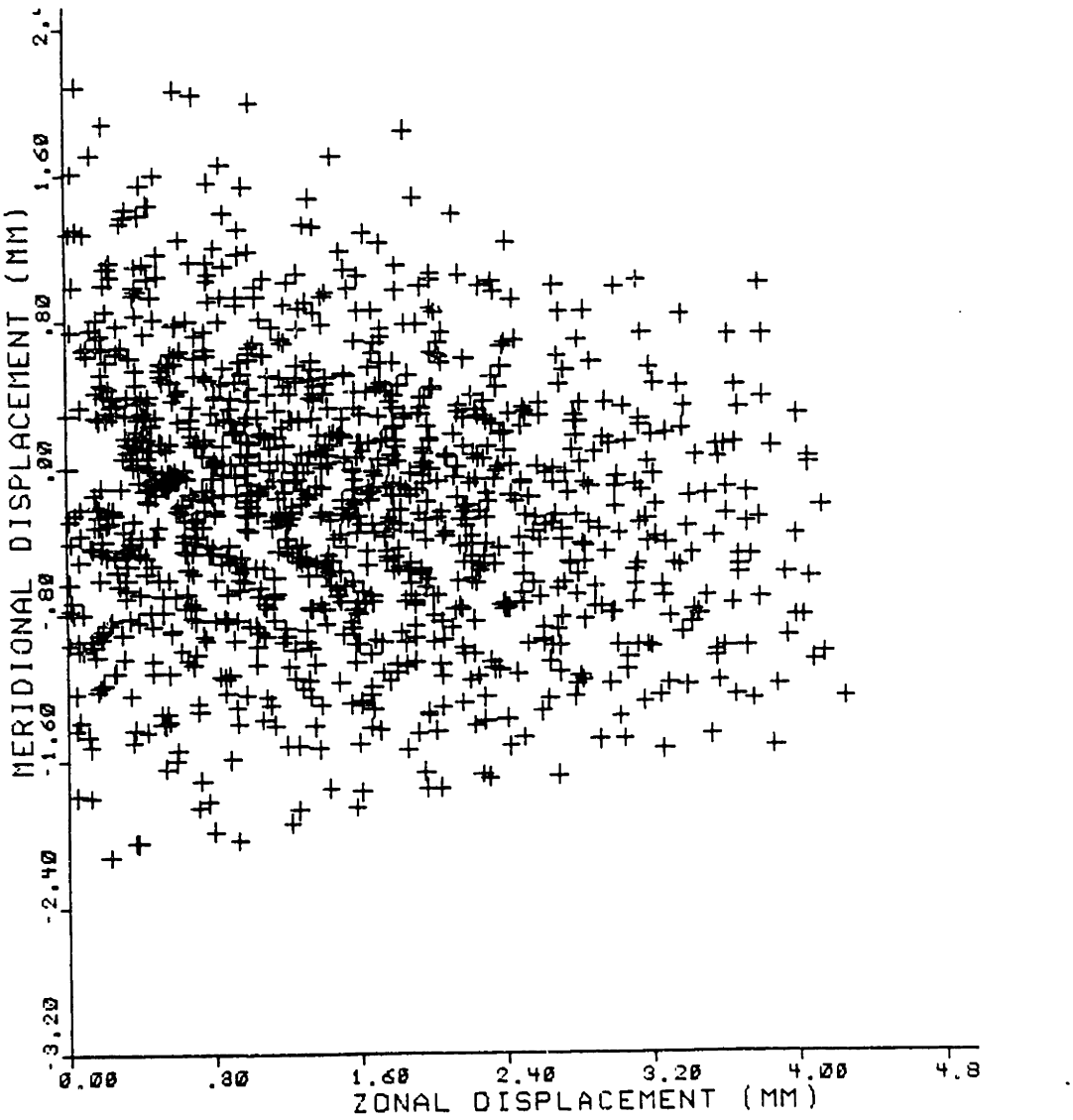


Figure IV.3.3: 1176 baselines corresponding to the CONUSA region (Mercator Projection)

The estimation for the isobaric domain follows the discussion of chapter III. That is the metric characterizing the relevant between two points is computed (it will determine the form of the isobaric correlation isopleths), then the isobaric covariance function *per se* is considered

IV.3.1 Anisotropy map

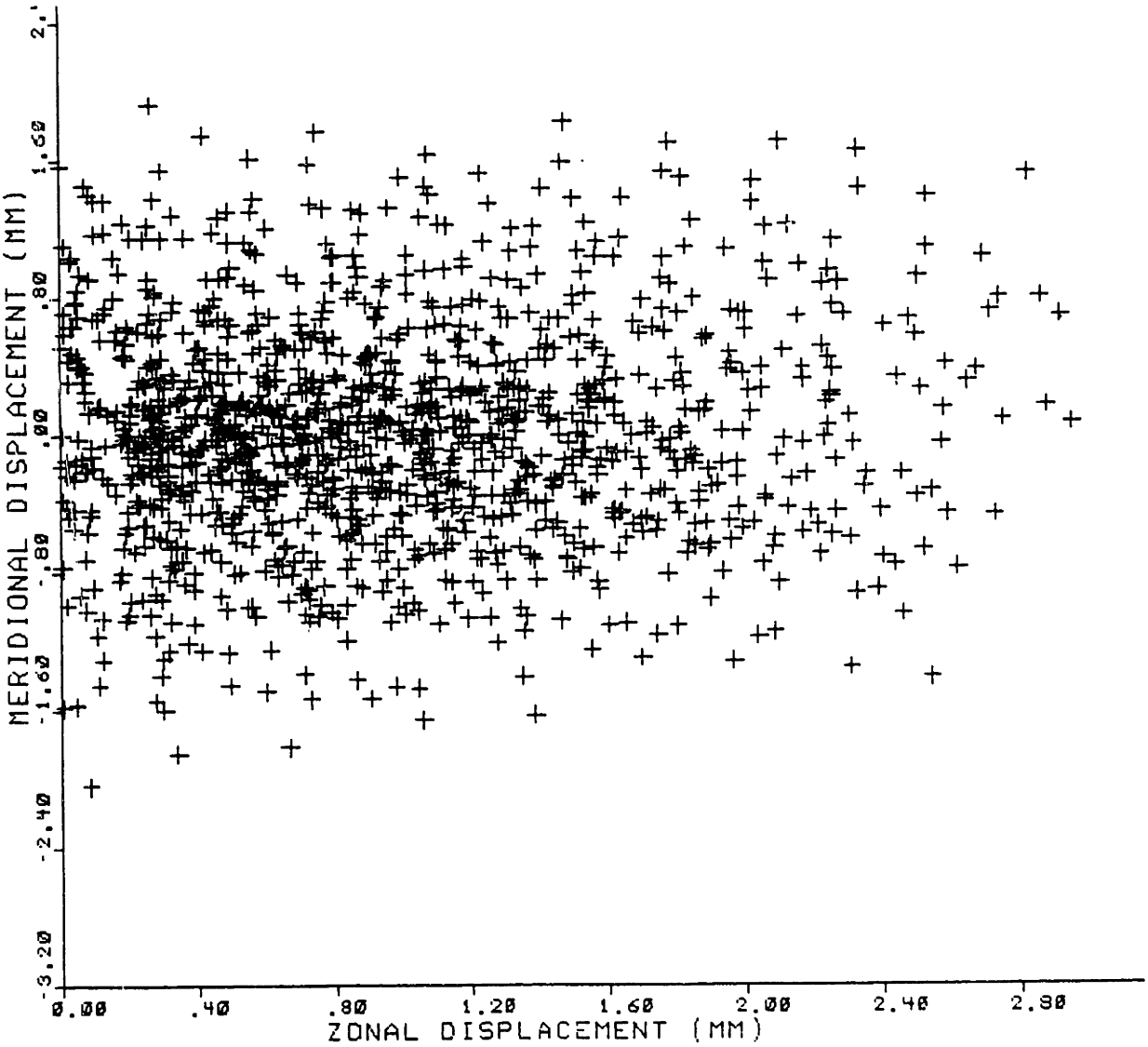


Figure IV.3.4: 1275 baselines corresponding to the EUROPE region (Mercator Projection)

The approach taken to the estimation of the anisotropy map is first to assume a model for the correlation isopleth independently from the function describing the isobaric covariance kernel, and then to derive an estimator for the parameters of the mapping. This estimation process could be made extremely sophisticated and yield interesting research topics. The purpose of this section is to derive some general considerations concerning this estimation problem and then derive a simple

estimator.

In the case of interest, the mapping is ellipsoidal and characterized by a contraction factor a and an azimutal orientation for the major axis v. Let us introduce the distance metric $m(\sigma_x, \sigma_y, a, v)$ defined as

$$m(\sigma_x, \sigma_y, a, v) = (\sigma_x^2 a^2 + \sigma_y^2)\cos^2 v + (\sigma_x^2 + a^2\sigma_y^2)\sin^2 v \qquad (IV.3.1)$$

The model for the isobaric covariance is $K_{TT}(r, \underline{\xi}(p))$ where r is the displacement from the origin computed using the metric m and $\underline{\xi}(p)$ is the now familiar complex pole. Let $K^o_{TT}(\sigma_{xi}, \sigma_{yi})$ $i = 1, \ldots, N$ denote the observed samples of the isobaric covariance function. Finally, let $||.||$ denote a norm (not necessarily Euclidian). Given the norm, the error between model and observed covariance can be assessed.

Given the norm $||.||$, the optimum elements of the mapping are found by solving

$$\hat{a}, \hat{v} \leftarrow \min_{a,v} ||K_{TT}(m(\sigma_x, \sigma_y, a, v), \underline{\xi}(p)) - K^o_{TT}(\sigma_x, \sigma_y)|| \qquad (IV.3.2)$$

Denoting by κ either a or v, the minimization can be found by setting the partial derivative of the error term with respect to κ to zero, namely

$$\frac{\partial||.||}{\partial K_{TT}} \frac{\partial K_{TT}}{\partial m} \frac{\partial m}{\partial \kappa} = 0 \qquad (IV.3.3)$$

The first term corresponds to the sensitivity of the error norm to its argument and is independent of the specific analytical expression of the isobaric covariance kernel. The last term depends on the metric m. It increases with the displacement from the origin as the square of the displacement. The second term involves the slope of the covariance kernel. This is natural since if the slope becomes infinite (or at least very large), it provides an excellent discriminator for the anisotropic mapping, independent of the specific norm used. Thus, although an accurate estimate of the anisotropic mapping's parameters does not require an exact knowledge of the

isobaric covariance, it is necessary to weight the error norm so that it peaks at the location where the variations of the derivative are the largest.

For a decaying exponential, the derivative is maximum at the origin and decreases with displacement. Because in the case of interest, the two partial derivatives $\frac{\partial K_{TT}}{\partial m}$ and $\frac{\partial m}{\partial \kappa}$ behave in opposite fashions as the displacement varies, the product of these two terms will present one (or several) maximum at which the estimation should be performed.

That displacement is found by maximizing the derivative of

$$\frac{\partial K_{TT}}{\partial m}\frac{\partial m}{\partial \kappa} \approx K_{TT}(0)\sigma^2 \exp(-\sigma) \qquad (IV.3.4)$$

The derivative is maximum at $2-\sqrt{2}$ and minimum at $2+\sqrt{2}$. These displacements correspond to the normalized correlations ($K_{TT}/K_{TT}(0)$ of 0.55 and 0.03 respectively. In most cases, the level $K_{TT} = 0.03K_{TT}(0)$ level is not of statistical importance. This is true, because at such distances, topographic effects and standing oscillations patterns (see section III.2.1) enter into account. This implies that the estimation of the parameters of the ellipsoidal mapping must be performed around the first of these two isopleths, that is around $K_{TT}/K_{TT}(0)$ equal 0.5.

The last element to consider is the error norm. This error norm is found by analyzing the variations of the radius of an isopleth as a function of the azimutal angle θ. This is another way of saying that the error criterion is tailored to the estimator. Luckily, as it will appear, this criterion is one commonly used in estimation theory. Simple trigonometry shows that along an ellipse the radius obeys the law

$$\begin{aligned} r^2 &= \cos^2(\theta - v) + a^2 \sin(\theta - v) \\ &= \frac{a^2+1}{2} + \frac{1-a^2}{2}\cos 2(\theta - v) \\ &= \frac{a^2+1}{2} + \frac{1-a^2}{2}\cos 2v \cos 2\theta + \frac{1-a^2}{2}\sin 2v \sin 2\theta \end{aligned} \qquad (IV.3.5)$$

This variation of r^2 is a simple trigonometric sum whose three components can be estimated using standard Fourier analysis. These coefficients can also be found

by minimizing the mean square error between the observed isopleth displacement squared and the trigonometric sum.

Let us, thus, take for norm $||.||$ in equation (IV.3.3) the mean square error between the square of the displacement of the observed $K_{TT}/K_{TT}(0)$ equal 0.5 isopleth and the trigonometric sum (IV.3.5). The observed isopleth is found by bi-linearly interpolating the observed correlation at different angles and displacements. The solution to this estimation is then given by

$$r^2 = \lambda_0 + \lambda_1 \cos 2\theta + \lambda_2 \sin 2\theta \qquad (IV.3.6)$$

where

$$\begin{aligned} \lambda_0 &= \frac{1}{2\pi} \int_0^{2\pi} r_o^2(\theta) d\theta \\ \lambda_1 &= \frac{1}{\pi} \int_0^{2\pi} r_o^2(\theta) \cos 2\theta d\theta \\ \lambda_2 &= \frac{1}{\pi} \int_0^{2\pi} r_o^2(\theta) \sin 2\theta d\theta \end{aligned} \qquad (IV.3.7)$$

The parameters of the ellipsoidal mapping are then

$$\hat{a} = \sqrt{\frac{1 - \frac{\sqrt{\lambda_1^2+\lambda_2^2}}{\lambda_0}}{1 - \sqrt{\frac{\lambda_1^2+\lambda_2^2}{\lambda_0}}}} \qquad (IV.3.8)$$

and for $\hat{a} \neq 1$

$$\hat{v} = \frac{1}{2} \tan^{-1} \frac{\lambda_2}{\lambda_1} \qquad (IV.3.9)$$

This estimator was implemented using a contouring routine to generate the isopleth $r_o(\theta)$ and the integrations were computed numerically using a 16-point quadrature.

This estimation technique proved to be extremely noisy for both a and v. A noisy (i.e., non-robust) estimate for the azimutal orientation of the major axis orientation is expected when the contraction factor a approaches one because v becomes undefined and the estimation pathological. The contraction factor was found to be between 0.9 and 1.2 depending on the pressure level. These values are in

agreement with the work reported in Chapter III. For the purpose of estimating the isobaric poles and the three-dimensional covariance kernel, the contraction factor was set equal to one for all levels, in all cases. That is to say, that the atmospheric temperature fields were assumed isotropic and the orientation of the satellite's track with respect to the earth does not enter into account in the computation of the retrieval operators.

IV.3.2 Isobaric complex poles

For each one of the regression regions, the cross correlation between each of the reporting stations is computed and mapped onto a linear displacement array (therefore assuming isotropy of the covariance field). Isobaric covariances are limited to the maximum baseline in the regression region or 3000 km, whichever is smaller. This linear displacement array is then partioned in 64 bins of width 50 km or smaller. The bin partitioning is in fact a low pass filtering of the correlation which reduces the variance of the estimates of the correlation by increasing the number of degreees of freedom entering each of the observed covariances.

The Fisher z transformation (Bulmer, 1967, Rao, 1973) was used to assess the statistical reliability of the estimates of the covariances and to allow an optimal χ^2 fit. Denoting by ρ the sample correlation (normalized at the origin) between two stations, the quantity z is defined as

$$z = \frac{1}{2}\log\left(\frac{1+\rho}{1-\rho}\right), \qquad (IV.3.10)$$

which is an unbiased asymptotically normal estimate of the transformation of the true correlation and has for variance $(N-1)^{-1}$, where N is the number of degrees of freedom in the series used to estimate ρ.

The estimation of the number of degrees of freedom is an intricate problem because the correlations are estimated as a space and time average. This estimation is performed in appendix IV.A, using a frozen atmosphere model.

When fitting correlations with noisy data, one must pay attention to the value of the correlation at zero displacement (Gandin, 1963, Rutherford, 1972). Let $z = x+\eta$ denote the observation equation where x is the true value of the parameter (in this case, the value of the temperature at a specific pressure level), z is the measured value, and η is the measurement noise, assumed white and uncorrelated with x. This latter assumption yields a best case analysis, since correlated error signals in neighboring observations reduce the informational content (Bergmann and Bonner, 1976). Let R_{xx} denote the (true) covariance of x and R_{zz} the measured covariance in z. The observation equation yields

$$R_{zz}(\tau) = R_{xx}(\tau) + R_{\eta\eta}(\tau) \qquad (IV.3.11)$$

where τ denotes the displacement.

The true normalized correlation is $R_{xx}(\tau)/R_{xx}(0)$, which equates one to the origin whereas the observed correlation is $R_{xx}(\tau)/(R_{xx}(0)+R_{\eta\eta}(0))$, which equates $R_{xx}(0)/(R_{xx}(0)+R_{\eta\eta}(0))$ to the origin.

The observed correlation should thus not be fitted to one for zero displacement. This DC value varies from the value of Julian (1975)

$$\rho_{Julian}(0) = \frac{R_{xx}(0) - R_{\eta\eta}(0)}{R_{xx}(0)} \qquad (IV.3.12)$$

The two expressions are nevertheless equivalent for large signal-to-noise ratios, for

$$\rho_{this\ work}(0) = \frac{1}{1+\frac{R_{\eta\eta}(0)}{R_{xx}(0)}} \approx 1 - \frac{R_{\eta\eta}(0)}{R_{xx}(0)} \qquad (IV.3.13)$$

Typical value for the covariance of radiosonde temperature measurement is of the order of 0.5 K^2. The covariance $R_{xx}(0)$ for compiled temperature profiles depends on the specific pressure level, the climatology, and the number of degrees of freedom in the compilation. For the July data over the continental United States, it varies from 56 K^2 at 1000 *mbar* to 5 K^2 at 30 *mbar*.

The isobaric parameters, decay constant α and oscillatory constant ω, were estimated by minimizing the mean square error between the observed correlation and the one derived from the model. Let $y[n]$ $(n = 0, 1, \ldots, 63)$ denote the sample correlation for each displacement in the linear displacement array (not all $y[n]$ are measured). Let $\tilde{y}[n]$ denote the sequence of estimated correlations computed using the model, namely

$$y[n] = \frac{R_{xx}(0)}{R_{xx}(0) + R_{\eta\eta}(0)} \exp(-\alpha dn) \cos(\omega dn) \qquad (IV.3.14)$$

where d is the bin width for the linear displacement array (about 50 km). The error norm being minimized is E

$$E = \sum_{n=0}^{63} b[n] w[n] (y[n] - \tilde{y}[n])^2 \qquad (IV.3.15)$$

where $b[n]$ is a binary function indicating if the correlation is measured or not, and $w[n]$ is a weight. If the weight equates the inverse of the variance in the observed data, This fit will be optimal from a χ^2 point of view (check section IV.5). Such a fit put more emphasis on observations measured with good accuracy. The weight can also be set uniformly to one in order to give more emphasis (all things considered) to correlations associated with long baselines. The former solution was chosen in this thesis.

Posed in these terms, this problem is equivalent to the time domain estimation of complex poles in time series, that is the estimation of damped sinusoidal signals buried in noise (Henderson, 1981). The efficient linear prediction techniques used to solve the time series problem could not be used in this specific identification problem because gaps were present in the observed correlation (Kumaresan and Tufts, 1982).

Moreover, since the error is not quadratic in the parameters, an iterative search for the optimal parameters was implemented using a conjugate gradient search (Ortega and Rheinbolt, 1970, Luenberg, 1976).

This latter method yields the maximum likelihood estimates of the complex pole $\underline{\xi} = \alpha + j\omega$, which in turns satisfies the Cramer Rao bound (Rife and Boorstyn, 1974).

In order to reach the global minimum of the error norm, the initial estimates must lie in the main valley of the error criterion. Because neighboring pressure levels have comparable poles, the a priori estimate at a specific pressure level was chosen to be the final estimate at the pressure directly above it (Blackmon, 1976, Blackmon *et al*, 1979). The highest pressure level (5 *mbar*) was fitted first and the lowest (1000 *mbar*) last. High pressure levels are badly sampled, resulting in fewer points, and thus are better conditioned to be fitted through.

To illustrate the behavior of the error term as a function of the fitted parameters, the error fields were computed for two extremes cases and the isoerror curves were plotted next to the observed correlation in Figures IV.3.1 and IV.3.2. Figure IV.3.1 presents the normalized error (that is, normalized to one at its minimum) in the case where the decay constant is much larger than the oscillatory term. The measurements are taken over the continental United States during the month of January 1979 at 100 mbar. Figure IV.3.2, on the other hand, presents the case where the decay and oscillatory constant are comparable (February 1981, USA, 800 mbar).

Since the cosine is an even function of its argument, the error field is symmetric with respect to the decay constant axis. The error is well behaved in the 0 to 2 Mm^{-1} range for both decay (α) and oscillatory (ω) terms, in the sense of a well defined unique valley. An interesting characteristic of the error field which will have implications in the accuracy of the estimates $\hat{\alpha}$ and $\hat{\omega}$ is the difference in the valley width in the α and the ω directions. Because the error increases slowly from its minimum value as ω departs from its optimum value, estimates of the oscillatory term will be noisier than estimates in the decay term. Also, the error field flattens

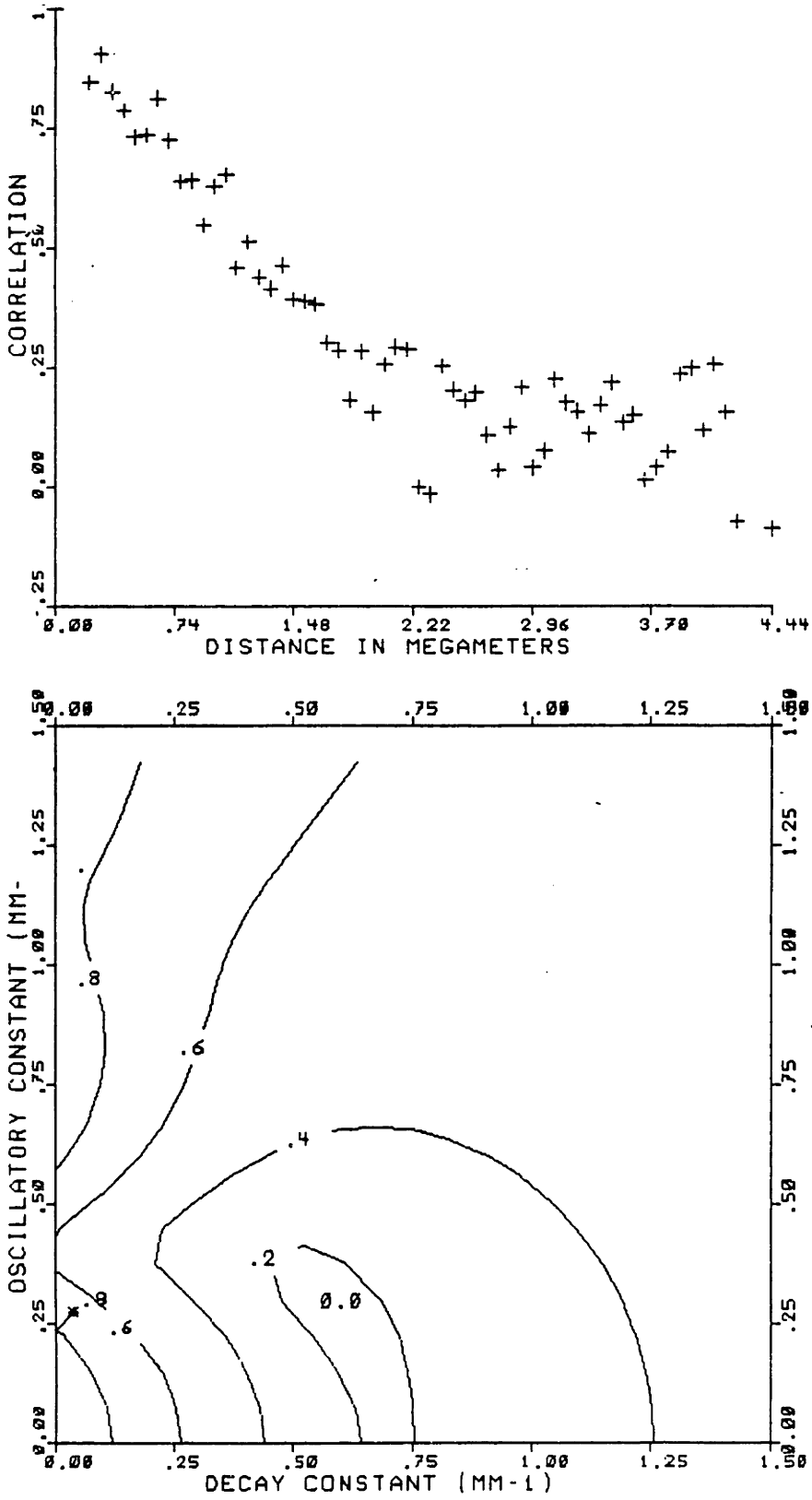


Figure IV.3.1: Observed correlation and isoerror curves for January data over the continental United States at 100 mbar

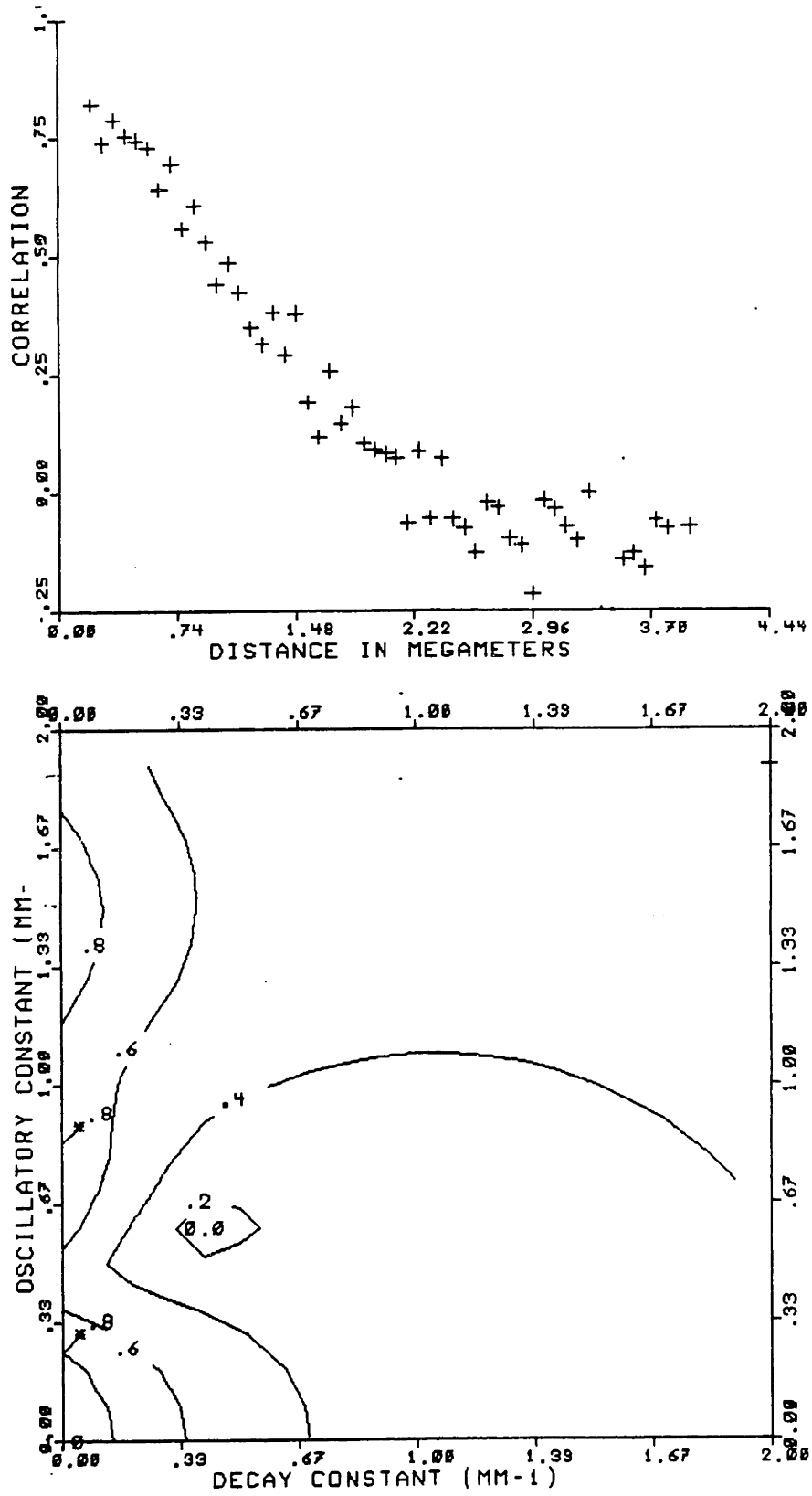


Figure IV.3.2: Observed correlation and isoerror curves for February data over the continental United States at 800 mbar

as the decay constant increases. This implies that in the case of noisy measurements or few points, estimates of α will have a tendency to be overestimated, thus resulting in short isobaric correlation lengths at those pressure levels. This will be the case at high altitudes which are rarely measured by radiosondes.

An important quantity associated with the isobaric complex pole is the "scale of fluctuation" (VanMarcke, 1983, Tennekes and Lumley, 1972) defined as (in the one dimensional case) as

$$\theta = \lim_{T \to \infty} \gamma(T) \qquad (IV.3.16)$$

where $\gamma(T)$ is the variance function of the process $x(t)$, defined as the ratio of the process integrated over the interval T to its variance. That is

$$\gamma(T) = \frac{1}{\sigma^2} < \left(\frac{1}{T} \int_{t-T/2}^{t+T/2} x(t')dt' \right) > \qquad (IV.3.17)$$

The scale of variation is important because it corresponds to the duration of an independent measurement, and it is important to characterize the time interval required to obtain low variance estimates of the mean of fluctuating quantities, and error bars to the estimates of their covariance function (see appendix IV.A).

This concept extends to two dimensional processes. For a two dimensional isotropic process with a correlation function such as $K \propto \exp(-\alpha\sigma)\cos(\omega\sigma)$, the scale of fluctuation is

$$\theta = \frac{\pi\alpha}{\alpha^2 + \omega^2} \qquad (IV.3.18)$$

IV.3.3 Vertical Covariance Matrix

The main issue in estimating the vertical covariance matrix is to ensure its positive definiteness. Non-positive definiteness can occur for two main reasons: not enough degrees of freedom in the compilation of temperature profiles and bad combination of different subcovariance matrices.

Avoiding the first of these problems requires that to estimate a covariance matrix of size l, one must compile at least linearly independent temperature profiles. This is true because the estimate $\overline{K}$ for the matrix is the mean value of outer product, namely

$$\overline{K} = \frac{1}{N} \sum_{i=0}^{N-1} \overline{T}_i \overline{T}_i^t = \frac{1}{N} \sum_{i=0}^{N-1} M_i \qquad (IV.3.19)$$

where the $\overline{T}_i$ are the temperature profiles (with their means subtracted) and ^{t} denotes the vector (matrix) transposition. Each matrix M_i (defined as the outer product $\overline{T}_i \overline{T}_i^t$) is of rank one and is positive semidefinite.

Because the rank of the sum of matrices is bounded above by the sum of the ranks of the matrices, at least l independent matrices M_i are necessary. This minimum number of profiles is the price paid for not having an analytical expression for the vertical covariance kernel. Should the number of degrees of freedom be too small, the resulting matrix will be positive semidefinite which is still acceptable.

The second problem arises when one estimates the covariance matrix of non-zero mean valued vectors whose components are not always sampled. This problem arises from the fact that because of missing pressure levels, cross products cannot be evaluated as often as the mean value of the profiles. Even when the running sum for the cross products and for the mean are weighted by the appropriate number of samples used to evaluate them, the resulting can have negative eigenvalues. This is the case of ADP temperature reports which are typically missing at high altitude, or when merging rocketsonde with radiosonde data. Because the highest registered level varies from report to report, they must be differentiated.

To illustrate this problem, let us consider the following example of the observations of three realizations of a two-component random process.

$$\overbrace{\begin{pmatrix} \bullet \\ 1 \end{pmatrix} \begin{pmatrix} 2 \\ 2 \end{pmatrix} \begin{pmatrix} 1 \\ \bullet \end{pmatrix}}^{\text{observations}} \qquad (IV.3.20)$$

The sample mean and covariance matrix for this realization based on all possible data is

$$\begin{pmatrix} \frac{3}{2} \\ \frac{3}{2} \end{pmatrix} \begin{pmatrix} \frac{1}{4} & \frac{7}{4} \\ \frac{7}{4} & \frac{1}{4} \end{pmatrix} \qquad (IV.3.21)$$

The covariance matrix has for eigenvalues 2 and -3/2, which is not positive definite.

One possible solution is to extrapolate the recorded level to the level missing above the highest recorded level and from the lowest recorded level down to the sea level (1015 *mbar*). This extrapolation can be statistical in nature or based on the concept of band-limited extrapolation.

Band-limited extrapolation assumes than the samples of the temperature profile are sufficient to reconstruct exactly the profile at the missing pressure levels. This sufficiency is introduced in terms of an equivalent Nyquist rate in the sampling (Papoulis, 1975, Shafer *et al.*, 1981) which allows us to switch back and forth between the spatial (vertical) representation of the profile and its Fourier transform. This method is not based on any physical understanding of the series being extrapolated, it is time consuming, and is extremely ill behaved when the samples of the profiles are close to each other. Furthermore, when the number of samples is too small, the resulting temperature profiles are very smooth, resulting in no inversion at the tropopause. When thought of as an interpolation method from a finite grid onto another finite grid, this technique is equivalent to mean-square fitting of a polynomial on irregularly spaced data. This latter approach has been characterized in terms of robustness and has been proven to be ill-conditioned (Berztiss, 1964). For these reasons, this method was abandoned.

Statistical extrapolation is based on static Bayesian estimation. Let T_{33} denote the temperature profile sampled at all 33 pressure levels and T_l the measured profile padded with zeroes at the pressure not measured. The observation equation is

$$T_l = \Sigma_l T_{33} \qquad (IV.3.22)$$

where Σ_l is a diagonal matrix of zeroes and ones. The Bayesian estimate of T_{33} is (Gelb, 1974)

$$\hat{T}_{33} = K^p_{TT}\Sigma_l(\Sigma_l K^p_{TT}\Sigma_l)^{-1}T_l \qquad (IV.3.23)$$

where K^p_{TT} is an a-priori vertical covariance matrix chosen on climatological grounds. This method presents the same shortcomings as do extrapolation algorithms for discrete signals (Jain and Rangantath, 1981, Sabri and Steenaart, 1978). They are ill behaved to measurement noise, which translates into temperature profiles being physically non feasible because of erroneous lapse rates. It further requires that the matrix K^p_{TT} be non diagonal with large off-diagonal terms. This type of matrix is characteristic of a very small number of degrees of freedom, and is therefore characteristic of a very specific climatology. Lastly, the vertical covariance matrix derived from these extrapolated profiles does not distinguish between profiles derived from different numbers of recorded pressure levels.

The solution to this problem is in the separation of the computation of the 33-times-33 covariance matrix into a series of evaluations of smaller covariance matrices.

Let us define a class C of profile as a collection of profiles sampled at the same pressure levels. For each class, the computation of the covariance matrix is straightforward. Namely

$$\begin{aligned} K^p_{TTC}(i,j) &= < T(i)T(j) > \\ &= \frac{1}{N_C}\sum_{k=1}^{N_C} T_k(i)T_k(j) - \frac{1}{N_C}\sum_{k=1}^{N_C} T_k(i)\frac{1}{N_C}\sum_{k=1}^{N_C} T_k(j) \end{aligned} \qquad (IV.3.24)$$

for i,j belonging to C. $T_k(i)$ denotes the k^{th} profile of the class C sampled at the i_{th} pressure level.

The computed product has the same expectation from class to class. That is, provided that the pressure levels i and j belong to both C1 and C2

$$< K_{TTC1}(i,j) >=< K_{TTC2}(i,j) > \qquad (IV.3.25)$$

Each of the computed submatrices yields a positive semidefinite matrix whose rank (number of non-zero eigenvalues) is bounded above by the number of pressure levels in that class.

One combines then all these submatrices into the estimate of the 33-times-33 vertical covariance matrix. This combination has to be weighted according to the number of degrees of freedom of each class. Assuming that all classes have the same temporal and spatial distributions, the number of degrees of freedom is proportional to the number of profiles used in the computation of the covariance matrix. Thus the estimate for the vertical covariance matrix $K_{TT}(i,j)$ is

$$\hat{K}_{TT}(i,j) = \frac{N_{C1}K_{TTC1}(i,j) + N_{C2}K_{TTC2}(i,j) + \ldots + N_{Cl}K_{TTCl}(i,j)}{N_{C1} + N_{C2} + \ldots + N_{Cl}} \qquad (IV.3.26)$$

where the sum is performed over all classes covering the i^{th} and j^{th} pressure levels. This estimate is an unbiased estimate of the true covariance, since it is an average of unbiased estimates. However, it does not optimally incorporate all measurements of the temperature profiles. An elaborate scheme optimally estimating (from a likelihood point of view) the covariance (and mean) of temperature profiles is presented in details in appendix IV.B. The estimate consists of the sum of all class covariance matrices multiplied by positive semi-definite gain matrices.

Because the estimate for the covariance is a sum of positive semi-definite matrices, it guarantees the resulting matrix is positive semi definite. Testing for the positive definiteness is done by performing the diagonalization of the matrix and checking the eigenvalues for positivity. It can also be checked by computing a series of 33 determinants (Sylvester criterion) using the fact that (Beckenbach and Bellman, 1965, Berg *et al.*, 1984) K^p_{TT} is positive definite if and only if

$$det((K^p_{TT}ij)i,j < k) > 0 \qquad (IV.3.27)$$

for $k = 1,\ldots,33$.

As described in section IV.3.1, all profiles used in this compilation have at least 5 pressure levels recorded. This leaves $23.24/2 = 406$ different classes. Each class requires the running accumulation of 33 terms for the mean and 561 for the cross products (using the symmetry of the covariance matrix). These calculations require 940 kbytes of memory on a typical 16-bit machine. This is larger than the memory capability of the computer used in this reduction (a Data General Nova 4X). To accommodate the computation to this reduced space, the number of classes has to be reduced.

The number of classes is reduced by extrapolating the temperature profiles but, contrary to the ill-conditioned signal theory oriented method described above, this extrapolation is physical in nature and follows the technique used by the National Meteorological Center: below the lowest recorded pressure level, half the adiabatic lapse rate (constant potential temperature) is assumed to extrapolate the profiles down to 1000 *mbar*. Namely, denoting by γ the ratio of specific heat at constant pressure C_p to the specific heat at constant volume C_V (γ values typically 5/3), the extrapolation equation is

$$p^{\frac{1-\gamma}{2\gamma}} T = constant \qquad (IV.3.28)$$

Above the highest recorded pressure level, the profile is not extrapolated. This adiabatic extrapolation reduces the number of classes to 28, which can be accommodated by the computer. This extrapolation tends to increase the vertical correlation in the lowest troposphere for datasets where the sea-level pressure is absent. This is the case for the Continental United States. This fact should be kept in mind when examining the rms error curves of chapter VI.

IV.4 Fitted parameters

Table IV.4.1 present the fitted associated with the four basic climatologies of this study, winter and summer over the United States and Europe. Figures IV.4.1,

	decay constant				oscillatory constant			
	(A)	(B)	(C)	(D)	(A)	(B)	(C)	(D)
1000 mbar	1.097	1.212	1.839	1.298	0.079	0.125	0.008	0.849
950 mbar	0.291	0.782	1.397	0.878	0.432	0.429	0.681	1.067
900 mbar	0.440	0.661	1.057	0.940	0.353	0.443	1.000	1.142
850 mbar	0.444	0.730	1.196	0.931	0.426	0.426	1.044	1.070
800 mbar	0.501	0.756	1.380	0.917	0.512	0.492	1.099	1.117
750 mbar	0.518	0.735	1.241	0.927	0.592	0.620	1.046	1.200
700 mbar	0.563	0.789	1.414	0.984	0.642	0.721	0.973	1.284
650 mbar	0.538	0.775	1.285	1.059	0.641	0.786	0.982	1.347
600 mbar	0.540	0.769	1.371	1.158	0.643	0.826	0.931	1.382
550 mbar	0.596	0.780	1.506	1.279	0.622	0.882	0.863	1.351
500 mbar	0.646	0.871	1.591	1.611	0.585	0.903	0.850	1.266
450 mbar	0.684	0.807	1.935	1.477	0.584	0.858	0.307	1.274
400 mbar	0.762	0.801	2.010	1.755	0.569	0.841	0.022	1.247
350 mbar	0.718	0.573	1.782	1.585	0.591	0.846	0.054	1.323
300 mbar	0.672	0.566	1.817	2.620	0.645	0.849	0.017	0.124
250 mbar	1.115	1.545	1.715	2.350	0.595	0.802	0.001	0.002
200 mbar	0.750	0.883	2.012	1.599	0.003	0.192	0.008	0.003
175 mbar	0.587	0.662	2.255	1.546	0.005	0.367	0.006	0.753
150 mbar	0.515	0.596	1.812	1.009	0.001	0.313	1.178	1.248
125 mbar	0.520	0.455	1.405	0.867	0.010	0.112	1.203	1.325
100 mbar	0.499	0.342	1.811	0.801	0.006	0.047	0.459	1.382
80 mbar	0.402	0.287	1.911	1.060	0.013	0.064	0.003	1.333
70 mbar	0.358	0.283	2.090	2.244	0.011	0.061	0.004	0.415
60 mbar	0.321	0.293	2.317	2.843	0.061	0.034	0.004	0.001
50 mbar	0.417	0.298	2.181	3.429	0.044	0.089	0.002	0.003
40 mbar	0.563	0.283	1.962	4.331	0.078	0.041	0.003	0.004
30 mbar	0.317	0.391	1.105	2.518	0.368	0.072	0.002	0.016
25 mbar	0.226	0.485	0.651	1.648	0.412	0.005	0.005	0.005
20 mbar	0.202	0.810	0.454	1.367	0.461	0.002	0.003	0.004
15 mbar	0.381	1.511	0.419	1.419	0.297	0.005	0.002	0.005
10 mbar	0.816	2.141	0.420	3.016	0.001	0.006	0.003	0.008
7 mbar	2.657	5.300	5.419	3.931	0.005	0.061	0.004	0.005
5 mbar	8.126	8.192	6.942	7.185	2.747	0.012	4.623	14.601

(A) winter USA (B) winter Europe (C) summer USA (D) summer Europe

Table IV.4.1: α and ω associated with summer and winter for Europe and the United States.

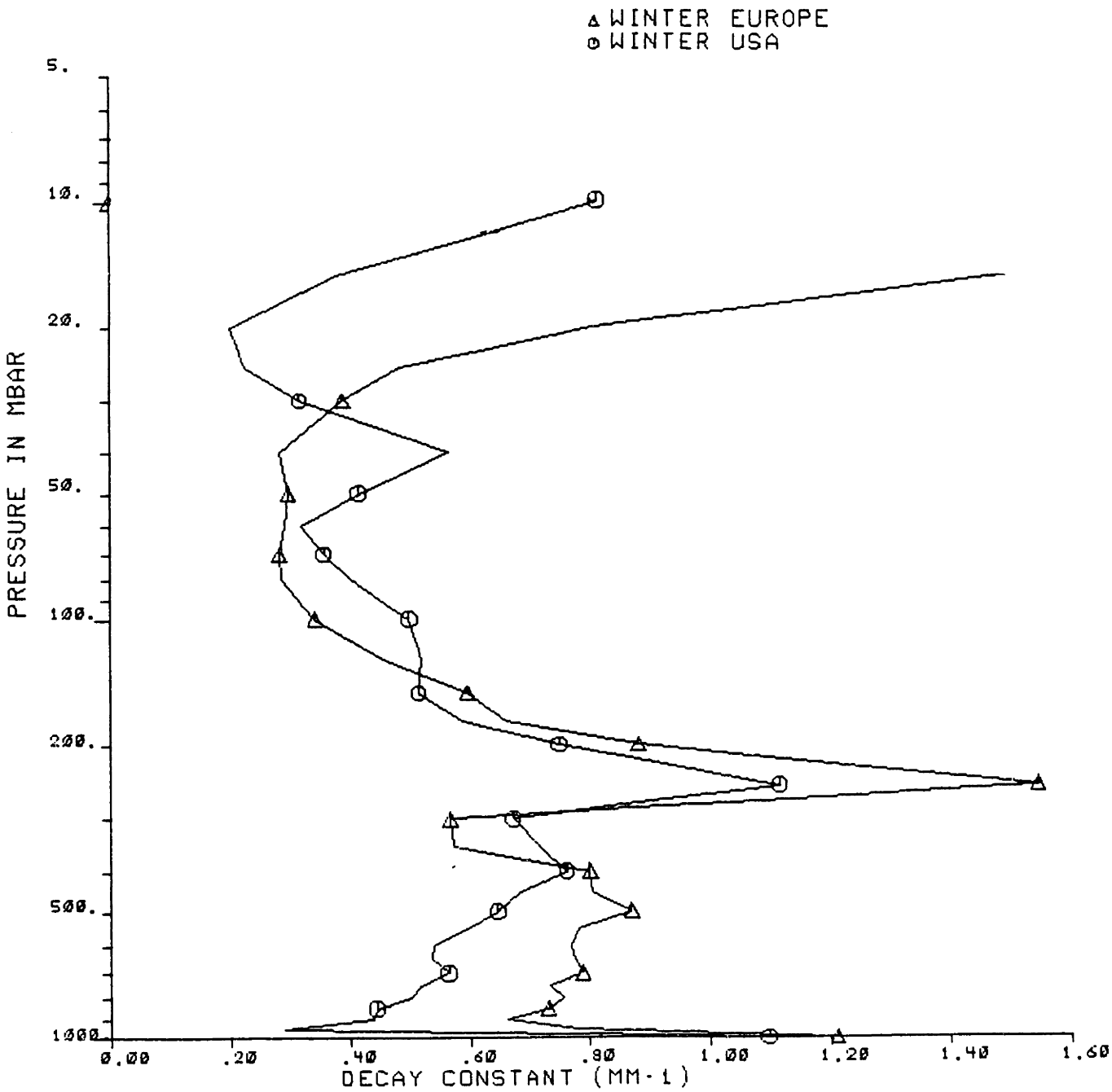


Figure IV.4.1: Winter decay constant over Europe and the USA

IV.4.2, and IV.4.3 plot these parameters as a function of season and location. As seen in Figure IV.4.1, the decay constant is minimum at the tropopause and close to the ground. For p less than 20 *mbar*, the increase in the fitted value of α comes from

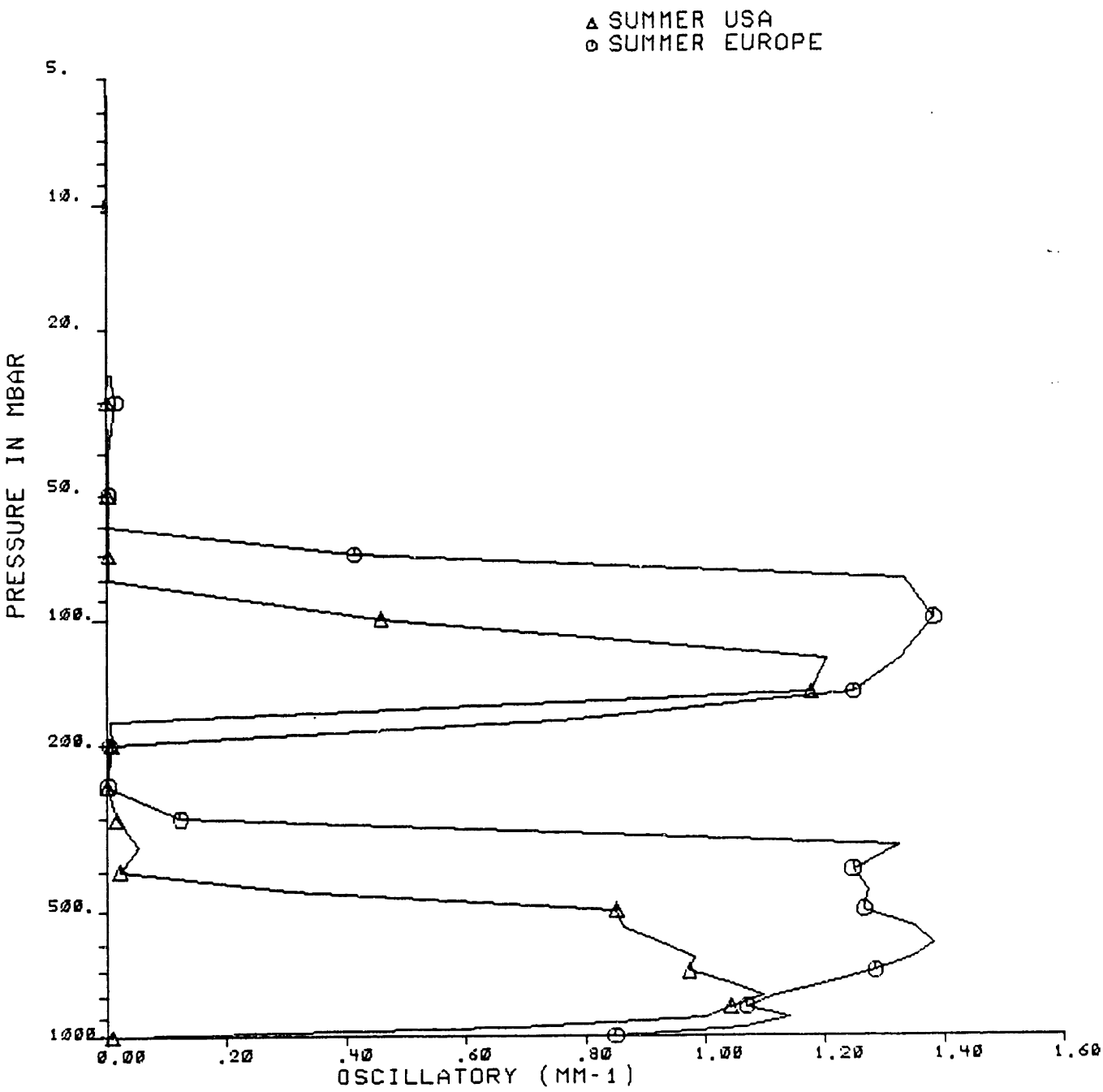


Figure IV.4.2: Summer oscillatory constant over Europe and the USA

the unreliable statistics. Retrieval operators derived using such statistics would not benefit from any horizontal correlation. In future chapters, both α and ω were set to their values at $p = 30mbar$ between 30 and 10 *mbar* in order to extend

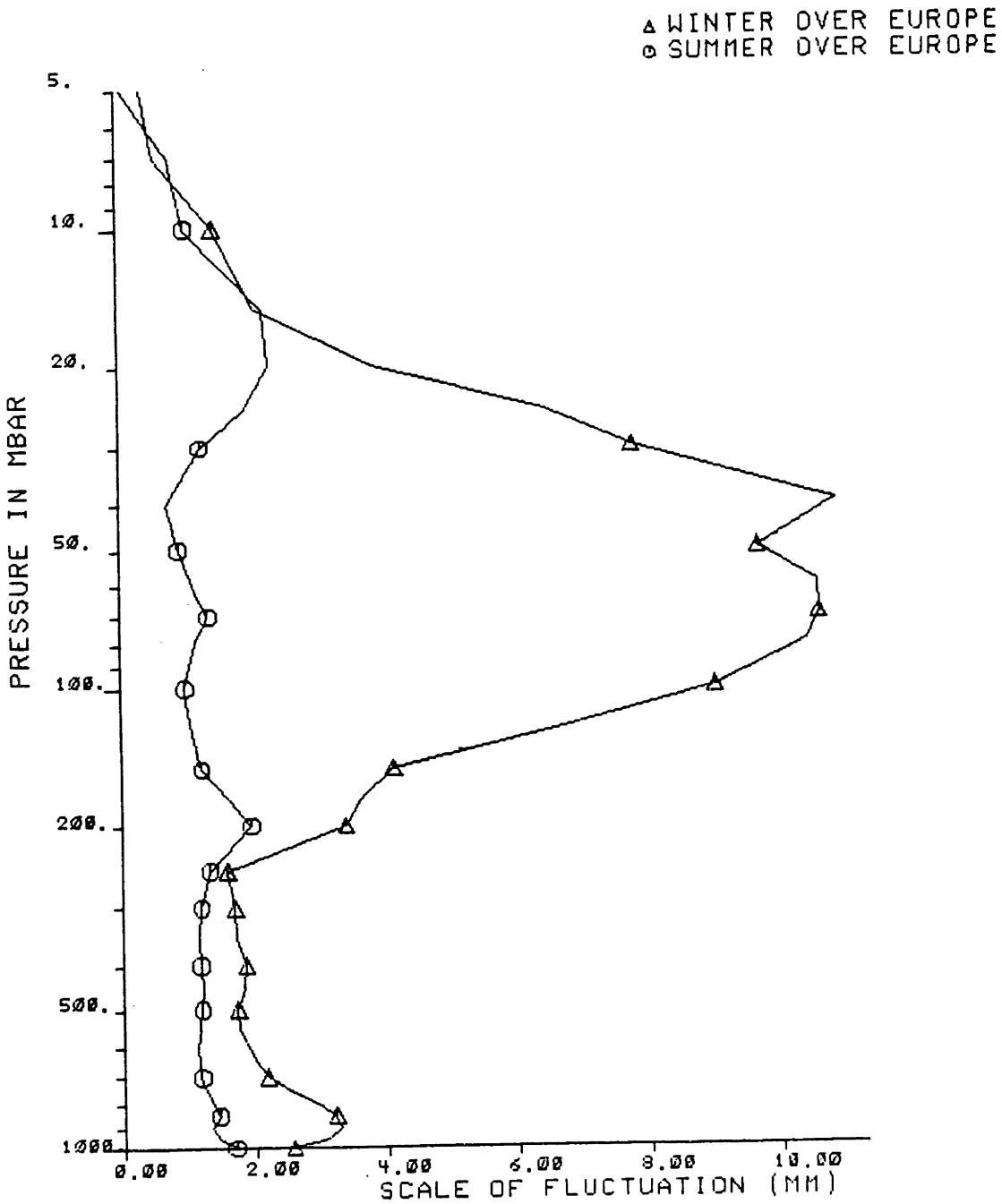


Figure IV.4.3: Scale of fluctuation over Europe during the winter and summer

the amount of horizontal correlation. The 5 *mbar* pressure level isobaric poles was left unchanged. Note that the horizontal correlation is stronger in the stratosphere than in the troposphere. Hence, based on intuition and the values of $\alpha(p)$, the

improvement for multidimensional operators should be larger in the stratosphere than in the troposphere. Also note the large decay constant at the ground level, this is caused by both missing data (elevated stations do no have $p = 1000mbar$), and by the effect of the topography which reduces correlation.

The impact of the tropopause is also visible when examining the vertical variations of ω. Whenever the horizontal correlation is small, the estimate for ω is small since the negative overshoot of that correlation is minimum.

The scale of fluctuation plotted in Figure IV.4.3 greatly differs from the troposphere to the stratosphere during the winter. This larger fluctuation length translates into a stronger horizontal correlation, hence, all things considered, into more improvement from the inclusion of horizontal correlation. However, these results are intuitive. Chapter V will test this intuition. Once again, the impact of missing data at high pressure level is perceived by a reduction of the scale of fluctuation above $p20mbar$.

IV.5 Validation

Given that the isobaric and crossbaric correlations look like the observed data and the works previously reported, and given the difficulty in obtaining reliable estimates for the correlation structure contraction mapping, a simple (but adequate) model validation technique is considered. This validation is based on the χ^2 test on the cross-covariance terms which are not considered when fitting the model but can still be compared to the value predicted by the model.

Because of the nature of the three-dimensional model, for a given dataset, the model matches the observed covariance perfectly. As the distance between points increases, the impact of the simple filtering definition of chapter III assumes more and more importance. Moreover, as the distance between point increases, so does the influence of orography and geographical trends. Thus, the domain of validity

of the model will be a pillbox whose radius will depend on the given dataset.

The purpose of the validation is to obtain an estimate of the pillbox radius. Consider a pillbox of radius R. This pillbox contains P possible bins of data, N of which with measurements (some baselines may not be sampled). The χ^2 statistics associated with the N bins of data is (Carnahan *et al.*, 1969)

$$\chi = \sum_{i=1}^{N} \left(\frac{1}{\delta_i} (K(\sigma_i) - \hat{K}(\sigma_i) \right) \qquad (IV.5.1)$$

where δ_i is the standard deviation associated with the measured covariance or cross-covariance. This unusual notation for the standard deviation of the samples cross-covariance is warranted by the usage of σ to denote isobaric displacements (sorry!).

δ_i is a function of the baseline sets, the temporal interval between measurements of the covariance, the bin width, the average velocity and the isobaric covariance kernel. The exact expression for δ_i (assuming a frozen atmosphere model as given by Appendix IV.A.)

To avoid dealing with tabulated probability distributions, the reduced χ^2 was considered. The reduced distribution is (Bevington, 1969)

$$\chi^2_\nu = \chi^2 / \nu \qquad (IV.4.2)$$

where $\nu = N - n$ is the number of degrees of freedom left after fitting N to a function with n parameters. A value for χ^2_ν equal to one or less correspond to adequate fit. Values greater than 1.5-2 indicates that the fit is no more adequate. With the present model, for each cross-covariance, five parameters are involved (the zero spacing covariance and the two complex poles).

Tables IV.5.1 and IV.5.2 present the reduced χ^2 associated with the cross-covariance for the 14 NMC pressure levels for isobaric distances up to 1000 km (N=18 bins) and, in the second case, distances up to 2500 km (N=44 bins). These

1000	0.1											
850	0.8	0.6										
700	0.7	0.8	0.2									
500	0.6	0.7	0.8	0.3								
400	0.5	0.6	0.8	0.8	0.6							
300	0.8	0.8	0.6	0.8	0.8	0.6						
250	1.2	1.1	0.8	0.9	0.8	0.8	0.4					
200	1.3	0.9	1.1	1.4	1.2	0.8	0.7	0.6				
150	1.2	0.6	0.7	0.7	0.7	0.7	0.8	0.3	0.6			
100	1.2	0.8	0.7	0.6	0.6	0.7	0.7	0.8	0.5	0.6		
70	1.2	1.2	1.2	0.8	0.9	1.0	0.5	0.8	0.3	0.7	0.6	
50	1.2	1.3	1.1	1.2	1.2	0.9	0.8	0.5	0.6	0.5	1.2	1.6
	1000	850	700	500	400	300	250	200	150	100	70	50

Table V.5.1: Table of χ^2 for the model fitted to the United States for distances up to 1.0 Mm

1000	0.1											
850	0.8	0.6										
700	1.4	0.8	0.4									
500	0.6	1.4	0.8	0.3								
400	1.1	0.9	0.8	0.8	1.2							
300	0.8	0.8	0.6	0.8	0.8	0.6						
250	1.9	1.1	0.8	0.9	0.8	0.8	1.5					
200	2.4	0.9	1.1	1.4	1.9	0.8	1.4	0.9				
150	1.9	0.6	1.4	1.2	1.4	1.4	0.8	0.7	1.1			
100	1.9	0.8	1.4	0.6	0.6	1.4	1.4	0.8	0.5	0.6		
70	1.9	1.9	1.9	0.8	0.9	1.0	0.6	0.8	0.9	1.4	0.6	
50	1.9	1.3	1.1	1.9	1.9	0.9	0.8	0.5	1.6	1.5	2.9	4.5
	1000	850	700	500	400	300	250	200	150	100	70	50

Table V.5.2: Table of χ^2 for the model fitted to the United States for distances up to 2.5 Mm

tables were generated according to the noise variance expressions developed in appendix IV.A and using the continental United States.

This table shows that the model is adequate for a distance of 1000 *km* with

the possible exception of the higher altitudes and of cross-covariance for elements strongly negatively correlated across the tropopause. For a distance of about 2500 km, the χ^2_ν coefficients depart significantly from one. Some the fits still appear good. Hence the distance of about 2500 km seems to be the limit of validity of the model. This distance also corresponds to reasonable limits for the assumption of stationarity of atmospheric temperature fields (Charney, 1973), which explains some of the terms along the diagonal.

IV.6 Applications of the Model (other than Remote Sensing)

The main purpose of the model for the three-dimensional covariance is for use in the design equation for the single-spot and multiple-spot retrieval operators developed in Chapter V and implemented in Chapters VI and VII. However, it can be used for other aspects of atmospheric science and also to provide us with a "feel" for when the inclusion of horizontal correlation will provide a significant improvement and when it will not.

IV.6.1 Covariance representations of temperature fields

The three dimensional model allows us to easily look at two different representations of the three dimensional covariance, the isobaric covariance varying as a function of pressure and the two-level (cross-level) cross-covariance as a function of pressure and isobaric displacement.

Figures IV.6.1 and IV.6.2 present contour plots of isobaric covariance $K_{TT}(\sigma, p, p)$ for summer over the United States and Europe. Figure IV.6.3 present the same quantity for the United States in winter. The means used in the evaluation of the deviations were computed on a monthly basis in order to reduce the impact of the difference in mean values between 1979 and 1981. These time "localized" means do not reduce the isobaric correlation lengths. The impact of the changes of

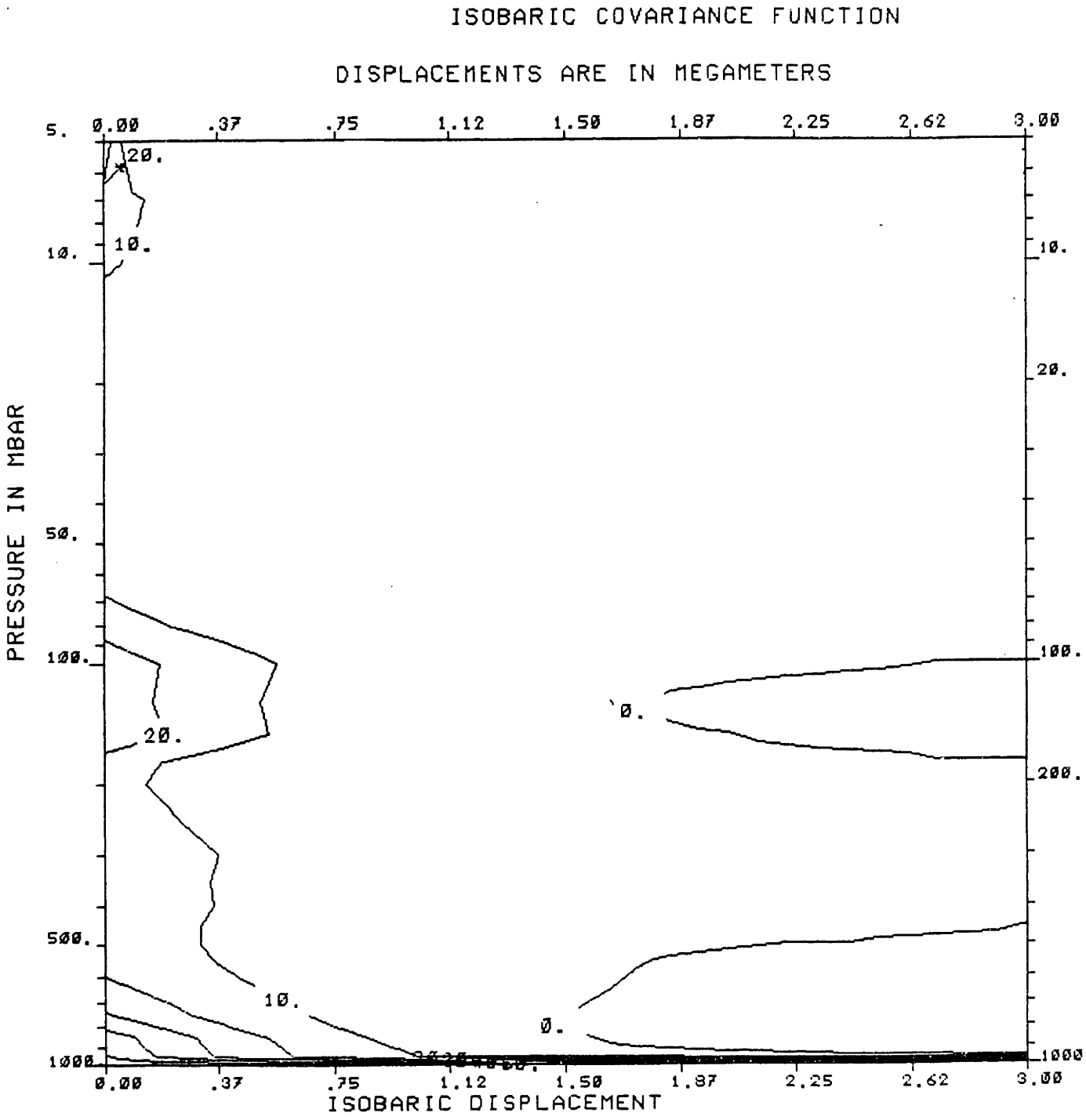


Figure IV.6.1: Isobaric covariance $K_{TT}(\sigma, p, p)$ as a function of isobaric displacement for the 33 pressure levels of this study over the United States during the summer

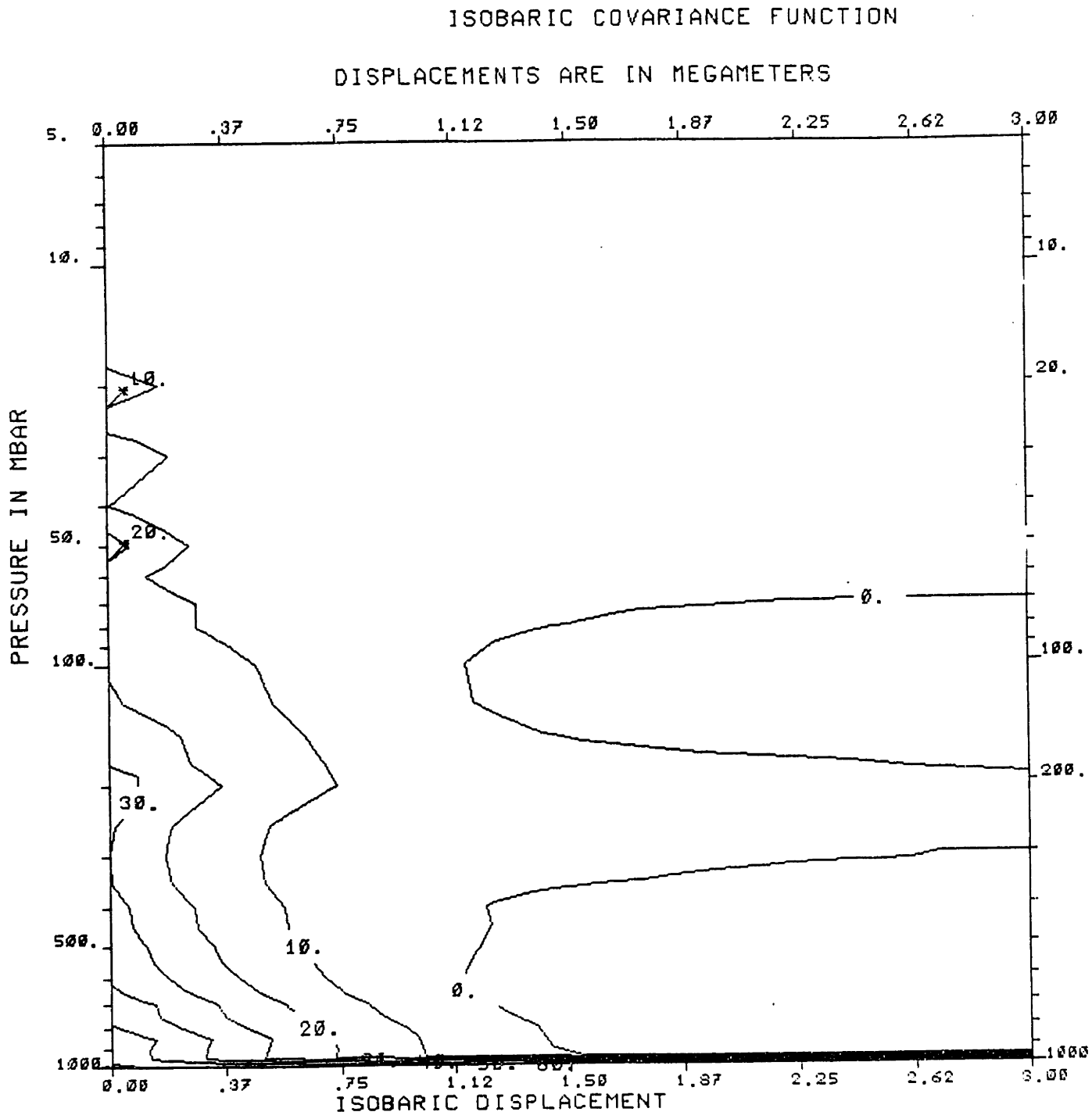


Figure IV.6.2: Isobaric covariance $K_{TT}(\sigma, p, p)$ as a function of isobaric displacement for the 33 pressure levels of this study over Europe during the summer

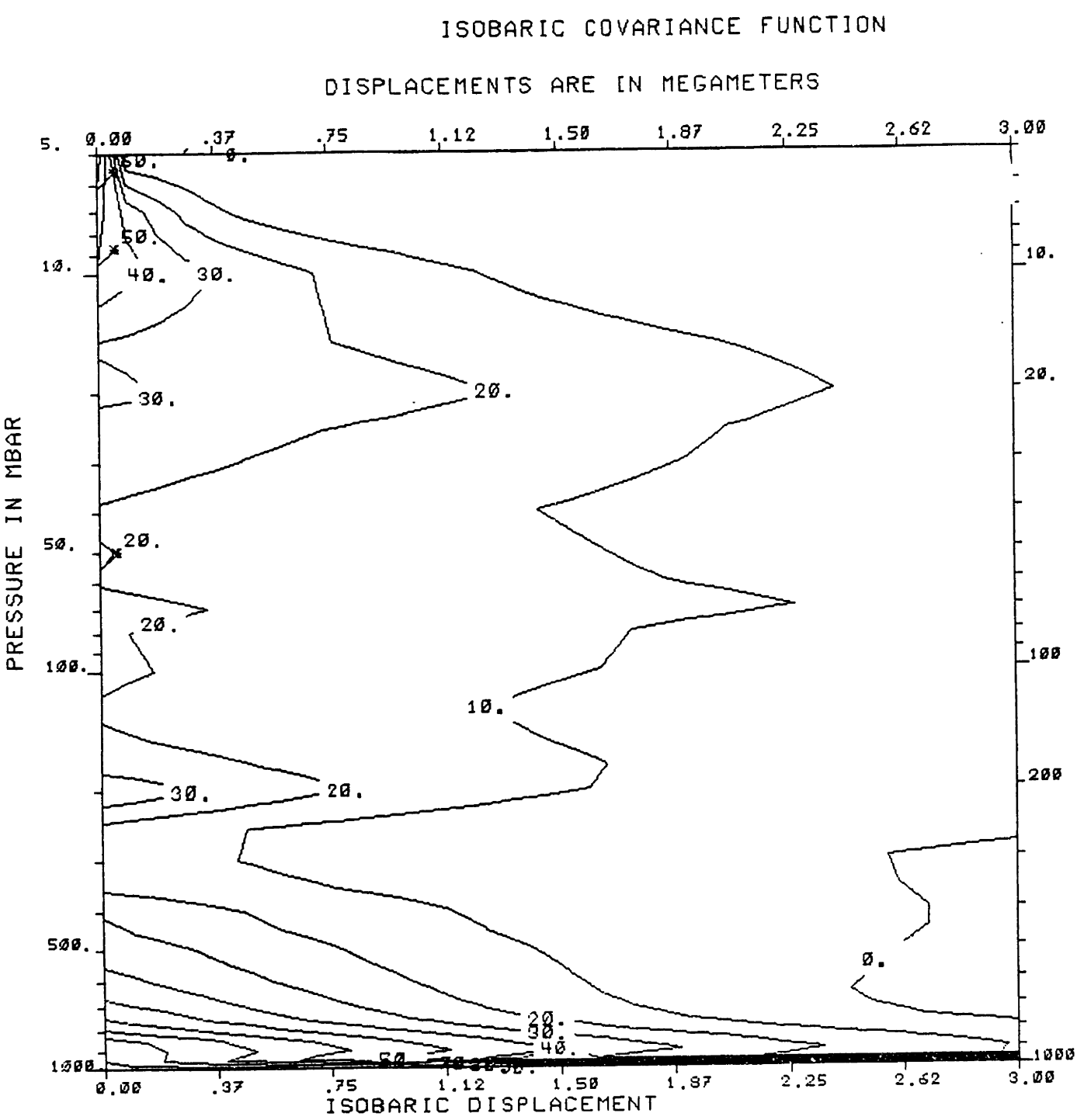


Figure IV.6.3: Isobaric covariance $K_{TT}(\sigma, p, p)$ as a function of isobaric displacement for the 33 pressure levels of this study over the United states during the winter

variance with pressure is seen as a variation in the power at zero displacement. The tropopause is seen by a shorter correlation length for pressure levels around 200-175 during the summer and 250-200 mbar during the winter. Thinking in terms of future remote sensing experiments, the large variance of the temperature close to the ground and above 20 mbar for the winter case will result in larger retrieval errors (all things considered) than in the troposphere. The European summer has less variance than the American summer (compare the power distribution close to the ground). The difference between the two summers is a reflection thay the European network is more compact that the US network; thus the stationarity assumption is more valid for Europe than for the US, hence the reduced covariance.

The normalized correlation $K_{TT}(\sigma, p, p)/K_{TT}(0, p, p)$ corresponding to the winter cases are plotted in Figures IV.6.4 and IV.6.5. This representation is equivalent to setting all the temperature variances equal. The longest correlation in the stratosphere reflects the already discussed increased correlation at high altitudes where the effects of the topography are diminishing and are replaced by the more steady large scale flow of the atmosphere. This strong correlation should yield large improvements when using multidimensional retrieval operators.

The two level correlation normalized to a specific pressure level and zero displacement is presented in Figures IV.6.6 to IV.6.8. The model has been fitted to measurements taken over the United States during February 1981. Three different reference pressure levels are exhibited: 800, 500, and 100 *mbar*.

A quantity directly related associated with the variance field is the structure field. The structure function is defined as (Tatarsky, 1961) as the expected value of the square difference between the temperature at two different locations. Namely

$$D_{TT}(\vec{r}, \vec{r}') = < (T(\vec{r}) - T(\vec{r}'))^2 > \qquad (IV.6.1)$$

The structure function presents the advantage over the covariance function that in order to be stationary, that is a function of the displacement $\vec{r} - \vec{r}'$, it requires the

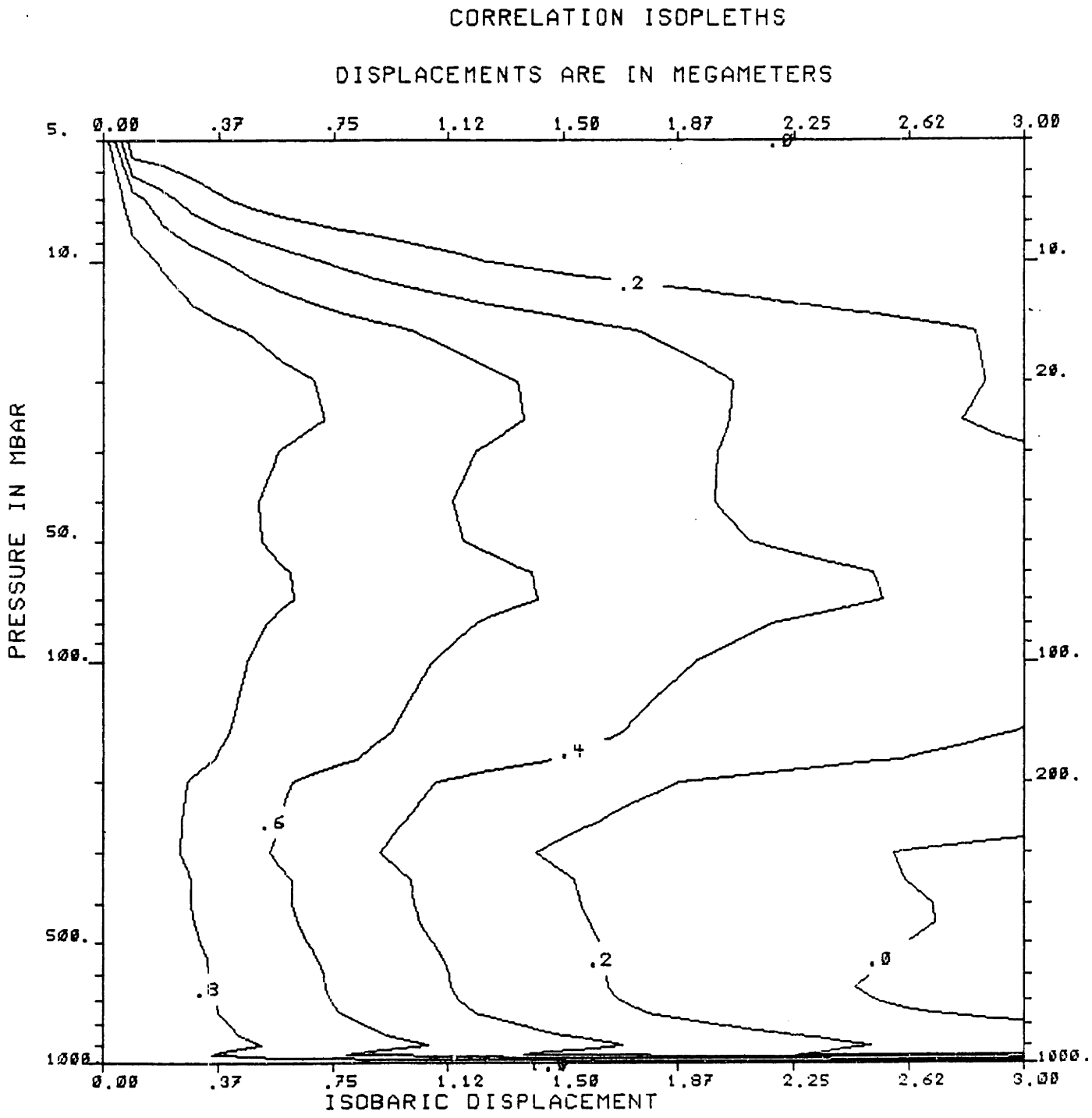


Figure IV.6.4: Normalized isobaric covariance $K_{TT}(\sigma, p, p)$ / $K_{TT}(0, p, p)$ as a function of isobaric displacement for the 33 pressure levels of this study over the United States during the winter

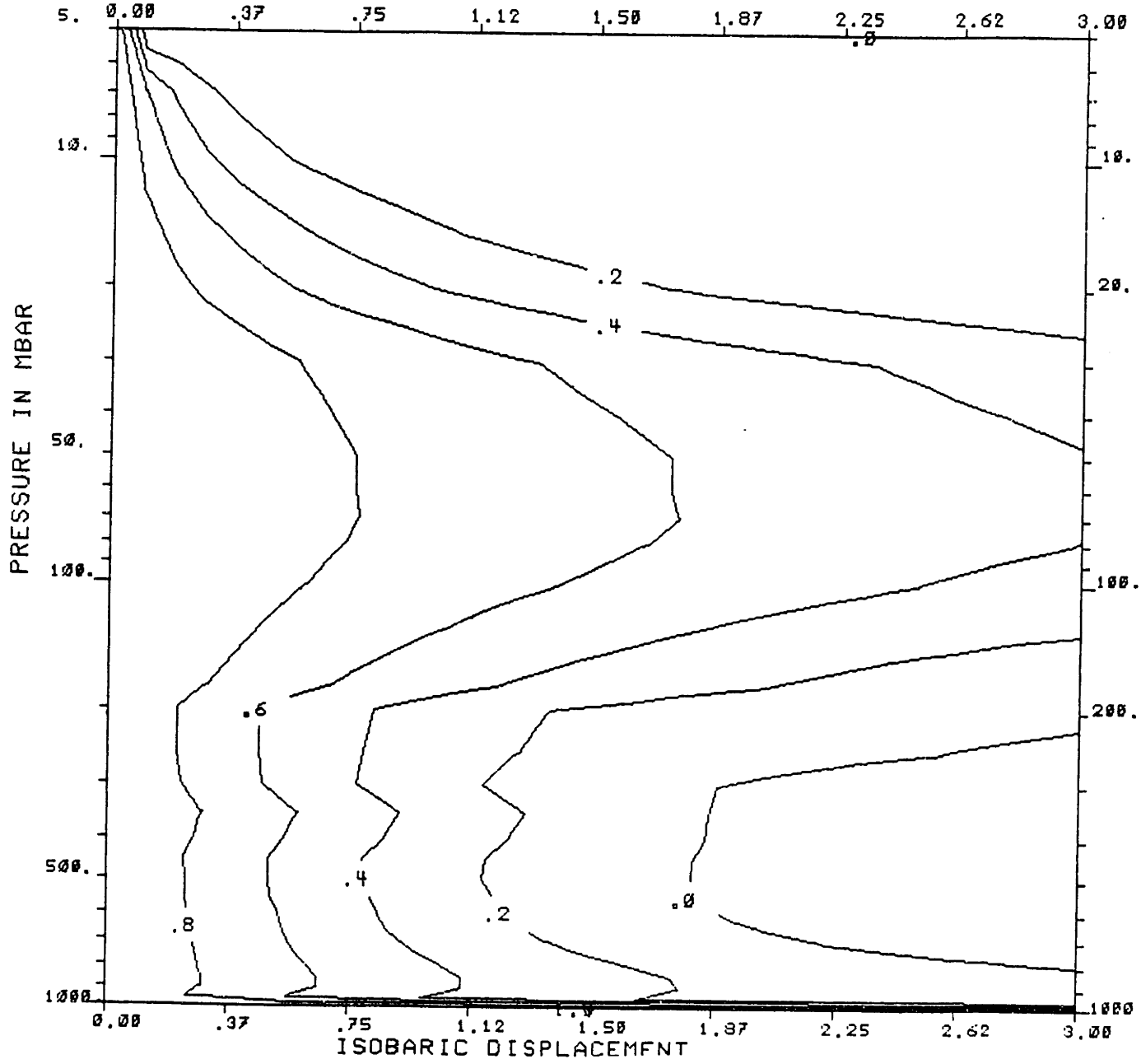


Figure IV.6.5: Normalized isobaric covariance $K_{TT}(\sigma, p, p)$ / $K_{TT}(0, p, p)$ as a function of isobaric displacement for the 33 pressure levels of this study over Europe during the winter

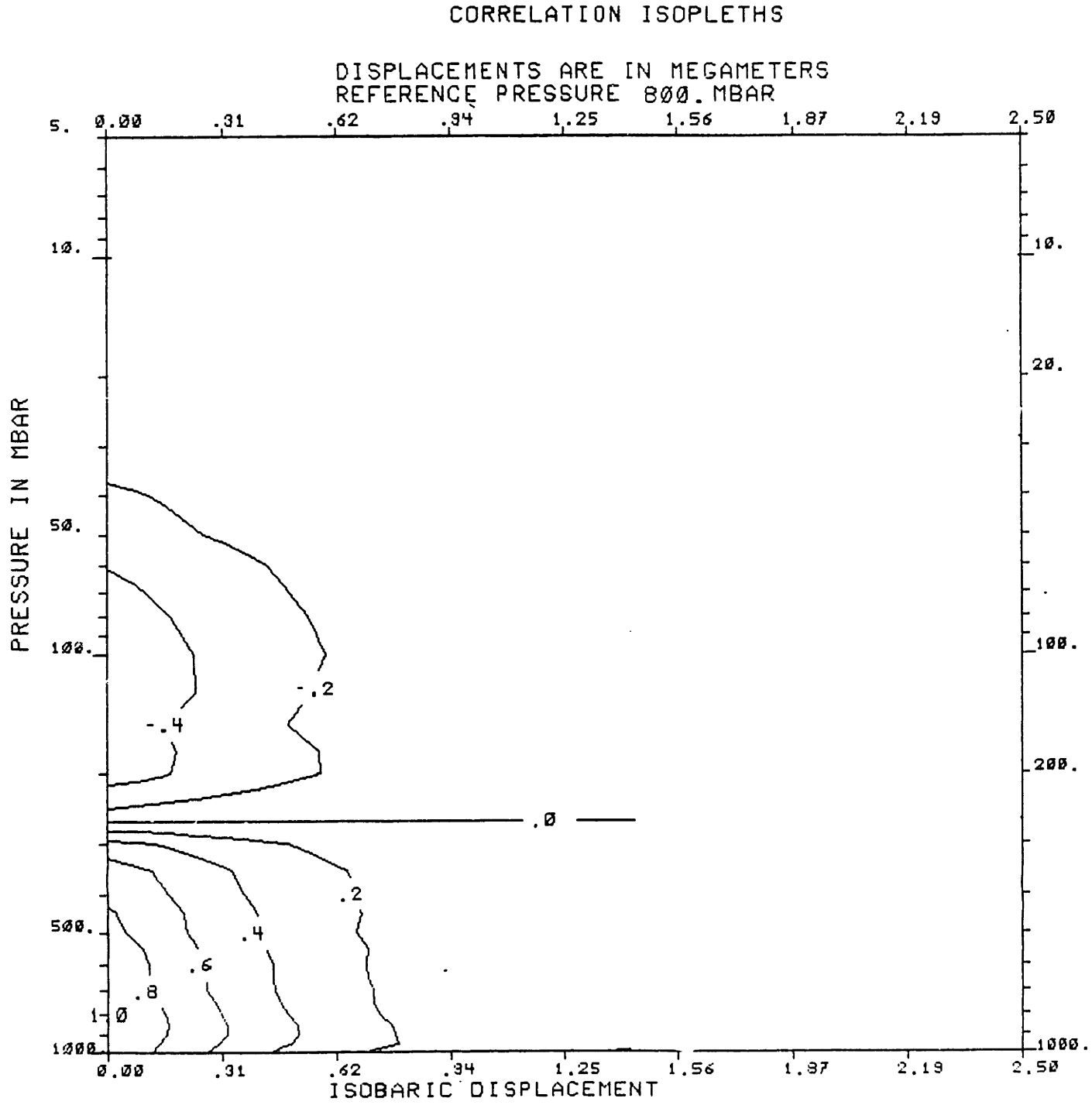


Figure IV.6.6: Two level correlation normalized to 800 mbar over the United States during February 1981

temperature field to be homogeneous at scales smaller than the distance between

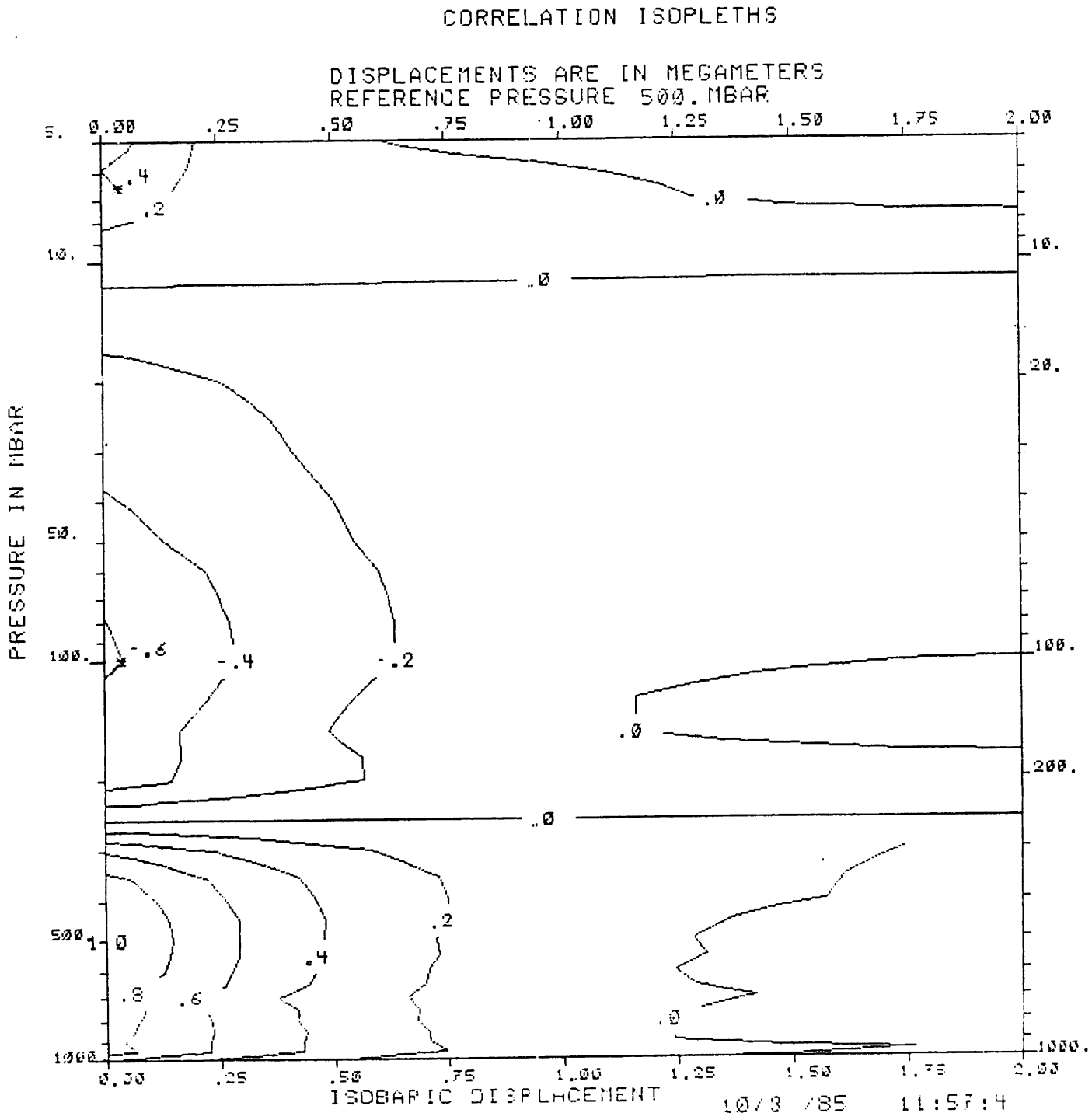


Figure IV.6.7: Two level correlation normalized to 500 mbar over the United States during February 1981

the two elements considered. This is true because the structure function is a high-

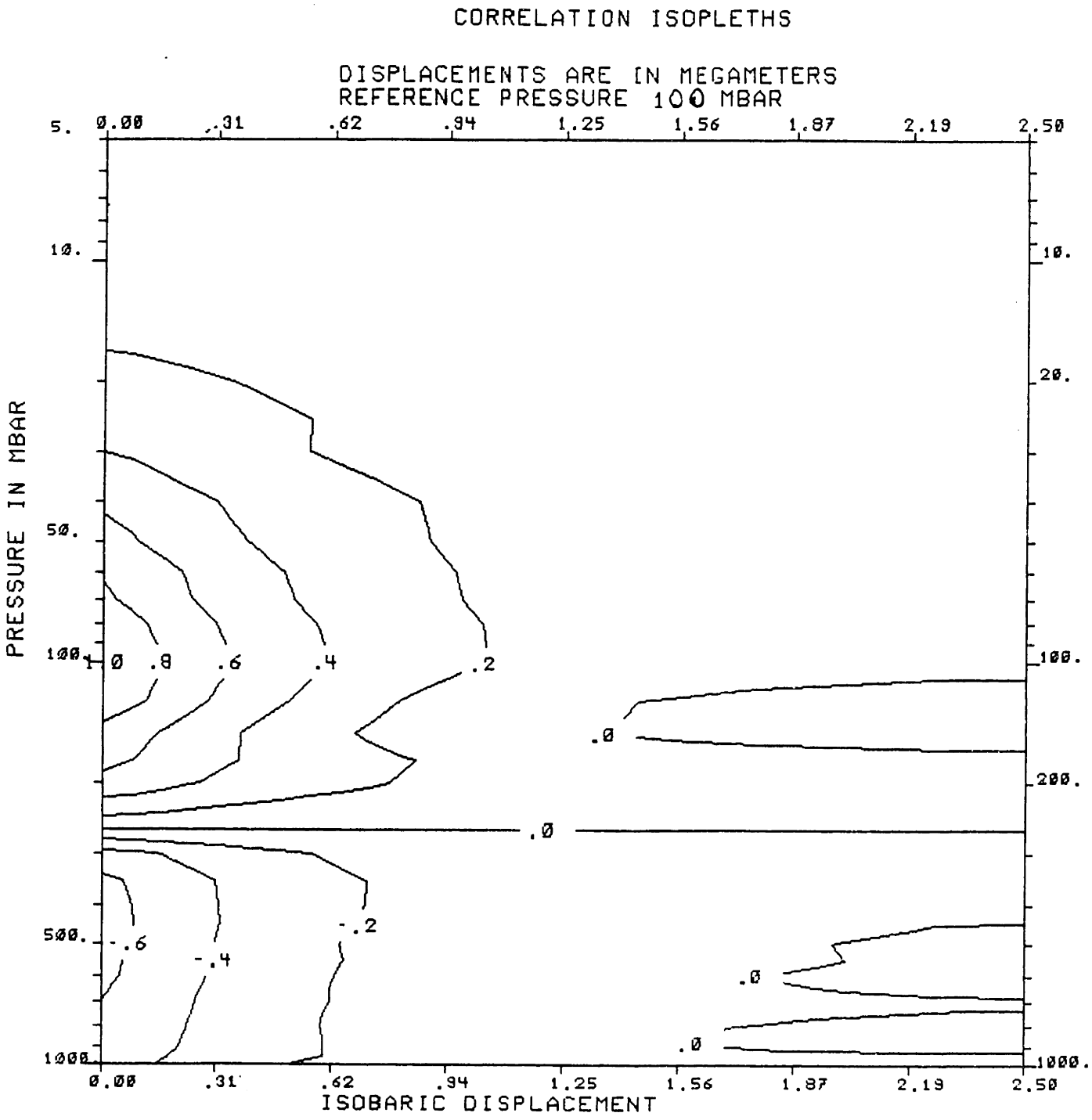


Figure IV.6.8: Two level correlation normalized to 100 mbar over the United States during February 1981

passed version of the traditional covariance. Low frequency terms which are rejected

correspond to large scale structures which create non stationarities. By rejecting these terms, the structure becomes less sensitive to non stationarities than the covariance function is. The structure can be readily evaluated.

$$D_{TT}(\vec{r},\vec{r}') = Var(T(\vec{r})) + Var(T(\vec{r}')) - 2Cov(T(\vec{r})T(\vec{r}')) \qquad (IV.6.2)$$

where Var denotes the variance of a variable and Cov the crosscovariance between two variables. Taking the two points σ Mm away in the isobaric direction, one at a pressure p and the other p' yields

$$D_{TT}(\sigma,p,p') = K_{TT}(0,p,p) + K_{TT}(\sigma,p',p') - 2K_{TT}(\sigma,p,p') \qquad (IV.6.3)$$

Figure IV.6.9 presents the structure function associated with the covariance of Figure IV.6.7 representing the United States during february. The tropopause inversion creates a large structure at zero isobaric displacement.

IV.6.2 Spectral representations of temperature fields

Taking the Fourier transform of the covariance kernel with respect to the isobaric displacement yields a spectral representation characteristic of two dimensional or nearly two dimensional flow. Figure IV.6.10 presents the isobaric spectrum for the month of February over the continental United States as a function of pressure at the standard 33 pressure levels. The isobaric log spectrum (defined as $10 \log S_{TT}(k_h,p,p)$) is plotted as a function of the horizontal wavenumber k_h.

At any pressure level, a maximum in the spectrum for a non zero wavenumber corresponds to a negative correlation in the isobaric field it models. Since the "knee" in the spectrum at which the spectrum begins is α, the decay constant of the covariance, it follows that the broader the spectrum, the narrower the spatial correlation. This broadening of the spectrum occurs at the tropopause and close to

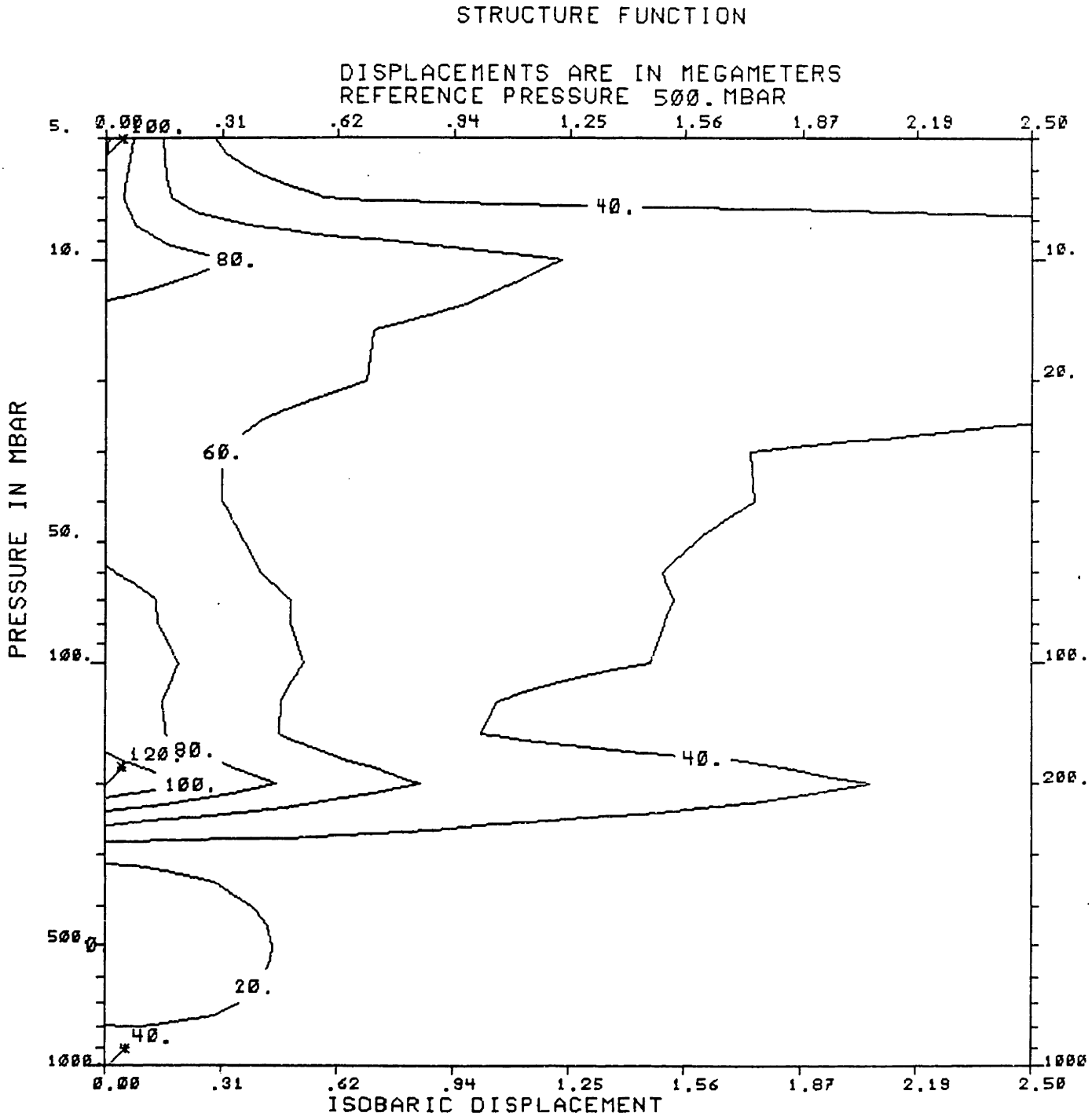


Figure IV.6.9: Structure function over the United States during February. reference pressure is 500 mbar

the ground. At high altitudes (p greater than 50 *mbar*), this broadening is rather

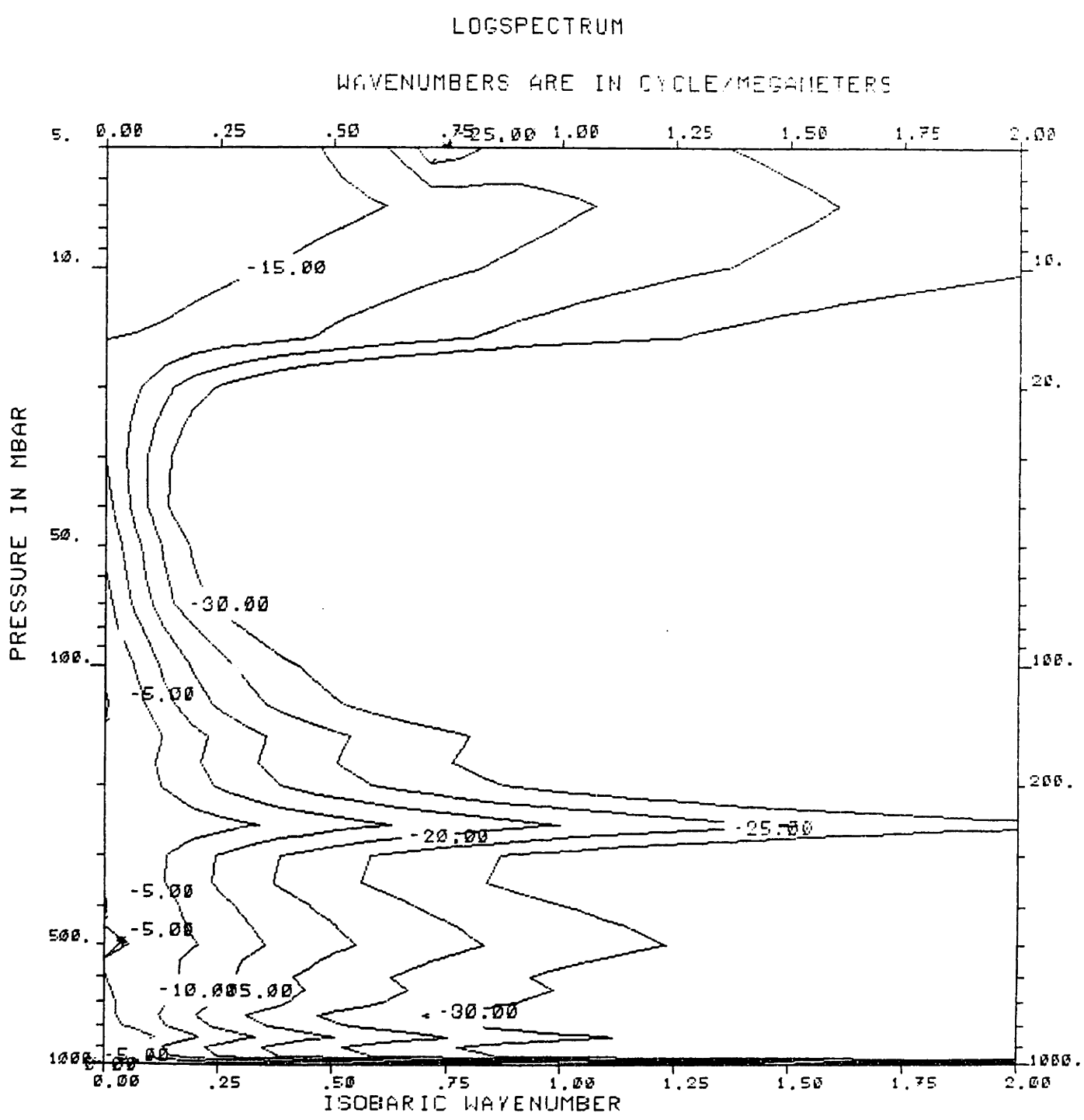


Figure IV.6.10: Isobaric Spectrum over the continental United States during February

a mere reflection of the bad quality of the data used in the fitting of the model.

IV.6.3 Objective analysis

Objective analysis solves the problem of estimating meteorological variables on a regular grid for forecast or climatological purposes from measurements taken at irregular locations by an heterogeneous set of observing sensors (rawinsondes, satellites, rockets, etc..) Optimal objective analysis, as developed by Gandin (1963), solves the problem of finding the coefficients for a linear combination of the observed meteorological parameters that will minimize the mean square error at the point of estimation. The analysis requires the specification of the covariance structure over the region of support of the analysis (Thiebaux 1975).

Objective analysis can be analyzed as static Bayesian estimation of a stochastic process from incomplete observations (Schweppe, 1973, Gelb, 1974). This latter approach will be used at the end of this section to exhibit how the knowledge of the covariance structure and of the radiative transfer equation permits the objective analysis directly using radiances rather than temperature profiles derived from these radiances. This approach can also be used to derive objective analysis schemes restricted in the pressure direction.

Let $\overline{T}_{rk}$ denote the true temperature profile expressed in vector form at the kth location of the M-point analysis grid (usually regular). Let $\breve{\overline{T}}_{oi}$ denote the observed temperature (or estimated in the case of profiles derived from radiances) at the ith sensor of the observation (usually irregular) raster. The observation is corrupted by noise, we denote $\overline{\eta}_i$ the vector of observation noise associated with the ith sensor.

$$\breve{\overline{T}}_{oi} = \overline{T}_{oi} + \overline{\eta}_i \qquad (IV.6.4)$$

. The observation error does not have to be uncorrelated with the temperature (Bergmann, 1979). The interpolation process can be written as

$$\hat{\overline{T}}_{rk} = \sum_{i=0}^{N} h_{ik} \breve{\overline{T}}_{oi} = \sum_{i=0}^{N} h_{ik} (\overline{T}_{oi} + \overline{\eta}_i) \qquad (IV.6.5)$$

where N denotes the number of stations used in the interpolation process. The optimum observation matrices h_{ik} are found by minimizing the mean square error at point k

$$E_k^2 = trace < (\overline{T}_{rk} - \hat{\overline{T}}_{rk})(\overline{T}_{rk}^t - \hat{\overline{T}}_{rk}^t) > \qquad (IV.6.6)$$

where ^{t} denotes the vector transposition and $< . >$ the probability expectation. Rather than multiplying the observation by the observational matrices, one considers only the deviations from an a-priori guess profile obtained from meteorological arguments. Should the temperature field be stationary, this a-priori guess would be its mean value. Denoting by $\tilde{\overline{T}}_{rk}$ and $\tilde{\overline{T}}_{oi}$ the a-priori guesses, the design equation for the observational matrices becomes the minimization of

$$E_k^2 = trace < \left(\overline{T}_{rk} - \tilde{\overline{T}}_{rk} - \sum_{i=0}^{N} h_{ik}(\hat{\overline{T}}_{oi} - \tilde{\overline{T}}_{oi})\right)\left(\overline{T}_{rk} - \tilde{\overline{T}}_{rk} - \sum_{i=0}^{N} h_{ik}(\hat{\overline{T}}_{oi} - \tilde{\overline{T}}_{oi})\right)^t > \qquad (IV.6.7)$$

Setting the partial derivative of the error with respect to the observational matrix to zero yields the set of equations

$$< \overline{T}_{oi}\overline{T}_{rk}^t > + < \overline{T}_{rk}\overline{\eta}_i^t >$$

$$= \sum_{j=1}^{N} h_{ij}\left(< \overline{T}_{oi}\overline{T}_{oj}^t > + < \overline{T}_{oi}\overline{\eta}_i^t > + < \overline{T}_{oj}\overline{\eta}_i^t > + < \overline{\eta}_i\overline{\eta}_j^t >\right) \qquad (IV.6.8)$$

This set of equation takes a more familiar form when the observation noise is uncorrelated with the observations. Namely,

$$< \overline{T}_{oi}\overline{T}_{rk}^t >= \sum_{j=1}^{N} h_{ij}\left(< \overline{T}_{oi}\overline{T}_{oj}^t > + < \overline{\eta}_i\overline{\eta}_j^t >\right) \qquad (IV.6.9)$$

Each element of the outer products of IV.6.7 can be evaluated using the three-dimensional model for the covariance of temperature fields. If an observation becomes uncorrelated with the point to be estimated, the weighting matrix h_{ik} tends to zero, and the estimate of the temperature $\hat{\overline{T}}_{rk}$ tends to the a-priori guess $\tilde{\overline{T}}_{rk}$.

Extensive work has been reported in the objective analysis and teleconnection literature on the impact of the observing locations, of the number of points used in the analysis onto the estimation error, and of the robustness of the observations matrices to changes and errors in the correlation $< \overline{T}_{oi}\overline{T}^t_{rk} >$. The interested reader is referred to these papers (Bergmann, 1979, Baedem *et al.*, 1976, Thiebaux, 1977, Ramanathan *et al.*, 1972, Rutherford, 1972, Schlatter, 1975, Schlatter *et al.*, 1976, Phillips, 1982).

Let us now express the objective analysis as a case of static Bayesian estimation to shed some light on the problem. Let us concatenate all observation vectors into one large observation vector $\overline{T}_o$ defined as $\overline{T}^t_o = (\overline{T}^t_{o1}, \overline{T}^t_{o2}, \ldots, \overline{T}^t_{oN})$. Likewise, let $\overline{T}_r$ denote the vector of temperature profiles at all locations where the estimation must be performed, $\overline{T}^t_r = (\overline{T}^t_{r1}, \overline{T}^t_{r2}, \ldots, \overline{T}^t_{rM})$. Finally, let $\overline{T}_s$ be the vector composed by concatenating $\overline{T}o$ and $\overline{T}_r$. That is, $\overline{T}^t_s = (\overline{T}^t_o, \overline{T}^t_r)$. The analysis problem can be written as the estimation of the vector $\overline{T}_s$ from $\overline{T}_o$ given the observation equation

$$\breve{\overline{T}}_o = (I \quad 0)\begin{pmatrix}\overline{T}_o \\ \overline{T}_s\end{pmatrix}\eta \qquad (IV.6.10)$$

where I is an identity matrix and 0 a null matrix. The matrix $(I \quad 0)$ can be interpreted as an observation matrix Υ which allows us to write

$$\breve{\overline{T}}_o = \Upsilon\begin{pmatrix}\overline{T}_o \\ \overline{T}_s\end{pmatrix} + \eta \qquad (IV.6.11)$$

In the case of observations uncorrelated with the measurements, the optimum estimate for $(\overline{T}^t_o, \overline{T}^t_s)$ is given by

$$\begin{pmatrix}\hat{\overline{T}}_o \\ \hat{\overline{T}}_s\end{pmatrix} = \overline{\overline{R}}\Upsilon^t(\Upsilon\overline{\overline{R}}\Upsilon^t + N_{\eta\eta})^{-1}\breve{\overline{T}}_o \qquad (IV.6.12)$$

where $\overline{\overline{R}}$ is the generalized covariance matrix defined as

$$\overline{\overline{R}} = \begin{pmatrix}< \overline{T}_o\overline{T}^t_o > & < \overline{T}_o\overline{T}^t_s > \\ < \overline{T}_s\overline{T}^t_o > & < \overline{T}_s\overline{T}^t_s >\end{pmatrix} \qquad (IV.6.13)$$

and $N_{\eta\eta}$ is the noise measurement covariance.

The matrix $\overline{\overline{R}}$ can easily be computed using the model. The inclusion of vertical and horizontal correlation will increase the error in estimating the temperature profile at the analysis raster locations.

One can generalize the objective analysis definition to optimally include brightness temperature measurements by noting that nothing requires the observation matrix Υ to be the identity. The only requirement is that Υ be a full rank matrix. This is required in order to be allowed to perform the matrix inversion of Equation (IV.6.17) for any measurement noise covariance matrix. The matrix Υ does not have to be diagonal to carry information about the temperature profiles. Some of the rows of Υ can be changed to represent the radiative transfer equation. Likewise, if one wishes to restrict the input (observation) vectors to specific pressure levels, the matrix Υ can be set to zero for those levels not used in the analysis. Let $\overline{T}_b$ denote the vector composed of all brightness measurements used in the analysis and $\overline{T}_{rad}$ denote the temperature profiles contributing to these radiances. The new observation equation becomes

$$\left(\breve{\overline{T}}_b, \breve{\overline{T}}_o. = (W \quad I \quad 0) \begin{pmatrix} \overline{T}_{rad} \\ \overline{T}_o \\ \overline{T}_s \end{pmatrix} + \right. \qquad \left. \eta \right) \qquad (IV.6.14)$$

which is still of the form (IV.6.16).

By performing the analysis using radiances rather than profiles derived from them, one saves one step in the assimilation process. Furthermore, one improves the accuracy in the estimate of $\overline{T}_s$ and in the estimate of $\overline{T}_{rad}$ by using correlation between profiles measured with different instruments. More of this analysis will be performed in Chapter V and Chapter VI of this dissertation.

The three-dimensional model combined with the static Bayesian analysis allows the combination of different sources of information concerning temperature

profiles. This allows for an optimal (in a mean square sense) assimilation of these data and in improved retrievals. As it will be discussed in Chapter VIII, these ideas can be extended to the analysis of statistics and to the inclusions of constraints on the retrieved or assimilated profiles.

IV.5. Conclusions

In this chapter, methods to estimate the different parameters associated with the model were developed. The most notable is the maximum-likelihood method to optimally use measurements of temperature profiles which present gaps varying in height from profile to profile. The model was submitted to a simple reduced χ^2 test and found to be acceptable from this quantitative point of view for displacements up to 3.0 Mm.

The fitted model was used to plot different representations of the atmospheric temperature second moment.

Appendix IV.A: Variance of the Estimate for the Isobaric Correlation based on a Frozen Atmosphere Model.

This appendix presents the calculation of the variance of the estimate for the isobaric covariance kernel based on the summation of the product of the deviation of the temperatures from their mean values at separated locations. This summation is performed on all the stations in a given geographic region that share the same baseline length and over a finite number of reports, typically one month worth of recordings.

The estimation of the variance is important to obtain the error bars associated with the measurements. This error characterization is important for two reasons: it provides an estimate on the validity of the parameters for the isobaric model and it allows for fits optimally weighted from a mean square point of view.

To perform this calculation, let us model $T(t;\vec{z})$ at a specific pressure level as a zero mean space/time Gaussian random variable varying with geographic location $\vec{z}$ and time index t. The spatio-temporal covariance of this process is defined as

$$K_{TT}(t_1,t_2;\vec{z}_1,\vec{z}_2) =< T(t_1;\vec{z}_1)T(t_2;\vec{z}_2) > \qquad (IV.A.1)$$

where $< . >$ denotes the expectation over all temperature fields specific to a climatology. This covariance is assumed wide sense stationary in the time domain, a fact corroborated by measurements, that is

$$K_{TT}(t_1,t_2;\vec{z}_1,\vec{z}_2) = K_{TT}(|t_1-t_2|;\vec{z}_1,\vec{z}_2) \qquad (IV.A.2)$$

As discussed in chapter III, the isobaric covariance is homogeneous in the spatial domain when the two measurements considered are coincident. Extending this homogeneity to measurements separated by space and time yields

$$K_{TT}(t_1,t_2;\vec{z}_1,\vec{z}_2) = K_{TT}(|t_1-t_2|;|\vec{z}_1-\vec{z}_2|) \qquad (IV.A.3)$$

where the distance $|\vec{z}_1 - \vec{z}_2|$ is to a first good approximation the Euclidian distance between the two points on the Earth sphere.

The covariance can be further simplified by extending the Taylor assumption (Taylor, 1935) to the covariance, hence assuming

$$K_{TT}(t_1, t_2; \vec{z}_1, \vec{z}_2) =< K_{TT}(U|t_1 - t_2| - |\vec{z}_1 - \vec{z}_2|) > \qquad (IV.A.4)$$

where the expectation $< . >$ is taken over all wind magnitudes U. If one assumes that the expectation over all temperature fields is independent of the expectation over all wind patterns, expectations can be commuted and

$$K_{TT}(t_1, t_2; \vec{z}_1, \vec{z}_2) = K_{TT}(< U > |t_1 - t_2| + |\vec{z}_1 - \vec{z}_2|) \qquad (IV.A.5)$$

This model assumes that in terms of overall spatio-temporal correlation, the atmosphere is isotropic. This conditions, in turn, requires the wind correlation to be isotropic. Such an isotropy has been observed (Rutherford, 1972). Average values of wind magnitudes are readily available in the literature (Oort and Rasmusson, 1971). Figure IV.A.1 presents the variations of the average wind magnitude as a function of height, in the northern hemisphere between the latitudes of 15 and 75 degrees. The magnitude is typically of the order of 5 m/s in the troposphere and stratosphere and peaks up to 10 to 15 m/s in the tropopause. This increase in wind speed will result in a loss of isobaric correlation around the tropopause. This result will have far reaching consequences in chapters VI and VII. A average wind magnitude of 10 m/s corresponds to about $1Mm/day$.

The estimate for the spatial covariance between points $\vec{z}_1$ and $\vec{z}_2$ is given by

$$\tilde{K}_{TT}(0; |\vec{z}_1 - \vec{z}_2|) = \frac{1}{N} \sum_{n=1}^{N} T(\tau n; \vec{z}_1) T(\tau n; \vec{z}_2) \qquad (IV.A.6)$$

where τ is the time interval at which regular measurements of the temperature are performed. For NMC observational data, the minimum value of τ is 12 hours, but

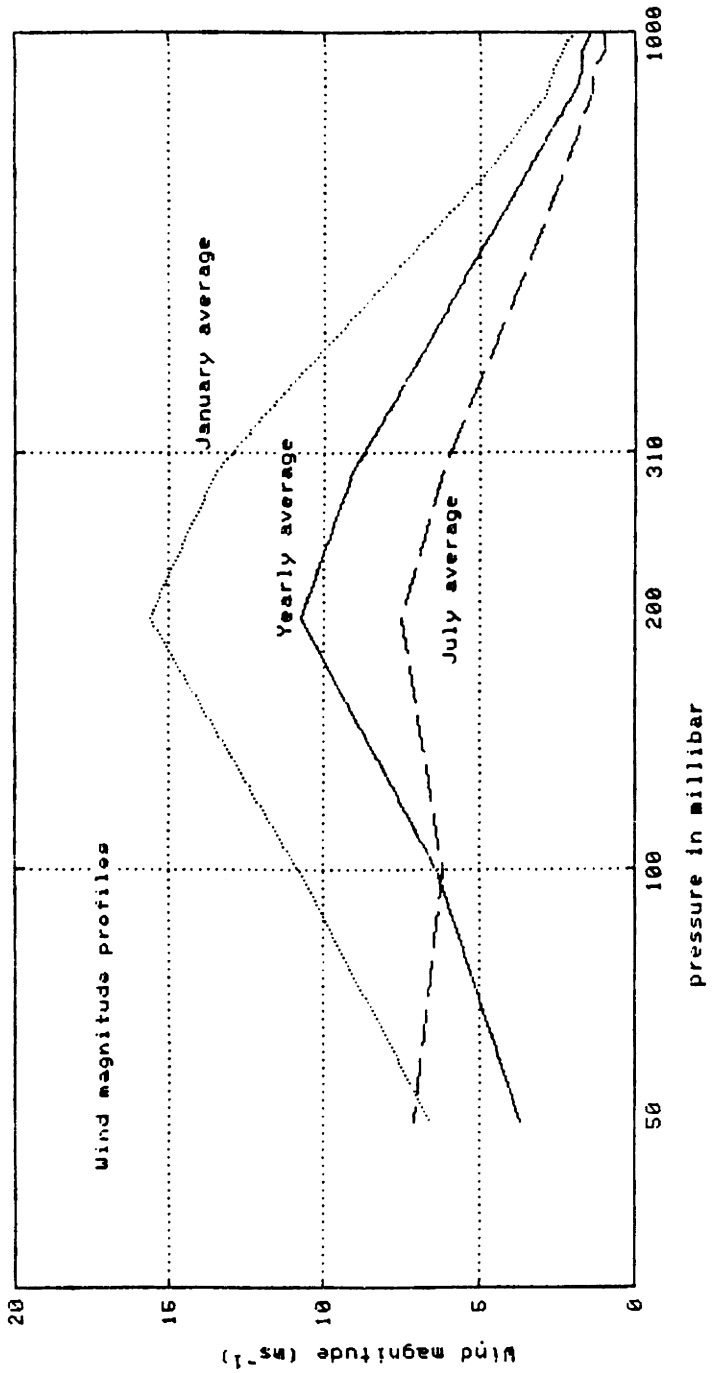


Figure IV.A.1: Variations of the wind magnitude as a function of height

it equals 24 hours for all fits performed in this research. Note that this estimation only takes into account one specific baseline of length $|\vec{z}_1 - \vec{z}_2|$. However, when one fits the homogeneous isobaric covariance kernel, all baselines with the same length (within a bin resolution) must be accounted for to reduce the variance of the

estimate. Moreover, using the symmetry of the isobaric covariance with respect to the origin, each baseline can be counted twice, one time in each direction ($\vec{z}_1 - \vec{z}_2$ and $\vec{z}_2 - \vec{z}_1$). This doubling of the number of pairs of stations used simplifies the calculation (mostly the coding part of it) of the estimate's variance. When the covariances for M such baselines are averaged, the estimate becomes

$$\tilde{K}_{TT}(0;|\vec{z}_1 - \vec{z}_2|) = \frac{1}{NM}\sum_{n=1}^{N}\sum_{m=1}^{M} T(\tau n;\vec{z}_{1m})T(\tau n;\vec{z}_{2m}) \qquad (IV.A.7)$$

The proposed estimate is clearly unbiased. Let us compute its variance $< \Delta\hat{K}^2_{TT}(0;|\vec{z}_{1m} - \vec{z}_{2m}|) >$ using the properties of Gaussian moment factorization.

$$< \hat{K}^2_{TT} >= \frac{1}{N^2M^2}\sum_{n=1}^{N}\sum_{n'=1}^{N}\sum_{m=1}^{M}\sum_{m'=1}^{M} < T(\tau n;\vec{z}_{1m})T(\tau n;\vec{z}_{2m})T(\tau n';\vec{z}_{1m'})T(\tau n';\vec{z}_{2m'}) > \qquad (IV.A.8)$$

After taking the appropriate expectations, this yields

$$\begin{aligned}
<\hat{K}^2_{TT} >&= K^2_{TT}(0;|\vec{z}_1 - \vec{z}_2|) \\
&+\frac{1}{N^2M^2}\sum_{n=1}^{N}\sum_{n'=1}^{N}\sum_{m=1}^{M}\sum_{m'=1}^{M} K_{TT}(\tau|n-n'|;|\vec{z}_{1m} - \vec{z}_{1m'}|)K_{TT}(\tau|n-n'|;|\vec{z}_{2m} - \vec{z}_{2m'}|) \\
&+\frac{1}{NM^2}\sum_{n=1}^{N}\sum_{n'=1}^{N}\sum_{m=1}^{M}\sum_{m'=1}^{M} K_{TT}(\tau|n-n'|;|\vec{z}_{1m} - \vec{z}_{2m'}|)K^2_{TT}(\tau|n-n'|;|\vec{z}_{1m'} - \vec{z}_{2m}|)
\end{aligned} \qquad (IV.A.9)$$

The first term in the RHS is the square of the expected value of the spatial covariance. Hence the second and third sums in the RHS form the proposed estimate's variance.

Expressing the variance in a closed form requires an analytical expression for the spatiotemporal covariance. In chapter III, the spatial covariance was modeled by

$$K_{TT}(0;|\vec{z}_1 - \vec{z}_2|) >= K_{TT}(0;0)\exp(-\alpha|\vec{z}_1 - \vec{z}_2|)\cos(\omega|\vec{z}_1 - \vec{z}_2|) \qquad (IV.A.10)$$

where $\alpha + j\omega$ is a complex pole characterizing the isobaric covariance which varies as a function of pressure. The fitted values of the complex pole (section IV.4) as well as the condition for positive definiteness show that the real part of the complex pole is usually much larger than the imaginary part. Hence one can approximate

$$K_{TT}(\tau|n-n'|;|\vec{z}_{1m}-\vec{z}_{2m'}|) >\approx K_{TT}(0;0)\exp(-\alpha|\vec{z}_{1m}-\vec{z}_{2m'}|-\alpha<U>\tau|n-n'|) \tag{IV.A.11}$$

Note that the Taylor approximation combined with the decaying exponential model yields a separable time/space temperature covariance, which permits us to separate the summation in time and in space.

Using this model, the summation in equation IV.A.11 becomes

$$\begin{aligned} &< \Delta\hat{K}_{TT}^2 >= \\ &\frac{1}{NM^2}\sum_{n=1}^{N}\sum_{n'=1}^{N}\sum_{m=1}^{M}\sum_{m'=1}^{M} K_{TT}^2(0;0)\exp(-\alpha(<U>\tau|n-n'|+|\vec{z}_{1m}-\vec{z}_{1m'}|)) \\ &\qquad \exp(-\alpha(<U>\tau|n-n'|+|\vec{z}_{2m}-\vec{z}_{2m'}|)) \\ &+\frac{1}{NM^2}\sum_{n=1}^{N}\sum_{n'=1}^{N}\sum_{m=1}^{M}\sum_{m'=1}^{M} K_{TT}^2(0;0)\exp(-\alpha(<U>\tau|n-n'|+|\vec{z}_{1m}-\vec{z}_{2m'}|)) \\ &\qquad \exp(<U>\tau|n-n'|+|\vec{z}_{1m'}-\vec{z}_{2m}|)) \end{aligned} \tag{IV.A.12}$$

The sum of decaying exponentials in time is solved in appendix VI.A and is not repeated here. After manipulations, assuming that the measurements are spread over a long period of time, the variance of the estimate can be expressed as

$$< \Delta\hat{K}_{TT}^2(0;|\vec{z}_1-\vec{z}_2|>= \frac{K_{TT}^2(0;0)}{N}\frac{1+\exp(-\alpha 2\tau<U>)}{1-\exp(-\alpha 2\tau<U>)}$$

$$\begin{aligned} &\left(\frac{1}{M^2}\sum_{m=1}^{M}\sum_{m'=1}^{M}\exp(\alpha|\vec{z}_{1m}-\vec{z}_{1m'}|-\alpha|\vec{z}_{2m}-\vec{z}_{2m'}|)\right. \\ &\left.+\frac{1}{M^2}\sum_{m=1}^{M}\sum_{m'=1}^{M}\exp(-\alpha|\vec{z}_{1m}-\vec{z}_{2m'}|-\alpha|\vec{z}_{1m'}-\vec{z}_{2m}|)\right) \end{aligned} \tag{IV.A.13}$$

The summation on the baselines cannot be reduced into a single expression because of the irregular locations of the recording stations. One can however define an equivalent number of stations $M_e(\alpha, \{z_i\})$ associated with a set of stations sharing the same baseline length according to

$$M_e(\alpha, \{z_i\}) = \left(\frac{1}{M^2} \sum_{m=1}^{M} \sum_{m'=1}^{M} \exp(\alpha |\vec{z}_{1m} - \vec{z}_{1m'}| - \alpha |\vec{z}_{2m} - \vec{z}_{2m'}|) \right.$$

$$\left. + \frac{1}{M^2} \sum_{m=1}^{M} \sum_{m'=1}^{M} \exp(-\alpha |\vec{z}_{1m} - \vec{z}_{2m'}| - \alpha |\vec{z}_{1m'} - \vec{z}_{2m}|) \right) \qquad (IV.A.14)$$

Table IV.A.1 presents the equivalent number of stations for specific fitting bins for the American and European networks. These bins vary from 400 km to 2.5 Mm and the isobaric decay constant equals 0.7, 1.0, and 1.4 MM^{-1}. The larger the decay constant, the more the independent measurements, which is as expected. Note that for the same bin and isobaric decay constant, the European and USA network do not provide the same number of measurements.

Given the equivalent number of stations associated with each baseline, the standard deviation of the estimate for the normalized correlation is hence

$$< \Delta \hat{K}_{TT}(0; |\vec{z}_1 - \vec{z}_2|) >= \frac{K_{TT}(0;0)}{\sqrt{N M_e(\alpha, \{z_i\})}} \sqrt{\frac{1 + \exp(-\alpha 2\tau < U >)}{1 - \exp(-\alpha 2\tau < U >)}}$$

$$\sqrt{1 + \exp(-\alpha 2 |\vec{z}_1 - \vec{z}_2|)} \qquad (IV.A.15)$$

This expression allows us to compute the standard deviation associated with a set (time and space) of measurements. If one thinks of the isobaric covariance as a signal function of the distance $\sigma = |\vec{z}_1 - \vec{z}_2|$, the signal to noise ratio associated with the (indirect) measurement of that signal and the proposed estimate is

$$SNR = N M_e \sinh(\alpha\tau < U >) \frac{1}{1 + \exp(2\alpha\sigma)} \qquad (IV.A.16)$$

The SNR improves with the number of sample taken, the spacing between the baselines, the average magnitude of the wind speed and the sampling time. It degrades as the baseline length increases and the correlation tends towards zero.

bin bounds	alpha= 0.7	1.0	1.4
[399- 449]	14 Me	1.9	2.8 4.1
[633- 683]	14 Me	2.2	3.3 4.9
[866- 916]	20 Me	2.7	4.4 7.1
[1099-1149]	16 Me	2.6	4.2 6.9
[1333-1383]	21 Me	2.6	4.3 7.2
[1566-1616]	27 Me	2.8	4.8 8.2
[1799-1849]	28 Me	3.2	5.5 9.2
[2033-2083]	22 Me	3.2	5.4 8.8
[2266-2316]	16 Me	2.8	4.3 6.2
[2499-2549]	19 Me	2.9	4.5 6.8

Continental United States network

bin bounds	alpha= 0.7	1.0	1.4
[399- 449]	26 Me	1.4	2.0 3.1
[633- 683]	37 Me	1.6	2.4 3.8
[866- 916]	30 Me	1.6	2.4 3.9
[1099-1149]	29 Me	1.8	2.9 4.8
[1333-1383]	39 Me	2.1	3.4 5.6
[1566-1616]	27 Me	2.3	3.7 6.0
[1799-1849]	20 Me	2.3	3.6 5.4
[2033-2083]	7 Me	2.4	3.4 4.7
[2266-2316]	8 Me	2.0	2.7 3.6
[2499-2549]	7 Me	2.1	2.8 3.6

Western Europe network

Table IV.A.1: Equivalent number of stations for the American and European network of observing stations for selected values of the isobaric decaying constant.

Appendix IV.B: Maximum Likelihood Estimation of First and Second Order Statistics from Incomplete and Sample Varying Observations.

The vertical kernel of the three-dimensional model for the covariance of temperature fields as well as the derivation of retrieval operators depends on the adequate evaluation of the first and second moments of temperature profiles.

Provided that the representation of temperature profiles by a vector with

33 components is adequate, this evaluation can be performed by compiling several profiles that belong to the same climatology. Unfortunately, these profiles are not perfectly measured. This is true because for locations of the globe with high elevation, the sea level pressure (1015 *mbar*) is absent and because rawinsondes (the main source of information when compiling the temperature statistics) do not measure high altitude temperature once the ballon breaks. Thus this compilation of statistics is to be performed from incomplete, sample varying observations.

The purpose of this appendix is to derive a method that provides an adequate characterization of the statistics of temperature profiles under these observing conditions.

Other problems encountered in engineering can be put in the same framework. A process can be observed through sensors which fail to record the information from time to time. Likewise, an large array of sensors is only partially sampled at each acquisition cycle because of data rate, signal to noise ratio, or time-bandwidth considerations.

The problem to be solved can be stated as follows: Given a set of incomplete linear measurements of a vectorial process and given a desired structure for its mean value and its covariance matrix, what are the best (in a sense yet to be defined) estimates of these moments that corresponds to the measurements?

The theoretical foundation for the estimation procedure is taken from Burg *et al* (1982). That is, take the random process whose characterization is desired to be a Gaussian process and derive a maximum likelihood estimate for the first and second moments.

Let $\{x_i\} i = 1, \ldots, N$ denote the N samples (realizations) of the M dimensional Gaussian random process $x = N(m, R_o)$ where both mean vector m and covariance matrix R_o are unknown. Let $\{z_i\} i = 1, \ldots, N$ denote the set of N incomplete measurements of the process where the observation equation $z_i = H_i x_i$.

The observation matrix H_i varies from sample to sample but is deterministic. The matrix $H_i \in \{H_{c1}, H_{c2}, \ldots, H_{cl}\}$ a set of l possible observation matrices.

In order to simplify the mathematics of the derivation, the matrices H_i are assumed to be full-rank, that is none of the rows can be written as a linear combinaison of the other rows.

In the case of rawinsondes, the observation matrix H_i is full of zeroes except for the pressure levels being sampled.

The estimation problem can be now stated as follows: Given $\{z_i, H_i\}, i = 1, \ldots, N$, what are m, R_o? Let us assume that the samples z_i are independent. The assumption of independence may be critized on the basis that intersample correlation would help to compensate for the incomplete observations. However, it is required to keep the mathematics of the derivation simple and provide an analytically tractable form. For possible extension of this method to non independent samples, the interested reader is referred to the work of Bhat (1974) and Yaakov (1971). The second hypothesis that can be critizised is the absence of observation noise in the measurements. From the point of view of collecting temperature data from radiosondes for the purpose of deriving statistics, this criticism is irrelevant since the radiosonde measurements are the ground truth to compare the estimates against and, as such, are assumed perfect. In other problems, the inclusion of measurement noise is possible and will be introduced at the end of this analysis.

The probability of the set of measurements $\{z_i\}$ is

$$p(z_1, z_2, \ldots, z_N | m, R_o) = p(z_1|m, R_o).p(z_2|m, R_o) \ldots p(z_N|m, R_o) \qquad (IV.B.1)$$

where

$$p(z_i|m, R_o) = \frac{1}{2\pi^{rank H_i/2}} \frac{1}{|H_i R_o H_i^t|} \exp\{-\frac{1}{2}(z_i - H_i m)^t (H_i R_o H_i^t)^{-1} (z_i - H_i m)\} \qquad (IV.B.2)$$

where $|.|$ denotes the determinant and ^{t} denotes the matrix transposition. Instead of maximizing (IV.B.2), one maximizes a strictly monotonic function of (IV.B.2) its natural logarithm. The objective function to maximize with respect to m and R_o is thus

$$\begin{aligned} L(\{z_i, H_i\}) = &- \frac{1}{2} \ln 2\pi \sum_{i=1}^{N} rank H_i - \frac{1}{2} \sum_{i=1}^{N} \ln |H_i R_o H_i^t| \\ &- \frac{1}{2} \sum_{i=1}^{N} trace\{(z_i - H_i m)^t (H_i R_o H_i^t)^{-1} (z_i - H_i m)\} \end{aligned} \qquad (IV.B.3)$$

Before deriving the optimum estimates of the moments, one must derive two matrix theorems. These theorems pertain to the derivatives of $|H_i R_o H_i^t|$ and of $trace\{(H_i R_o H_i^t)^{-1} A\}$ with respect to R_o when the matrix H_i is not square. These derivations are required because these results are not readily available in the litterature.

Let us consider first the singular value decomposition of the observation matrix H (the index i is dropped for the purpose of clarity). The observation matrix can be written as

$$H = \mathcal{H}_l \begin{pmatrix} \Lambda & 0 \end{pmatrix} \mathcal{H}_r{}^t \qquad (IV.B.4)$$

where the matrix $\mathcal{H}_l$ is an orthonormal matrix of size $rank H.rank H$, the matrix $\mathcal{H}_r$ an orthornomal matrix of size $M.M$, and the matrix Λ a diagonal matrix with non zero entries on its main diagonal (a reflection of the fact that the observation matrices are taken to be full-rank). The matrix $\mathcal{H}_r$ projects the vectorial process x along the row space and null space of the observation matrix H. Namely

$$\mathcal{H}_r m = \begin{pmatrix} m_r \\ m_k \end{pmatrix} \qquad (IV.B.5)$$

and

$$\mathcal{H}_r R_o \mathcal{H}_r{}^t = \begin{pmatrix} R_{orr} & R_{ork} \\ R_{okr} & R_{okk} \end{pmatrix} \qquad (IV.B6)$$

where the indexes r and k indicate row and kernel space respectively.

We can now evaluate the determinant $|HR_oH^t|$ using the row space-null space representation. That is,

$$\begin{aligned} |HR_oH^t| &= |\mathcal{H}_l \begin{pmatrix} \Lambda & 0 \end{pmatrix} \mathcal{H}_r R_o \mathcal{H}_r{}^t \begin{pmatrix} \Lambda \\ 0 \end{pmatrix} \mathcal{H}_l{}^t| \\ &= |\mathcal{H}_l \begin{pmatrix} \Lambda & 0 \end{pmatrix} \mathcal{H}_r \begin{pmatrix} R_{orr} & R_{ork} \\ R_{okr} & R_{okk} \end{pmatrix} \mathcal{H}_r{}^t \begin{pmatrix} \Lambda \\ 0 \end{pmatrix} \mathcal{H}_l{}^t| \\ &= |\mathcal{H}_l \Lambda R_{orr} \Lambda \mathcal{H}_l{}^t| \end{aligned} \qquad (IV.B.7)$$

The matrices $\mathcal{H}_l$, Λ, and R_{orr} are all square which allows us to use the Binet Cauchy formula for the determinant of product of matrices (Gantmacher, 1959). Noticing that $\mathcal{H}_l$ is unitary, this yields

$$|HR_oH^t| = |\Lambda|^2 |R_{orr}| \qquad (IV.B.8)$$

Hence the derivative is

$$\frac{d}{dR_o} \ln |HR_oH| = \frac{d}{dR_o} \ln |R_{orr}| \qquad (IV.B.9)$$

This latter derivative is known (Athans and Schweppe, 1965) Thus

$$\frac{d}{dR_o} \ln |HR_oH^t| = \mathcal{H}_r{}^t \begin{pmatrix} R_{orr}^{-1} & 0 \\ 0 & 0 \end{pmatrix} \mathcal{H}_r \qquad (IV.B.10)$$

Let us use the properties of the trace operator to derive the second theorem.

$$\begin{aligned} \mathcal{T} &= trace\{(z - Hm)^t (HR_oH^t)^{-1} (z - Hm)\} \\ &= trace\{(HR_oH^t)^{-1} (z - Hm)(z - Hm)^t\} \end{aligned} \qquad (IV.B.11)$$

Using the SVD decomposition of the observation matrix H yields

$$\begin{aligned} \mathcal{T} &= trace\{(\mathcal{H}_l \Lambda R_{orr} \Lambda \mathcal{H}_l{}^t)^{-1} (z - Hm)(z - Hm)^t\} \\ &= trace\{R_{orr}^{-1} \Lambda^{-1} \mathcal{H}_l{}^t (z - Hm)(z - Hm)^t \mathcal{H}_l \Lambda^{-1}\} \end{aligned} \qquad (IV.B.12)$$

The derivative with respect to the covariance matrix

$$\frac{d}{dR_o} \mathcal{T} = \frac{d}{dR_o} trace\{R_{orr}^{-1} \Lambda^{-1} \mathcal{H}_l{}^t (z - Hm)(z - Hm)^t \mathcal{H}_l \Lambda^{-1}\} \qquad (IV.B.13)$$

The derivative on the RHS has been tabulated (Athans and Schweppe, 1965)

$$\frac{d}{dR_o} = -\mathcal{H}_r{}^t \begin{pmatrix} R_{orr}^{-1}\Lambda^{-1}\mathcal{H}_l{}^t(z-Hm)(z-Hm)^t\mathcal{H}_l\Lambda^{-1}R_{orr}^{-1} & 0 \\ 0 & 0 \end{pmatrix} \mathcal{H}_r \qquad (IV.B.14)$$

The derivative of the trace with respect to the mean can also be found using the SVD decomposition. Let us expand the trace in terms of the mean.

$$\begin{aligned} \mathcal{T} =& trace\{R_{orr}^{-1}\Lambda^{-1}\mathcal{H}_l{}^t(zz^t - Hmz^t - zm^tH^t + Hmm^tH^t)\mathcal{H}_l\Lambda^{-1}\} \\ =& trace\{R_{orr}^{-1}\Lambda^{-1}\mathcal{H}_l{}^tH(mm^t)H^t\mathcal{H}_l\Lambda^{-1}\} - 2trace\{R_{orr}^{-1}\Lambda^{-1}\mathcal{H}_l{}^t(Hmz^t)\mathcal{H}_l\Lambda^{-1}\} \\ &+ trace\{R_{orr}^{-1}\Lambda^{-1}\mathcal{H}_l{}^t(zz^t)\mathcal{H}_l\Lambda^{-1}\} \end{aligned} \qquad (IV.B.15)$$

After substitution

$$\mathcal{T} = trace\{R_{orr}^{-1}m_rm_r^t\} - 2trace\{R_{orr}^{-1}m_rz^t\mathcal{H}_l\Lambda^{-1}\} + trace\{R_{orr}^{-1}\Lambda^{-1}\mathcal{H}_l{}^t(zz^t)\mathcal{H}_l\Lambda^{-1}\} \qquad (IV.B.16)$$

where m_r is, as before, the projection of the mean vector into the row space of the observation matrix H. The vector m_r will therefore be referred to as the row space mean. This yields the following derivative

$$\frac{d}{dm}\mathcal{T} = \mathcal{H}_r{}^t \begin{pmatrix} 2R_{orr}^{-1}m_r \\ 0 \end{pmatrix} - 2\mathcal{H}_r{}^t \begin{pmatrix} R_{orr}^{-1}\Lambda^{-1}\mathcal{H}_l{}^tz \\ 0 \end{pmatrix} \qquad (IV.B.17)$$

These derivatives having being established, it is now possible to derive the optimum moments associated with the measurements provided that they exists. Setting the partial derivatives of the likelihood function $L(\{z_i, H_i\})$ with respect to m and R_o to zero yields (The caret symbol will not be written for this derivation except for the final expression of the design expression for the estimators and when examining the bias of the estimates)

$$\sum_{i=1}^{N} \mathcal{H}_{ri}{}^t \begin{pmatrix} R_{orr}^{-1} & 0 \\ 0 & 0 \end{pmatrix} \mathcal{H}_{ri} =$$

$$\sum_{i=1}^{N} \mathcal{H}_{ri}{}^t \begin{pmatrix} R_{orr}^{-1}\Lambda^{-1}\mathcal{H}_{li}{}^t(z_i - H_im)(z_i - H_im)^t\mathcal{H}_{li}\Lambda^{-1}R_{orr}^{-1} & 0 \\ 0 & 0 \end{pmatrix} \mathcal{H}_{ri} \qquad (IV.B18)$$

and

$$\sum_{i=1}^{N} \mathcal{H}_{ri}{}^{t} \begin{pmatrix} 2R_{orr}^{-1} m_r \\ 0 \end{pmatrix} = 2 \sum_{i=1}^{N} \begin{pmatrix} R_{orr}^{-1} \Lambda^{-1} \mathcal{H}_{li}{}^{t} z_i \\ 0 \end{pmatrix} \qquad (IV.B.19)$$

Let us introduce the concept of observation class to reduce these equations. A class of observation is associated with a set of observation matrices H_i that share the same row space. These matrices have all the same number of rows (since singular weighting matrices are forbidden in this analysis). If two matrices H_1 and H_2 belong to the same class, there is an invertible linear operator $\mathcal{A}_{(1-2)}$ satisfying

$$H_1 = \mathcal{A}_{(1-2)} H_2, \qquad H_2 = \mathcal{A}_{(1-2)}^{-1} H_1 \qquad (IV.B.20)$$

It is possible to group different matrices into a single class because the measurements are assumed noiseless. A class is then characterized by a unique representative matrix H_C. All measurements from that class must be mapped onto that observation matrix using the unique mapping between a matrix of the class and the reference matrix. Using the notation introduced above, the measurement z_2 performed using H_2 is mapping into $\tilde{z}_2$ according to

$$\tilde{z}_2 = \mathcal{A}_{(1-2)} z_2 \qquad (IV.B.21)$$

Let N_c denote the number of samples of the observation z in the class C. The sample mean vector is defined as

$$m_{zc} = \frac{1}{N_c} \sum_{i \in C} z_i \qquad (IV.B.22)$$

where the measurements are equivalenced to the reference matrix of the class. Likewise the sample covariance matrix is

$$R_{zc} = \frac{1}{N_c} \sum_{i \in C} (z_i - m_{zc})(z_i - m_{zc})^t \qquad (IV.B.23)$$

Note that these two moments are defined in terms of the OBSERVATIONS z_i.

The concept of class is important since it allows the combination of mean and covariance of observation of a same process under different observing conditions while reducing the memory storage associated with the estimation. The estimation is performed without requiring each sample value to be used. In some instances, no real measurements are taken and a class's first and second moments are given. This is the case when a priori information about the process is included in the estimation process. If the estimation problems starts with the first two moments rather than with the samples, the number N_c corresponds to a weighting factor associated with a given class.

Plugging (IV.B.22) and (IV.B.23) into (IV.B.18) and (IV.B.19) yields the design equation for the covariance matrix

$$\sum_{c=1}^{l} N_c \mathcal{H}_{rc}{}^t \begin{pmatrix} R_{orr}^{-1} & 0 \\ 0 & 0 \end{pmatrix} \mathcal{H}_{rc} =$$

$$\sum_{c=1}^{l} N_c \mathcal{H}_{rc}{}^t \begin{pmatrix} R_{orr}^{-1}\Lambda^{-1}\mathcal{H}_{lc}{}^t(R_{zc} + (H_c m - m_c)(H_c m - m_c)^t)\mathcal{H}_{lc}\Lambda^{-1}R_{orr}^{-1} & 0 \\ 0 & 0 \end{pmatrix} \mathcal{H}_{rc} \tag{IV.B24}$$

and for the mean

$$\sum_{c=1}^{l} N_c \mathcal{H}_{rc}{}^t \begin{pmatrix} 2R_{orr}^{-1} m_r \\ 0 \end{pmatrix} = \sum_{c=1}^{l} N_c \mathcal{H}_{rc}{}^t \begin{pmatrix} R_{orr}^{-1}\Lambda^{-1}\mathcal{H}_{lc}{}^t m_{zc} \\ 0 \end{pmatrix} \tag{IV.B.25}$$

Let us first examine the equation defining the mean estimate. Let us multiply the RHS and the LHS of the equation on the left by the covariance matrix R_o and let us replace the row space mean m_r as a function of the mean m.

$$\sum_{c=1}^{l} N_c \mathcal{H}_{rc}{}^t \begin{pmatrix} I & 0 \\ R_{okr}R_{orr}^{-1} & 0 \end{pmatrix} \mathcal{H}_{rc} m = \sum_{c=1}^{l} N_c \mathcal{H}_{rc}{}^t \begin{pmatrix} I \\ R_{okr}R_{orr}^{-1} \end{pmatrix} \Lambda^{-1}\mathcal{H}_{lc} m_{zc} \tag{IV.B.26}$$

The matrix $\Gamma(R_o, H_c)$ defined as

$$\begin{aligned} \Gamma(R_o, H_c) =& \mathcal{H}_{rc}{}^t \begin{pmatrix} I & 0 \\ R_{okr}R_{orr}^{-1} & 0 \end{pmatrix} \mathcal{H}_{rc} \\ =& R_o H_c^t (H_c R_o H_c^t)^{-1} H_c \end{aligned} \tag{IV.B.27}$$

will heterofore referred as an extrapolation matrix. The matrix Γ satisfies the equality

$$\Gamma(R_o, H_c)\Gamma(R_1, H_c) = \Gamma(R_o, H_c) \qquad (IV.B.28)$$

and is therefore a projection matrix, its eigenvalues are zero and ones independtly of the matrix R_o, the eigenvectors associated with the eigenvalues equal to one are also independent on the covariance matrix R_o, and it is positive semidefinite regardless of the covariance matrix. The trace of Γ equates the rank of the matrix H_c. Extrapolation matrix as it will soon appear are related to the Bayesian estimation matrix of a vector x_i from its sampled components z_i.

The estimate for the mean is uniquely defined if the sum of extrapolation matrices

$$S = \sum_{c=1}^{l} N_c \Gamma(R_o, H_c) \qquad (IV.B.29)$$

admits an inverse . When this is the case, the mean vector estimate is

$$\begin{aligned} \hat{m} &= \Big(\sum_{c=1}^{l} N_c \Gamma(\hat{R}_o, H_c)\Big)^{-1} \sum_{c=1}^{l} N_c \mathcal{H}_{rc}{}^t \begin{pmatrix} I \\ \hat{R}_{okr}\hat{R}_{orr}^{-1} \end{pmatrix} \Lambda^{-1} \mathcal{H}_{lc} m_{zc} \\ &= \Big(\sum_{c=1}^{l} N_c \Gamma(\hat{R}_o, H_c)\Big)^{-1} \sum_{c=1}^{l} N_c \Gamma(R_o, H_c) H_c^t (H_c H_c^t)^{-1} m_{zc} \end{aligned} \qquad (IV.B.30)$$

The estimate for the mean is a weighted average of the sample means, once they have been rotated and scaled aloong the singular vectors of the observations matrices, extrapolated along the kernel space and finally rotated back in the common natural representation.

The condition for the existence of a unique estimate of the mean requires that all the different pressure levels (or sensors) be at least sampled once to provide information about all components of the vector x. Furthermore, it requires a certain degree of overlapping between the observation matrices. To relate this uniqueness condition to the design of test signals for identification of linear systems, the different

classes can be thought as "modes" of the sampling system. The mean vector m can be identified only if enough modes of the sampling system are present.

Each one of the matrices in (IV.B.29) are positive semidefinite (since the extrapolation matrices are positive definite), their sum is positive semidefinite. Positive definitness of the sum guarantee an invertible sum and a unique estimate for the mean. This is the case if one class has for observation matrix a full-rank square matrix.

The invertability condition shows that if certain components of the vector x are systematically missing (such, as say, pressure levels above 50 $mbar$), the estimate for the mean is not unique. Under these conditions, an unique estimate of the mean can be found only by imposing an other condition for the estimate. A minimum norm solution, for instance, would multiply the RHS of (IV.B.26) by the Penrose-Moore pseudo-inverse of S to provide such an estimate (Rao, 1973, Kailath, 1976). When the estimate is not unique and requires further asssumptions to be defined, it indicates that the sampling of the vectorial process is not adequate.

When defined, the estimator for the mean is unbiased, independtly of the estimate for the covariance matrix R_o. This is shown by setting the sample matrix m_{rc} to its expectation

$$m_{rc} = \mathcal{H}_{lc}{}^{t} \begin{pmatrix} \Lambda & 0 \end{pmatrix} \mathcal{H}_{rc} m \qquad (IV.B.31)$$

in equation (IV.B.) and by factoring the mean m and by noting that

$$\begin{pmatrix} I \\ \hat{R}_{okr}\hat{R}_{orr}^{-1} \end{pmatrix} \Lambda^{-1} \mathcal{H}_{lc} \mathcal{H}_{lc}{}^{t} \begin{pmatrix} \Lambda & 0 \end{pmatrix} \mathcal{H}_{rc} m = \Gamma(\hat{R}_o, H_c) m \qquad (IV.B.32)$$

The unbiasness of the optimal estimate for the mean independly of the covariance is noteworthy because first and second moments are independent quantities for most pdf, in particular for the Gaussian pdf. However, to be optimal, the estimate of the mean still requires the optimal estimate for the covariance. Note, once the optimal covariance matrix is found, the estimate for the mean vector is an explicit function of the sample means.

The solution of the maximum likelihood estimation problem is not as simple for the covariance as it is for the mean, even in the simplified case where the mean is known exactly. As for the estimation of the mean, let us multiply RHS and LHS of (IV.B.24) by R_o on the left and right and introduce the extrapolation matrices Γ. After manipulation, this yields

$$\sum_{c=1}^{l} N_c \Gamma(\hat{R}_o, H_c)\hat{R}_o = \hat{R}_o \sum_{c=1}^{l} N_c \Gamma(\hat{R}_o, H_c) = \sum_{c=1}^{l} N_c \Gamma(\hat{R}_o, H_c) \mathcal{H}_{rc}$$

$$\begin{pmatrix} \Lambda^{-1} \mathcal{H}_{lc}{}^t (R_{zc} + (H_c\hat{m} - m_{zc})(H_c\hat{m} - m_{zc})^t) \mathcal{H}_{lc} \Lambda^{-1} & 0 \\ 0 & 0 \end{pmatrix} \mathcal{H}_{rc}{}^t \Gamma^t(\hat{R}_o, H_c) \tag{IV.B.33}$$

This design equation is similar to the one defining the estimate for the mean but to the contrary of the mean estimate, except in special cases, such as perfect observations, this design equation does not define the estimate for the covariance explicitely because the extrapolation matrix does depend on the estimate of the covariance matrix itself.

The condition for an unique estimate is the same than for the mean (Equation (IV.B.29)) and the estimate is

$$\hat{R}_o = \Big(\sum_{c=1}^{l} N_c \Gamma(\hat{R}_o, H_c)\Big)^{-1} \sum_{c=1}^{l} N_c \Gamma(\hat{R}_o, H_c) \mathcal{H}_{rc}$$

$$\begin{pmatrix} \Lambda^{-1} \mathcal{H}_{lc}{}^t (R_{zc} + (H_c\hat{m} - m_{zc})(H_c\hat{m} - m_{zc})^t) \Lambda^{-1} \mathcal{H}_{lc} & 0 \\ 0 & 0 \end{pmatrix} \mathcal{H}_{rc}{}^t \Gamma^t(\hat{R}_o, H_c) \tag{IV.B.34}$$

$$= \Big(\sum_{c=1}^{l} N_c \Gamma(\hat{R}_o, H_c)\Big)^{-1} \sum_{c=1}^{l} N_c \Gamma(\hat{R}_o, H_c) \mathcal{H}_{rc} \Xi_{zc}(\hat{m}) \Gamma^t(R_o, H_c)$$

The quantity $\Xi_{zc}(\hat{m})$ is the covariance of the signal (x) found by taking the Moore-Penrose pseudo inverse of the observation sample covariance associated with the class' observation matrix. It depends on the observation class C and on the estimated mean vector $\hat{m}$. This matrix will be referred to as the Moore-Penrose equivalent (MPE) sample covariance.

As for the mean vector, the covariance cannot be evaluated for systematically umsampled components. As it will latter be discussed, the estimate can be made unique by the adjunction in the algorithm of an a-priori covariance. From an information theoretic point of view, in the case of non unique estimates, the uncharacterized terms should be assumed uncorrelated with an infinite variance ("minimum information").

Let us characterize the expected value of the covariance to check if the estimate is biased or not. Let us replace the sample covariance R_{zc} by its expected value as a function of the true covariance matrix R_o according to

$$R_{zc} = \mathcal{H}_{lc} \Lambda R_{orr} \Lambda \mathcal{H}_{lc}{}^t \qquad (IV.B.35)$$

Since the estimate for the mean is unbiased, the equation defining the covariance becomes

$$\sum_{c=1}^{l} N_c \mathcal{H}_{rc}{}^t \begin{pmatrix} \hat{R}_{orr} & \hat{R}_{ork} \\ \hat{R}_{okr} & \hat{R}_{okr}\hat{R}_{orr}^{-1}\hat{R}_{ork} \end{pmatrix} \mathcal{H}_{rc} = \sum_{c=1}^{l} \Gamma(\hat{R}_o, H_c) \mathcal{H}_{rc} \begin{pmatrix} R_o & 0 \\ 0 & 0 \end{pmatrix} \mathcal{H}_{rc}{}^t \Gamma^t(\hat{R}_o, H_c) \qquad (IV.B.36)$$

or

$$\sum_{c=1}^{l} N_c \mathcal{H}_{rc}{}^t \begin{pmatrix} \hat{R}_{orr} & \hat{R}_{ork} \\ \hat{R}_{okr} & \hat{R}_{okr}\hat{R}_{orr}^{-1}\hat{R}_{ork} \end{pmatrix} \mathcal{H}_{rc} =$$

$$\sum_{c=1}^{l} N_c \mathcal{H}_{rc}{}^t \begin{pmatrix} R_{orr} & R_{orr}\hat{R}_{orr}^{-1}\hat{R}_{ork} \\ \hat{R}_{okr}\hat{R}_{orr}^{-1}R_{orr} & \hat{R}_{okr}\hat{R}_{orr}^{-1}R_{orr}\hat{R}_{orr}^{-1}\hat{R}_{ork} \end{pmatrix} \mathcal{H}_{rc} \qquad (IV.B.37)$$

As discussed, when the estimate is defined, it is unique. Thus any solution to (IV.B.37) is the unique estimate. Replacing $\hat{R}_{orr}$ by R_{orr} satisfies (IV.B.37), the estimate of the covariance is also unbiased.

Let us note that in the special case of perfect observations ($H_c = I$), or more generally the observation matrix is invertible, both mean and covariances estimates are the sample mean and covariance, a result already known from time series analysis (Oppenheim and Schaffer, 1975). Note that equation (IV.B.30) which

defines the maximum likelihood estimate for the mean requires the knowledge of the estimate for the covariance which itself depends on the estimate of the mean through an implicit equation (IV.B.34).

To solve these two estimates, one must use an iterative approach. At each point in the recursion, let us denote the estimate for the mean by $\hat{m}_n$ and the estimate for the covariance by $\hat{R}_n$. The proposed recursion can be written as

$$\hat{R}_{n+1} = \left(\sum N_c \Gamma(\hat{R}_n, H_c)\right)^{-1} \sum_{c=1}^{l} N_c \Gamma(\hat{R}_n, H_c) \Xi_{zc}(\hat{m}_n) \Gamma^t(\hat{R}_n, H_c)\Big)^{-1} \quad (IV.B.38)$$

and for the mean vector

$$\hat{m}_{n+1} = \left(\sum_{c=1}^{l} N_c \Gamma(\hat{R}_n, H_c)\right)^{-1} \sum_{c=1}^{l} N_c \begin{pmatrix} I \\ \hat{R}_{nkr}\hat{R}_{nrr}^{-1} \end{pmatrix} \mathcal{H}_{rc} \Lambda^{-1} \mathcal{H}_{lc} m_{zc} \quad (IV.B.39)$$

This type of iteration is known as the Picard iteration (Ortega and Rheinboldt, 1970).

The algorithm is initialized at step 0 by chosing a class varying mean satisfying

$$H_c \hat{m}_o = m_{zc} \quad (IV.B.40)$$

and an a-priori covariance matrix based on climatology arguments. A conservative choice for the a-priori covariance would be a diagonal matrix with all entries on the diagonal equal. This corresponds to a "minimum information" solution in the sense that no structure about the covariance matrix is inferred in the algorithm.

Convergence proof of this iterative algorithm is intricate because it relies on the properties of infinite sequences in compact sets. The only hypothesis on the recursion is that the sampling of the vectorial process is rich enough so that the quantity $S = \sum_{c=1}^{l} N_c \Gamma(\hat{R}_n, H_c)$ is invertible. The existence of the inverse of S does not depend on the estimated covariance $\hat{R}_n$.

From the above equations, it should be evident that if the sequence of covariance matrices converges, so does the sequence of means. For this reason, we will concentrate our effort on the covariance matrix.

First, let us demonstrate that the estimate $\hat{R}_n$ remains bounded. Using Cauchy-Schwartz inequality, equation (IV.B.38) yields

$$||\hat{R}_{n+1}|| \leq ||\Big(\sum_{c=1}^{l} N_c \Gamma(\hat{R}_n, H_c)\Big)^{-1}|| \; ||\sum_{c=1}^{l} N_c \Gamma(\hat{R}_n, H_c) \Xi_{zc}(\hat{m}_n) \Gamma(\hat{R}_n, H_c)|| \tag{IV.B.41}$$

where $||.||$ denotes the Euclidian norm defined over the vectorial space of matrices. Because each extrapolation matrix Γ is positive definite, the sum in the RHS can be bounded above by

$$||\sum_{c=1}^{l} N_c \Gamma(\hat{R}_n, H_c) \Xi_{zq}(\hat{m}_n) \Gamma^t(R_n, H_q)|| \tag{IV.B.42}$$

where **Q** is one of the observation classes. Using Cauchy-Schwartz again yields

$$\begin{aligned} ||\hat{R}_{n+1}|| &\leq ||\Big(\sum_{c=1}^{l} N_c \Gamma(\hat{R}_n, H_c)\Big)^{-1}|| \; ||\sum_{c=1}^{l} N_c \Gamma(\hat{R}_n, H_c) \Xi_{zq}(\hat{m}_n) \Gamma^t(R_n, H_q)|| \\ &\leq ||\Xi_{zq}(\hat{m}_n) \Gamma^t(R_n, H_q)|| \end{aligned} \tag{IV.B.43}$$

Projections operators are non expansive continuous operators, hence

$$||X|| \leq ||PX|| \tag{IV.B.44}$$

This implies that

$$||\hat{R}_{n+1}|| \leq ||\Xi_{zq}(\hat{m}_n)|| \leq ||\Xi_{zc}(\hat{m}_n)||_{max} \tag{IV.B.45}$$

where the maximum is taken over all MPE sample covariance matrices. The same type of argumentation can be used to bound the estimate of the mean vector, that is

$$||\hat{m}_n|| \leq ||m_{zq'}|| \leq ||m_{zc}||_{max} \tag{IV.B.46}$$

This bound implies that the MPE sample covariance with the largest norm is finite and that, at any iteration,

$$\begin{aligned} ||\hat{R}_n|| &\leq \mathcal{M}_R \\ ||\hat{m}_n|| &\leq \mathcal{M}_m \end{aligned} \tag{IV.B.47}$$

The Bolzano-Weierstrass theorem (Bertsekas, 1976) implies that there exists two subsequences of estimates $\hat{R}_n^*$ and $\hat{m}_n^*$ such that

$$\{\hat{R}_n^*\} \subseteq \{R_n\}, \qquad \{\hat{m}_n^*\} \subseteq \{m_n\} \tag{IV.B.48}$$

which converge to the stationary point of the recursion formula which is the unique optimum estimator of the moments under the working assumptions.

Let us now concentrate on determining the step size of the recursion in the covariance matrix. Replacing $\Xi_{zc}(\hat{m}_n)$ in equation (IV.B.38) by $\Xi_{zc}(\hat{m}_n) - \hat{R}_n + \hat{R}_n$ yields

$$\hat{R}_{n+1} - \hat{R}_n = \left(\sum_{c=1}^{l} N_c \Gamma(\hat{R}_n, H_c)\right)^{-1} \sum_{c=1}^{l} N_c \Gamma(\hat{R}_n, H_c)(\hat{R}_n - \Xi_{zc}(\hat{m}_n))\Gamma(\hat{R}_n, H_c) \tag{IV.B.49}$$

Computing the norms for both sides yields, after invoking Cauchy-Schwartz inequality as above

$$||\hat{R}_{n+1} - \hat{R}_n|| \leq ||\hat{R}_n - \Xi_{zc}(\hat{m}_n)||_{max} \tag{IV.B.50}$$

where the maximization is taken over all classes C. If the RHS difference decreases with the index n, so will the LHS. Consider now a specific class C' out of all classes. The difference $\hat{R}_{n+1} - \Xi_{zc'}(\hat{m}_n)$ can be expressed as

$$\hat{R}_{n+1} - \Xi_{zc'}(\hat{m}_n) = \hat{R}_n - \Xi_{zc'}(\hat{m}_n) - \left(\sum_{c=1}^{l} N_c \Gamma(\hat{R}_n, H_c)\right)^{-1}$$

$$\left(\sum_{c=1}^{l} N_c \Gamma(\hat{R}_n, H_c)(\hat{R}_n - \Xi_{zc}(\hat{m}_n))\Gamma(\hat{R}_n, H_c)\right) \tag{IV.B.51}$$

after manipulations,

$$\hat{R}_{n+1} - \Xi_{zc'}(\hat{m}_n) = \left(\sum_{c=1}^{l} N_c \Gamma(\hat{R}_n, H_c)\right)^{-1} \left[\sum_{c=1}^{l} N_c \Gamma(\hat{R}_n, H_c)\right.$$

$$\left.\left(\hat{R}_n - \Xi_{zc'}(\hat{m}_n) - \hat{R}_n \Gamma(\hat{R}_n, H_c) + \Xi_{zc}(\hat{m}_n)\Gamma(\hat{R}_n, H_c)\right)\right] \tag{IV.B.52}$$

and

$$\hat{R}_{n+1} - \Xi_{zc'}(\hat{m}_n) = \left(\sum_{c=1}^{l} N_c \Gamma(\hat{R}_n, H_c)\right)^{-1} \sum_{c=1}^{l} N_c \Gamma(\hat{R}_n, H_c) \left(\Xi_{zc}(\hat{m}_n) - \Xi_{zc'}(\hat{m}_n)\right)$$

$$+ \left(\sum_{c=1}^{l} N_c \Gamma(\hat{R}_n, H_c)\right)^{-1} \sum_{c=1}^{l} N_c \Gamma(\hat{R}_n, H_c) \left(\hat{R}_n - \Xi_{zc}(\hat{m}_n)\right) \left(1 - \Gamma(\hat{R}_n, H_c)\right) \quad (IV.B.53)$$

Using the triangular and the Cauchy-Schwartz inequalities yields

$$||\hat{R}_{n+1} - \Xi_{zc'}(\hat{m}_n)|| \leq ||\Xi_{zc}(\hat{m}_n) - \Xi_{zc'}(\hat{m}_n)||_{max} + ||\hat{R}_n - \Xi_{zc}(\hat{m}_n)||_{max} \quad (IV.B.54)$$

Two different cases must be analysed depending on the distance between MPE sample covariances. If the term $||\Xi_{zc}(\hat{m}_n) - \Xi_{zc'}(\hat{m}_n)||_{max}$ equals zero, the difference between $\hat{R}_n$ and $\Xi_{zc'}(\hat{m}_n)$ remains the same for all classes. This difference is zero if all classes have the same MPE sample covariances. However, by construction, two MPE sample covariances can only be the same if the associated observation matrices share the same row space, that is if they belong to the same class. Hence, this situation can only occur in instance where there is only one observation class which must be perfect under the working assumptions. In such a case, the optimal estimates for the covariance and mean are the sample covariance and mean (once equivalenced to the identity matrix according to (IV.B.21)), and the difference between $\hat{R}_{n+1}$ and $\Xi_{zc'}(\hat{m}_n)$ is zero. The iterative algorithm then converges at step zero.

In general, however, there are different observation classes and

$$||\Xi_{zc}(\hat{m}_n) - \Xi_{zc'}(\hat{m}_n)||_{max} \neq 0 \quad (IV.B.55)$$

Therefore

$$||\hat{R}_{n+1} - \Xi_{zc'}(\hat{m}_n)|| \leq ||\hat{R}_n - \Xi_{zc}(\hat{m}_n)||_{max} \quad (IV.B.56)$$

This is true for all classes C'. Hence

$$||\hat{R}_{n+1} - \Xi_{zc}(\hat{m}_n)||_{max} \leq ||\hat{R}_n - \Xi_{zc}(\hat{m}_n)||_{max} \quad (IV.B.57)$$

This implies that the sequence of differences $||\hat{R}_{n+1} - \hat{R}_n||$ is bounded above by a decreasing sequence, then tends to zero as n increases, and $\hat{R}_n$ is Cauchy sequence. Such a sequence converges at the optimum estimate for the covariance R according to the Bolzano-Weierstrass theorem. The sequence of mean vector estimates also converges towards the optimum.

The Picard recursion for mean and covariance estimates converge to a single (optimal) solution as long as the eigenvalues of the sum $S = \sum_{c=1}^{l} N_c \Gamma(\hat{R}_n, H_c)$ are all greater that zero, that is this quantity admits an inverse. This convergence proof was verified experimentally in the context of compiling NMC observational data.

Let us compare the maximum likelihood estimators with the ones derived from statistical Bayesian extrapolation. Statistical signal extrapolation will optimally estimate in a mean square sense each of the realization x_i from the incomplete observations z_i. Then based on these extrapolated profiles, it computes the standard mean and covariances. This Bayesian extrapolation based method guarantes positive semidefinite covariance matrices.

Let m_a and R_a denote the a-priori mean and covariances associated with the extrapolation. The optimum estimate $\hat{x}_i$ of the realization z_i is given by (Schweppe, 1973)

$$\hat{x}_i = m_a + R_a H_i^t (H_i R_a H_i^t)^{-1} (z_i - H_i m_a) \qquad (IV.B.58)$$

The estimate for the mean is found using the extrapolated profiles

$$\hat{m} = \frac{1}{N} \sum_{i=1}^{N} \hat{x}_i = \frac{1}{N} m_a + \frac{1}{N} \sum_{i=1}^{N} R_a H_i^t (H_i R_a H_i^t)^{-1} (z_i - H_i m_a) \qquad (IV.B.59)$$

which becomes once one introduces the concept of class

$$\hat{m} = m_a + \frac{1}{N} \sum_{c=1}^{l} N_c R_a H_c^t (H_c R_a H_c^t)^{-1} (m_{zc} - H_c m_a) \qquad (IV.B.60)$$

Introducing the extrapolations matrices Γ yields

$$\begin{aligned}\hat{m} =& m_a + \frac{1}{N}\sum_{c=1}^{l} N_c \Gamma(R_a, H_c)(\eta_c - m_a) \\ =& \frac{1}{N}\sum_{c=1}^{l} N_c \Gamma(R_a, H_c)\eta_c + \frac{1}{N}\sum_{c=1}^{l}(I - \Gamma(R_a, H_c))m_a\end{aligned} \tag{IV.B.61}$$

where the vector η_c is defined as

$$\eta_c = \mathcal{H}_{rc}\Lambda^{-1}\mathcal{H}_{lc}m_{zc} \tag{IV.B.62}$$

The vector η_c plays for the mean the role that Ξ_{zc} played for the covariance.

Contrary to the maximum likelihood estimate, this estimate of the mean is a biased one except if the a-priori mean is the true mean of the process or if all observations classes are perfect (invertible observation matrices).

The Bayesian extrapolation estimate for the covariance is

$$\hat{R} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{m})(x_i - \hat{m})^t \tag{IV.B.63}$$

which can be expressed using the MPE sample covariances as

$$\begin{aligned}\hat{R} &= \frac{1}{N}\sum_{c=1}^{l} N_c \Gamma(R_a, H_c)(\Lambda^{-1}\mathcal{H}_{lc}{}^t R_{zc}\mathcal{H}_{lc}\Lambda^{-1} + (\hat{m} - \eta_c)(\hat{m} - \eta_c)^t)\Gamma(R_a, H_c)^t \\ &= \frac{1}{N}\sum_{c=1}^{l} N_c \Gamma(R_a, H_c)\Xi_{zc}(m_a)\Gamma^t(R_a, H_c)\end{aligned} \tag{IV.B.64}$$

This estimate is similar to the maximum likelihood estimate except for the normalization. For the maximum likelihood estimate, the normalization of the sum of weighted sample covariance matrices is $\sum_{c=1}^{l} N_c \Gamma(R_o, H_c)^{-1}$, whereas for the Bayesian extrapolation estimate, the normalizing factor is simply N^{-1}. The Bayesian extrapolation estimate treats all observation identically irrespectively of the dimension of row and kernel space. The two estimates will be close to one another when the observation matrices H_c are full rank for most of the observations.

This is true, because the closer H_c is from having M independent rows, the closer the projection matrix Γ will be from the identity. For the other limiting case, where the observation matrices only possess a few channels, the Bayesian extrapolation estimate will be more sensible to error than the maximum likelihood estimate because it treats all class independently of their number of channels.

Contrary to the maximum likelihood estimates, the Bayesian extrapolation estimates are always uniquely defined, but it is biased except in the unlikely event where m_a and R_a are the true values of the moments to be estimated. Using a Picard type iterative implementation of the Bayesian estimation of mean or covariance does not yield unbiased estimates except if

$$\frac{1}{N}\sum N_c\Gamma(R_o,H_c)=I \qquad (IV.B.65)$$

This condition is more restrictive than the condition of existence ,and unbiasness, of the maximum likelihood estimates which only requires the eigenvalues of $1/N\sum_{c=1}^{l} N_c\Gamma(R_o,H_c)$ to be strictly greater than zero (they are less than or equal to one by construction).

For both types of estimates, it is possible to include some a-priori knowledge by using pseudo measurements associated with a specific class. The relative weight of the a-priori knowledge is controlled by N_{ca} the number of pseudo points. The adjunction of a-priori knowledge with an invertible observation matrix will also assure the invertibility of the sum in equation (IV.B.29) and thus unique maximum likelihood estimates for the mean and covariance. Thus, if a climatological description of the atmosphere is available (standard atmosphere or idealized description), one can derive a unique estimate for the mean and covariance even if some levels are always unsampled. This adjunction garantee an unique but may be not relevant estimate. Note that in certain cases, it may be inappropriate to add a-priori knowledge or pseudo-measurements solely on the basis of garanteing the existence of unique estimators. This is true because these estimates will not be robust to changes in

the sample mean and covariance when the number of pseudomeasurements (N_{ca}) is small.

The issue of robustness of the estimators, although relevant to this analysis, was not examined because of flood and time constraints. Likewise, no conclusion were drawn concerning the advantage of incorporating in the estimate all sample means and covariances rather than discarding some.

Finally, let us consider the impact of measurement noise on the optimum estimates and on their calculation.

The observation equations are modified to include the noise according to

$$z_i = H_i x_i + \epsilon_i \qquad (IV.B.66)$$

where the observation noise power varies, as everything else, as a function of the observation index.The matrix Q_i denotes the covariance of the noise at measurement i ($Q_i =< \epsilon_i \epsilon_i >$). The sensing noise is assumed independent of the vectorial process x.

The probability of the set of measurements $\{z_1\}$ (equation (IV.B.1)) remains the product of the individual probabilities $p(z_i|m, R_o, Q_i)$ which becomes

$$p(z_i|m, R_o, Q_i) = \frac{1}{2\pi^{rank H_i/2}} \frac{1}{|H_i R_O H_i^t + Q_i|}$$

$$\exp\{-\frac{1}{2}(z_i - H_i m)^t (H_i R_o H_I^t + Q_i)^{-1}(z_i - H_i m)\} \qquad (IV.B.67)$$

The covariance of the observation is modified by adding the covariance matrix of the observation noise to the covariance matrix of the signal. The SVD representation of this sum is

$$H_i R_o H_i + Q_i = \mathcal{H}_{ri} \Lambda_i (R_{orr} + \mathcal{H}_{ri}{}^t \Lambda_i^{-1} Q_i \Lambda_i^{-1} \mathcal{H}_{ri}) \Lambda_i \mathcal{H}_{ri} \qquad (IV.B.68)$$

A observation class can now be defined as a set of matrices that share the same row space and for which the product $\mathcal{H}_{ri}{}^t \Lambda_i^{-1} Q_i \Lambda_i^{-1} \mathcal{H}_{ri}$ is the same. The computation of the diverse derivatives remain the same, except that in equations

(IV.B.10), (IV.B.14), and (IV.B.17), the matrix R_{orr}^{-1} is replaced by $(R_{orr} + \mathcal{H}_{ri}{}^t \Lambda_i^{-1} Q_i \Lambda_i^{-1} \mathcal{H}_{ri})^{-1}$. The extrapolation matrices become

$$\Gamma(R_o, H_c) \rightarrow \Gamma'(R_o, H_c) = R_o H_c^t (H_c R_o H_c^t + Q_c)^{-1} H_c \qquad (IV.B.69)$$

which are no more projection matrices (the eigenvalues of these matrices are either zero or less than one). The optimum estimates become

$$\hat{m} = \Big(\sum_{c=1}^{l} N_c \Gamma'(\hat{R}_o, H_c)\Big)^{-1} \sum_{c=1}^{l} N_c \Gamma(R_o, H_c) H_c^t (H_c H_c^t)^{-1} m_{zc} \qquad (IV.B.70)$$

for the mean and

$$\hat{R}_o = \Big(\sum_{c=1}^{l} N_c \Gamma'(\hat{R}_o, H_c)\Big)^{-1} \sum_{c=1}^{l} N_c \Gamma(\hat{R}_o, H_c) \mathcal{H}_{rc}$$

$$\begin{pmatrix} \Lambda^{-1} \mathcal{H}_{lc}{}^t (R_{zc} + (H_c \hat{m} - m_{zc})(H_c \hat{m} - m_{zc})^t) \Lambda^{-1} \mathcal{H}_{lc} & 0 \\ 0 & 0 \end{pmatrix} \mathcal{H}_{rc}{}^t \Gamma'^t(\hat{R}_o, H_c) \qquad (IV.B.71)$$

for the covariance. These two estimates are found using the same picard iterative algorithms than the one introduced in the absence of observation noise. The convergence proof remains the same, except that it requires the signal to noise ratio to be finite at each observation. If the noise becomes infinite, the measurement must be discarded.

To summarize, in this appendix, novel estimates for both mean and covariance of vectorial signal observed under varying conditions were introduced and characterized. An efficient iterative method to solve for these estimates was presented and its convergence proved. Applications of this estimation to the compilation of radiosonde measurements in presented in the core of chapter V.

CHAPTER V
MULTIDIMENSIONAL REMOTE SENSING OF TEMPERATURE FIELDS

"They say: one, and one, and one is three." John Lennon, Paul Mc Cartney

"Even in the shadow of Death, two and two is not six." Leo Tolstoy

V.1 Introduction

This Chapter examines the theoretical aspects of remote sensing - the single spot retrieval problem as well as the multi-spot retrieval problem. The organization of this Chapter is as follows. The optimal single spot retrieval operator as well as the associated error is derived; then both quantities are studied for various limiting forms of the statistical description of the temperature fields. Since the temperature profile is discretized, the estimation can be performed in any basis representation. The most natural one is the one created by the discretization which one could coin 'user-oriented'. A more appropriate basis representation is found considering the singular value decomposition of the observation matrix; in that sense it is 'instrument-oriented'. Because of polar satellite sounding patterns, relevant considerations and notations are introduced. The Chapter proceeds with the analysis of the two-spot retrieval problem in either basis, and examines the improvement created by the inclusion of horizontal statistics. The design equation for multidimensional retrieval operators is derived and analyzed in the context of diagonalizing operators. This latter interpretation yields a fast design method for instruments with fine resolution, and permits a connection with previous work when the support of the retrieval operator increases. Finally, in an appendix, the 'instrument-oriented' interpretation of the retrieval algorithm is used to explain the

improvement in previous research dealing with Kalman filtering along the track of the satellite.

V.2 One Dimensional Retrievals

In this section, the single spot optimal retrieval operator is derived and analysed in the natural representation of temperature profiles and then along two of the fundamental subspaces of the observation matrix.

V.2.1 Observation Equation.

Equation II.4.7 showed that once the temperature profiles have been discretized, the weighting functions can themselves discretized and the observation equation becomes

$$\overline{T}_b = W\overline{T} + \overline{\epsilon} \qquad (V.2.1)$$

where $\overline{T}$ is the 33×1 vector where each component corresponds to the temperature at one of the pressure levels introduced in section III.4.1, $\overline{\epsilon}$ is the receiver noise, $\overline{T}_b$ is the observed brightness temperature and W is the weighting matrix. The weighting matrix depends on the functions used to interpolate between sampled points.

As noted in section III.3.1, in the troposphere the temperature profile can be modeled as piecewise linear. Extending this modeling to the stratosphere, one represents the temperature profiles as a series of slabs, determined by the endpoints of the slabs centered around the 33 pressure levels, and by a lapse rate. Figure V.2.1 presents the discrete weighting matrix associated with this interpolation method between sample levels for MSU over land. The temperature profile used to compute the weighting matrix is the US standard atmosphere (US Standard Atmosphere Supplements, 1966)

V.2.2 Derivation of Retrieval Operator and Error.

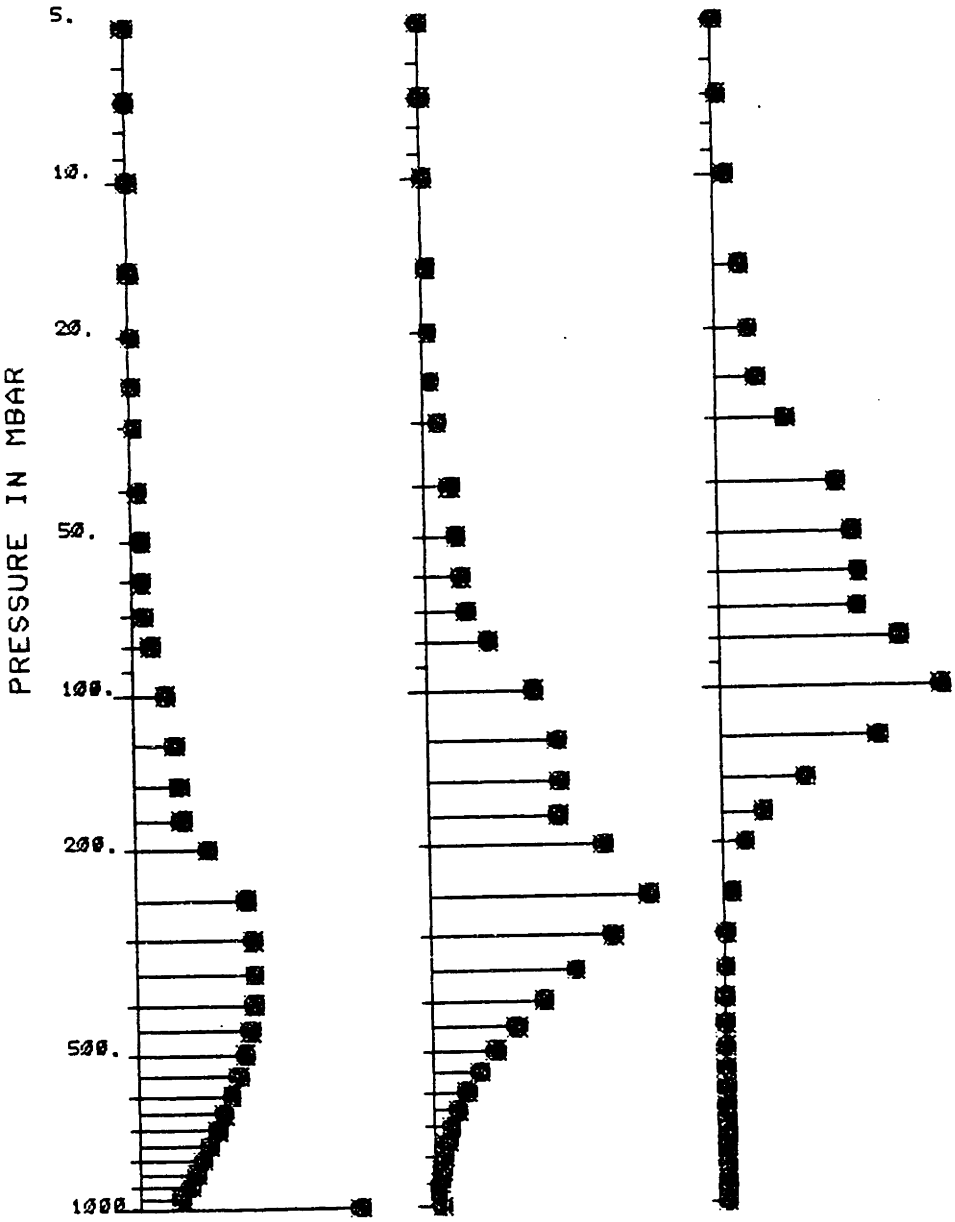


Figure V.2.1: Discrete weighting matrix for MSU over land

This derivation of the optimal (in a mean square sense) one-spot retrieval operator as well as the multi-spot retrieval operator is by no means novel. However, in view of the present literature on static Wiener filtering (Kailath, 1981, and references therein), the emphasis on the issue of the a-priori mean is new. This

emphasis is justified by the fact that estimation of a stochastic process in general, and remote sensing of temperature in particular, is usually an affine rather than a linear operation, thus requiring offset and gain.

The problem with the estimation of the mean is caused by the fact that atmospheric temperature does not present a constant mean for all locations and typically varies on a latitudinal and seasonal basis. This is true because of the different scales of variations in the evolution of temperature fields. Hence, the traditional claim of estimation theory and signal linear processing that the mean is a secondary problem and that only the estimation of the deviations of the signal from that mean must be considered is inadequate in the context of remote sensing. In other words, the mean profile cannot be treated as the deterministic part of the signal, but rather as a slowly varying stochastic part. Moreover, because atmospheric temperature is not a stationary process, the mere definition of the mean, of deviations thereof, and of optimal filters is a difficult task.

The issue of the a priori mean also arises when the stochastic process under consideration is wide-sense stationary, allowing for the optimal design of linear space invariant filters. This is true because the estimations of simulated or real signals are implemented on finite sets. These sets are derived from specific realizations of underlying random process. Even if the underlying process is assumed to be ergodic, the probability that the specific finite realization of that process has a sample mean equal to the first moment of the probability density function of the process is small. Should the process adopt continuous values, this probability would become zero.

Optimal mean-square optimization is performed in the context of stochastic processing and thus requires the modeling of temperature fields and their resulting brightness temperature fields as random variables.

The framework is as follows: Let x, y denote a jointly stochastic process, where x is the parameter to be estimated and y is the observation. Also, x and y are

multivariate and multidimensional. Their joint probability density function (pdf) $p(x,y)$ is assumed given. This quantity allows the computation of the conditional pdf of the observation given the parameter $p(y|x)$.

The optimal estimate $\hat{x}(y)$ of the variable x in a mean square sense (MMSE) is defined by the minimization of the conditional cost function

$$< ((\hat{x}(y) - x)^2|y) >= \int_X (\hat{x}(y) - x)^2 p(y|x)dx \qquad (V.2.2)$$

where $< . >$ denotes the expectation across all possible values of the doublet (x,y). Computing this cost function,

$$E = Var(x|y) + (< (x|y) > -\hat{x}(y))^2 \qquad (V.2.3)$$

where $Var(x|y)$ denotes the conditional of the variable x given y and $< (x|y) >$ its conditional expectation. Thus, the optimal estimate (MMSE) for the variable x is its conditional expectation given the observation. It requires knowledge of the entire pdf of the joint process, or equivalently of all its moments.

A suboptimal solution is sought in the form

$$\hat{x}(y) = a + by \qquad (V.2.4)$$

where a and b have appropriate size. This is the linear MMSE or "LMMSE". The computation of the LMMSE does require the first and the second moment of the joint pdf, which also requires the entire pdf but can readily be estimated from specific realizations. The LMMSE will be optimal for those pdf where the conditional expectation is an affine function of one of the variables. Such is the case for the Gaussian pdf.

Let us take the doublet $(< \overline{T} >, R)$ to denote the mean and vertical covariance matrix of the temperature field being retrieved. As noted, both the mean and covariance are unknown. The impact of the error in the covariance matrix and ways

to correct for it will be discussed in Chapter VII. For the present analysis, the covariance matrix will be assumed known exactly. To analyze the impact of the offset term in the affine estimation, let $\overline{T}_a$ denote the a-priori mean ("guess profile") for the temperature distribution.

The retrieval equation can be written as

$$\hat{\overline{T}} = D\overline{T}_b = D(W\overline{T} + \overline{\eta}) \qquad (V.2.5)$$

Note that both the temperature profile $\overline{T}$ and the measurement noise $\overline{\eta}$ are taken as stochastic processes. The product of the retrieval matrix multiplied by the weighting matrix is referred to as the transfer matrix of the retrieval process by analogy with linear theory (Nathan, 1983). It is not a full rank matrix because the number of channels is less (for most remote sensing systems in general, for MSU and AMSU in particular) than the number of pressure levels used in the computation of the brightness temperature.

The error vector for the retrieval is given by

$$\overline{e} = \hat{\overline{T}} - \overline{T} = D\overline{T}_b - \overline{T} = D(W\overline{T} + \overline{\eta}) - \overline{T} \qquad (V.2.6)$$

The criterion used in the minimization is the square of the norm or energy of the error vector, namely

$$E = < e^2 > = < \overline{e}^t\overline{e} > = trace\{< \overline{e}\overline{e}^t >\} = trace\{\mathcal{E}\} \qquad (V.2.7)$$

where $< . >$ denotes the expectation and $\mathcal{E}$ denotes the error matrix whose trace is the error criterion. The introduction of the error covariance matrix will become more obvious in the decomposition of the error onto observable and unobservable components and other bases of profile representations. The error at a given pressure level or for a linear combination of such pressure levels is found by multiplying the error matrix left and right by the appropriate vector.

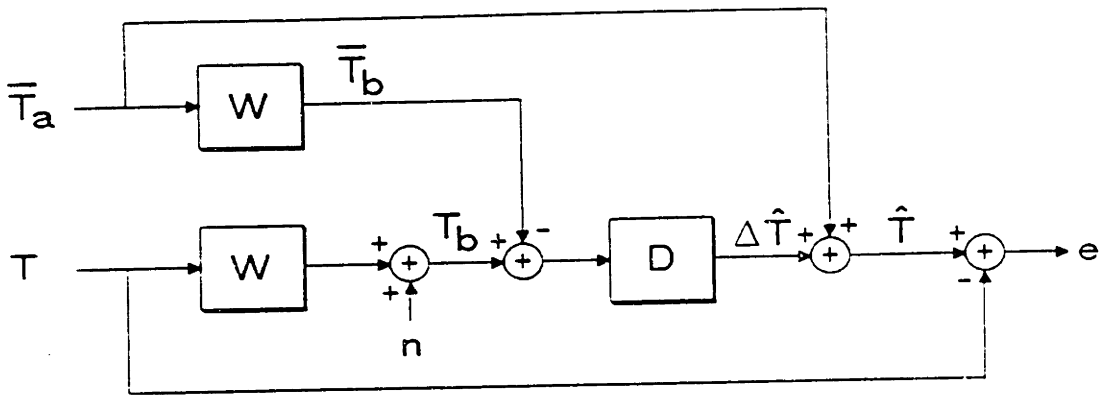


Figure V.2.2: Block Diagram Representation of the Single Spot Retrieval Problem

The retrieval operator does not operate on the brightness temperature, but on the deviation of the brightness temperature from the a priori mean brightness temperature value computed from the a priori temperature profile. This is done to reduce the energy in the signal and consequently the energy in the error. Moreover, reducing the signal energy reduces the impact of non-linearities in the radiative transfer equation. Thus

$$\hat{\overline{T}} - \overline{T}_a = DW\overline{T} + D\overline{\eta} - DW\overline{T}_a \qquad (V.2.8)$$

and

$$\hat{\overline{T}} = (I - DW)\overline{T}_a + DW\overline{T} + D\overline{\eta} \qquad (V.2.9)$$

The error vector becomes

$$\overline{e} = (I - DW)(\overline{T}_a - \overline{T}) + D\eta \qquad (V.2.10)$$

An overall picture of the retrieval problem is given by the block diagram in Figure V.2.2. By analogy with linear system theory, the product of the matrix DW is called the transfer or transmission matrix of the retrieval process; it will heretofore

be denoted by $\mathcal{T}$. The transfer matrix is related to the concept of impulse response introduced in the context of remote sensing by Rosenkranz (1971). The impulse response of the retrieval process for a pressure level is defined as the retrieved temperature profile corresponding to the change in brightness temperature created by a 1 K area impulse located at that pressure. Each column of the transfer matrix is the impulse response of the corresponding pressure level.

The error criterion is the square of the average Euclidian norm of the error vector.

$$E = trace\{\mathcal{E}\} = trace\{(I-DW) < (\overline{T}_a-\overline{T})(\overline{T}_a^t-\overline{T}^t) > (I-W^tD^t)+D < \overline{\eta} > D^t\} \tag{V.2.11}$$

The trace of the matrix $< (\overline{T}_a - \overline{T})(\overline{T}_a^t - \overline{T}^t) >$ corresponds to the energy of the signal being estimated. In the above equation, the expectations are taken over all temperature profiles in a given climatology and realizations of the sensing noise.

The optimum a priori mean is found by setting the partial derivative of the error with respect to the vector $\overline{T}_a$ to zero.

$$\begin{aligned} \frac{\partial E}{\partial \overline{T}_a} &= \overline{T}_a 2(I - WD) - < \overline{T} > 2(I - DW) \\ &= (\overline{T}_a - < \overline{T} >)2(I - DW) \end{aligned} \tag{V.2.12}$$

The optimal a priori mean profile will not be unique if the matrix in the right hand side of (V.2.12) is not full rank. As will be shown in the next section, such is the case whenever some of the channels do not have sensing noise and thus perfectly measure a linear combination of the components of the temperature profile. Perfect measurement constraint the temperature profiles to lay in a specific hyperplane of $\mathbf{R}^k$ where k is the dimension of the vectorized profile (k equals 33 in this dissertation). This analysis is closely related to the future analysis in terms of observable and unobservable components. That is, the true mean is always an optimum choice for the a priori mean. Note that the convex set of optimal a priori

mean profiles implicitly depends on the vertical covariance of the signal through the D matrix.

In all cases, the solution $\overline{T}_a =< \overline{T} >$ is always valid as might have been expected. Whenever the sensor noise has a full rank covariance matrix, it is the unique solution. When the noise covariance is not full rank, linear combinations of the observations can be found with no measurement noise associated with them. Hence, the transmission matrix $\mathcal{T} = DW$ has one of its eigenvalues equal to one, and the RHS of (V.2.12) is not full rank.

The optimal D matrix (for historical reasons, D stands for determination) is found by differentiation. Namely

$$\frac{\partial E}{\partial D} = 2W < (\overline{T}_a - \overline{T})(\overline{T}_a^t - \overline{T}^t) > W^t D^t - 2W < (\overline{T}_a - \overline{T})(\overline{T}_a^t - \overline{T}^t) > +2ND^t \qquad (V.2.13)$$

where N denotes the covariance matrix (diagonal) of the sensor noise. The optimal D matrix is thus

$$D = W < (\overline{T}_a - \overline{T})(\overline{T}_a^t - \overline{T}^t) > (W < (\overline{T}_a - \overline{T})(\overline{T}_a^t - \overline{T}^t) > W^t + N)^{-1} \qquad (V.2.14)$$

The optimal determination matrix is the product of the cross-covariance between temperature and brightness temperature by the inverse of the covariance matrix of brightness temperature with itself. Note that no assumptions were made about the shape of the weighting matrix (number of rows vs number of columns). The weighting matrix has to be full rank only if the sensing noise has no variance (perfect observations) or in certain cases of positive semidefinite noise covariance matrices. Under reasonable circumstances, a sufficient condition for the D matrix to exist is the pressure covariance matrix R is positive semi-definite. This is true, because in such a case, the eigenvalues of the product $W < (\overline{T}_a - \overline{T})(\overline{T}_a^t - \overline{T}^t) > W^t + N$ are all strictly positive, and that product admits an inverse. Because the partial derivative of the error can be recognized as the expectation of the inner product of

the error vector with the observations, the linear mean square filter is often derived from the "principle of orthogonality" which states that this inner product on the average is zero.

The disadvantage with such an interpretation of the LMMSE is that it does not allow for the incorporation of auxiliary information, or adaption of filters to changes in statistics and constraints.

If the a priori mean is chosen to be the expectation of the temperature field over the region of retrieval (because this was known in advance), the determination takes the familiar form

$$D = RW^t(WRW^t + N)^{-1} \qquad (V.2.15)$$

or using the Sherman-Woodbury formula (Athans and Schweppe, 1965)

$$D = (W^tN^{-1}W + R^{-1})^{-1}W^tN^{-1} \qquad (V.2.16)$$

The latter expression is valid only if the sensing noise covariance matrix is positive definite which is the case if the channels are independent of each other. It also requires the existence of the matrix R^{-1} which in turns requires the vertical covariance matrix to be positive definite.

Denoting by K_a the signal energy matrix $< (\overline{T}_a - \overline{T})(\overline{T}_a^t - \overline{T}^t) >$, the error becomes

$$\begin{aligned} E =& trace\{(I - K_aW^t(WK_aW^t + N)^{-1}W)K_a(I - W^t(WK_aW^t + N)^{-1}WK_a)\} \\ &+ trace\{K_aW^t(WK_aW^t + N)^{-1}N(WK_aW^t + N)^{-1}WK_a\} \\ =& trace\{(K_a - K_aW^t(WK_aW^t + N)^{-1}WK_a\} \\ =& trace\{(I - DW) < (\overline{T}_a - \overline{T})(\overline{T}_a^t - \overline{T}^t) >\} \\ =& trace\{(< (\overline{T}_a - \overline{T})(\overline{T}_a^t - \overline{T}^t) > (I - DW)\} \end{aligned} \qquad (V.2.17)$$

Using the Sherman-Woodbury formula it reduces to

$$E = trace\{(< (\overline{T}_a - \overline{T})(\overline{T}_a^t - \overline{T}^t) >^{-1} + W^tN^{-1}W)^{-1}\} \qquad (V.2.18)$$

This expression can now be optimized with respect to the weighting matrix W to further minimize the retrieval error. This problem is similar to the mathematical design of experiments. Several problems occur when doing so. First, the criterion is neither a linear nor a quadratic function of the weighting functions. Thus the optimization may not be analytically tractable. Second, the presence of noise in the sensing process leads to solutions where the signal-to-noise ratio of the sounding equation is boosted by having the optimal weighting matrix with large entries. Finally, the only control on the weighting functions is provided by the choice of frequencies and sounding angles, which is decidedly not a large factor. The physics of the atmosphere controls the shape of the weighting functions. The only control available to the user is the location of the peak of the weighting function. All things being equal, the retrieval error should be lower at those locations than at other locations which requires that the weighting functions be spaced vertically.

The bias error as defined by the expected value of the error vector is

$$\bar{b} = <\bar{e}> = (I - DW)(\overline{T}_a - <\overline{T}>) \qquad (V.2.19)$$

The LMMSE is thus in general a biased estimate unless the sample mean is known. The optimum a priori mean eliminates this bias error.

Let us point out that as the signal-to-noise ratio of the sounding process degrades, the determination matrix tends to zero, and the estimate becomes the a prior mean $\overline{T}_a$. In addition, the determination matrix computed using the covariance of the temperature field around its true mean $< (\overline{T} - <\overline{T}>)(\overline{T} - <\overline{T}>)^t >$ is not optimal when the a priori mean is not the true mean. The optimal matrix is thus found by taking the entire signal energy K_a. However, none of these quantities is known exactly.

The minimum information solution (Fleming and Smith, 1972, Fritz *et al* 1972) to the remote sensing problem is found by setting the vertical covariance matrix to a diagonal matrix (no interlevel correlation) and driving the variances to

infinity allowing them to stay equal, thus achieving an infinite signal-to-noise ratio.

$$D \rightarrow M = W^t(WW^t)^{-1} \qquad (V.2.20)$$

M is referred to as the Minimum Information Estimator (MIE). This solution does not involve any second order statistics about the temperature profiles (the mean has yet to be known) and is thus referred to as "physical". It is only defined for underspecified problems (number of pressure levels greater or equal than sensing channels) when the weighting matrix is full rank. The minimum information solution can also be reached from non probabilistic arguments by taking a minimum norm solution to the equation

$$\overline{T} = W\overline{T}_b \qquad (V.2.21)$$

which is only defined for undetermined problems (Rao, 1973). As usual, only the deviations of the two vectors from their respective means are considered. The minimum norm solution realizes the Moore Penrose pseudo inverse of the underspecified weighting matrix W.

Let us examine what happens as the vertical covariance matrix becomes infinitely diagonal in (V.2.16).

$$D \rightarrow F = (W^tN^{-1}W)^{-1}W^tN^{-1} \qquad (V.2.22)$$

This is what Schweppe (1976) refers to as the Fisher estimate when the temperature profile $\overline{T}$ is modeled as a complete unknown variable. This estimate is only valid for overspecified problems. One can take the Fisher estimate one step further by taking the noise as also being completely unknown. The noise covariance matrix becomes diagonal infinite and

$$F \rightarrow M' = (W^tW)^{-1}W^t \qquad (V.2.23)$$

which is the dual of the minimum information solution in the sense that WW^t is replaced by W^tW. M' is also the Moore Penrose pseudoinverse of the weighting matrix W.

The popularity of the Bayesian LMMSE can be appreciated by the extensive number of interpretations and variations on the theme. Three good reviews of these interpretations can be found in Schweppe (1976), Baumann (1980), and Fymat (1977). Among those one can name the regularization approach to the solution of the radiative transfer equation (Wahba, 1981, Tikhonov, 1977) where the mean square term plays the role of a "regularizing functional". One can also interpret the solution as the optimal merging of the physical solution with the pseudomeasurement provided by the mean (Rodgers, 1976) which is a method to go from Fisher to Bayesian estimation.

Other popular interpretations of the mean square solution to the retrieval problem include the maximum likelihood solution, the Gauss Markov least square theorem for correlated variables, the maximum a posteriori estimation for Gaussian variables, the generalized linear problem and the minimum norm solution under the extended inner product $x^t R^{-1} y$.

The principal advantage of the mean square estimation procedure over other types of inversion techniques is that it incorporates knowledge about the profiles derived from climatological arguments and can thus do well in estimating high frequency components of the temperature fields such as the tropopause inversion.

An interesting and somewhat rarely used variation of the LMMSE is that could be denoted as the "no a priori information" solution to the remote sensing problem. To the best of the author's knowledge, the following interpretation is novel. This no a priori information solution is an "intermediate" between the minimum information solution and the true mean square linear estimate in the sense that it does not require any a priori knowledge of the second order statistics of the temperature profiles.

The minimum information solution to the radiative transfer equation allows the mapping of an estimate of temperature profile to a new estimate of the same

profile according the relationship

$$\overline{T}' = W^t(WW^t)^{-1}W\hat{\overline{T}} \qquad (V.2.24)$$

The product of matrices $W^t(WW^t)^{-1}W$ is a projection operator satisfying $P^2 = P$. An interesting case of this mapping is obtained when the estimated profile is obtained from the traditional D matrix. Namely,

$$\begin{aligned}\overline{T}' &= W^t(WW^t)^{-1}WRW^t(WRW^t + N)^{-1}\overline{T}_b \\ &= \left(W^t(WW^t)^{-1}\right)\left(WRW^t\right)\left(WRW^t + N\right)^{-1}\overline{T}_b \qquad (V.2.25) \\ &= A\overline{T}_b\end{aligned}$$

This estimator is called the Non A Priori Information Estimator (NAPIE). The advantage of this retrieval operator is that, assuming that over the set of temperature profiles being estimated the sample covariance matrix of the sensor noise is nominal, all the elements of the retrieval operator can be computed. As will be demonstrated in the following section, this retrieval operator is the LMMSE estimator associated with a specific modeling of the vertical covariance matrix of the temperature field. Because the statistics of this retrieval operator are directly computed from the observations, this retrieval operator can readily be adapted to changes in statistics.

V.2.3 Usage of the Singular Value Decomposition of Weighting Matrices.

The purpose of this section is to express the different operators and quantities developed in the previous section in a form that separates the influence of the weighting matrices from that of the signal statistics. This task will be accomplished by considering the singular value decomposition of the weighting matrix.

Since remote sensing systems usually have fewer channels than parameters to estimate, this discussion will limit itself to weighting matrices with fewer rows than

columns. Furthermore, to avoid dealing with degenerate cases, these matrices will be taken to be full row rank. That is, no weighting function can be written as a linear combination of other weighting functions. This degenerate case could arise if two channel nominal operating frequencies were symmetric with respect to a resonance line of one of the atmosphere's absorbants. Such is the case for the two sidebands of the high altitude probing channels of AMSU. Under such circumstances, two bands present the same response. They are combined and treated as one single channel Lastly, the weighting matrices are assumed to be real.

Any $r \times N$ matrix W of rank r can be decomposed as (Ben-Israel and Greville, 1974)

$$W = \sum_{i=1}^{r} \sigma_i \overline{\psi}_i \overline{\varphi}_i^t \qquad (V.2.26)$$

where the vectors $\overline{\psi}_i$ are the orthonormal eigenvectors of the square matrix WW^t and $\overline{\varphi}_i$ are the eigenvectors of W^tW. The singular values σ_i are the positive square roots of the eigenvalues corresponding to $\overline{\psi}$ which coincide with the r first eigenvalues corresponding to r and weighting functions of channels $\overline{\varphi}$. Namely

$$\begin{aligned} WW^t\overline{\psi}_i &= \sigma_i^2 \overline{\psi}_i, \qquad i = 1, 2, \ldots, r \\ W^tW\overline{\varphi}_i &= \sigma_i^2 \overline{\varphi}_i, \qquad i = 1, 2, \ldots, r \end{aligned} \qquad (V.2.27)$$

For full rank matrices, all singular values are greater than zero. This decomposition is not unique because the $N - r$ vectors φ can be chosen arbitrarily. This is true because they all share the same eigenvalue 0.

The eigenvectors $\overline{\psi}$ and $\overline{\varphi}$ can be rearranged as two orthonormal matrices $\mathcal{W}_l$ and $\mathcal{W}_r$ where

$$\mathcal{W}_r = (\overline{\varphi}_1 \quad \overline{\varphi}_2 \quad \ldots \overline{\varphi}_N) \qquad (V.2.28)$$

and

$$\mathcal{W}_l = (\overline{\psi}_1 \quad \overline{\psi}_2 \quad \ldots \overline{\psi}_r) \qquad (V.2.29)$$

The weighting matrix can thus be expressed as the product of three matrices

$$W = \mathcal{W}_l \begin{pmatrix} \Sigma & 0 \end{pmatrix} \mathcal{W}_r^t \qquad (V.2.30)$$

where the matrix Σ is the diagonal matrix composed of the singular values σ_i and O a null $N - r$ by r matrix. The above decomposition is called the Singular Value Decomposition (SVD) and has been extensively used in regression theory and other aspects of signal processing (Tsitsiklis and Levy, 1981, Levy and Tsitsiklis, 1982). The singular values are generally stable with respect to changes in the matrix elements and thus do not require perfect knowledge of the weighting matrix (Fymat, 1977).

This decomposition of the weighting matrix is strongly related to the decomposition of the temperature profile onto the row space and kernel space of the observation matrices which are orthogonal to each other. For reasons that will soon become apparent, the row space of the observation matrix will be heretofore referred to as the observable space of the weighting matrix (sometimes also coined as the "seen" space of the satellite) and the kernel space of W as the unobservable space ("hidden" space of the satellite).

To show the relationship between observable and unobservable subspaces and the SVD, let us decompose the temperature profile using the $\overline{\varphi}$ basis.

$$\overline{T} = \sum_{i=1}^{N} < \overline{T}^t \overline{\varphi}_i > \overline{\varphi}_i \qquad (V.2.31)$$

The resulting brightness temperature is shown by plugging (V.2.31) into (V.2.1)

$$\begin{aligned} \overline{T}_b &= \sum_{i=1}^{r} \sigma_i \overline{\psi}_i \overline{\varphi}_i^t \sum_{j=1}^{N} < \overline{T}^t \overline{\varphi}_j > \overline{\varphi}_j \\ &= \sum_{i=1}^{r} \sum_{j=1}^{N} \sigma_i \overline{\psi}_i < \overline{\varphi}_i^t \overline{\varphi}_j > < \overline{T}^t \overline{\varphi}_j > \qquad (V.2.32) \\ &= \sum_{i=1}^{r} \sigma_i \overline{\psi}_i < \overline{T}^t \overline{\varphi}_i > \end{aligned}$$

This equation demonstrates that only the first r components of the brightness temperature expressed only the $\overline{\varphi}$ basis contribute to the brightness temperature. Hence, it belongs to the row space of the weighting matrix. Hence, the observable subspace of the matrix W is spanned by the first r rows of the matrix $\mathcal{W}_r$. The brightness temperature vector resulting from the observed temperature profile is nothing else than the observable components of the temperature profile operated on by a full rank square matrix Ω. Namely

$$\overline{T}_b = \sum_{i=1}^{r} \sigma_i \overline{\psi}_i < \overline{T}^t \overline{\varphi}_i >= \Omega \overline{T}_u \qquad (V.2.33)$$

where the matrix Ω is a reduced version of the original weighting matrix projected onto the observable subspace. That is,

$$\Omega = \sum_{i=1}^{r} \sigma_i \overline{\psi}_i \qquad (V.2.34)$$

The remaining $N - r$ vectors $\overline{\varphi}$ span the kernel of the weighting matrix and form the basis for the unobservable space of W.

From the above analysis, it is apparent that the retrieval problem can rephrased in the following context: Given the measurement of the observable components of the temperature profile operated on by a full rank matrix and corrupted by additive noise, what is the best estimate of the components of the temperature profile in the observable and in the unobservable subspaces of the weighting matrices?

Before answering this question in the context of the single-spot estimation problem, the construction of the SVD basis of the weighting matrix should be considered.

The software package LINPACK (Dongarra *et al.*, 1979) provides a series of routines that computes the singular value decomposition. Unfortunately, they do not provide any control on the basis vectors of the unobservable subspace of W

which is of dimension $N - r$ and can be thus be generated by an infinite number of basis vectors. This is a problem when used in the context of remote sensing because the shape of these vectors is relevant to the analysis of the retrieval process.

The basis vectors for the SVD basis (forming the matrix $\mathcal{W}_r$) are computed using the Gram-Schmidt orthogonalization algorithm. This algorithm starts from a set of linearly independent vectors and computes an orthonormal basis. It does not provide total control for the shape of the vector φ but leaves some freedom by choosing the initially guessed vectors.

The first r vectors (3 for MSU) of $\mathcal{W}_r$ are chosen to span the row space of the weighting matrix and are simply taken as the row of the matrix. To make this construction canonical, the row pertaining to the lowest probing channel of the atmosphere (as measured by the largest entry) is written as the first line, the next to the lowest probing channel is the second line, etc. More freedom is possible in terms of the initial vectors in the unobservable subspace. The only condition to be satisfied is that these vectors be independent of each other.

One possibility is to take the initial vectors to be identically zero except at a specific pressure level where the vector's component is set to one. The problem with this choice is that it is close to the natural basis representation and is thus misleading. A better choice is to take a series of vertical sine wave which vary with increasing frequency; this frequency can be made proportional to the vector index. Such vectors are orthogonal to each other. Thus, one can expect the Gram-Schmidt orthonormalization to leave them relatively unchanged as their index increases.

There are two advantages to this choice: It would allow a bridge between this analysis and the one based on the eigenvectors of the vertical covariance kernel (Smith and Woolf, 1976). This is true because the eigenvectors of such matrices tend to oscillate with increasing frequency (not necessarily in a sinusoidal fashion) as their index grows. Hence, the high index unobservable basis vectors will be closer

(in an inner product sense) to the high index eigenvectors of the vertical covariance kernel than to the lower index ones. The second reason, of less importance, pertains to the modal analysis of specific implementations of Kalman filtering of temperature fields which are analyzed in Appendix V.A.

In summary, the initial vectors for the Gram-Schmidt algorithm are taken to be

$$\begin{aligned} \hat{\varphi}_{i(j)} &= w(i,j), \qquad for\ i = 1,\ldots,r, j = 1,\ldots,N \\ \hat{\varphi}_{i(j)} &= \sin(\frac{2\pi ij}{N}), \qquad for\ i = r+1,\ldots,N, j = 1,\ldots,N \end{aligned} \tag{V.2.35}$$

The remaining part of the singular value decomposition ($\mathcal{W}_l$) first applies the matrix $\mathcal{W}_r^t$ to the left of W to obtain Ω

$$\Omega = W\mathcal{W}_r^t = \mathcal{W}_l \begin{pmatrix} \Sigma & 0 \end{pmatrix} \mathcal{W}_r \mathcal{W}_r^t = \mathcal{W}_l \begin{pmatrix} \Sigma & 0 \end{pmatrix} \tag{V.2.36}$$

then by calling the LINPACK routine SVVDC to obtain $\mathcal{W}_l$.

For this analysis, the temperature profiles were represented by 33×1 vectors ($N = 33$). In the natural representation, the components were the temperatures at the pressure levels introduced in Chapter IV (1000, 950, 900, 850, 750, 700, 650, 600, 550, 500, 450, 400, 350, 300, 250, 200, 175, 150, 125, 100, 80, 70, 60, 50, 40, 30, 25, 20, 15, 10, 7, and 5 *mbar*).

Figure V.2.3 presents the basis functions of the observable components of the temperature profiles for the Microwave Sounding System at three different sounding angles. Note that the difference between the basis vectors is not as important between nadir and 32^o as it is between 32^o and 56^o. These bases are computed for weighting matrices computed over a summer atmosphere and above ground.

Figures V.2.3 and V.2.5 present the 30 basis vectors of the unobservable subspace for the same conditions at nadir. Note that as the index of4the basis vector increases, so does the frequency of oscillation.

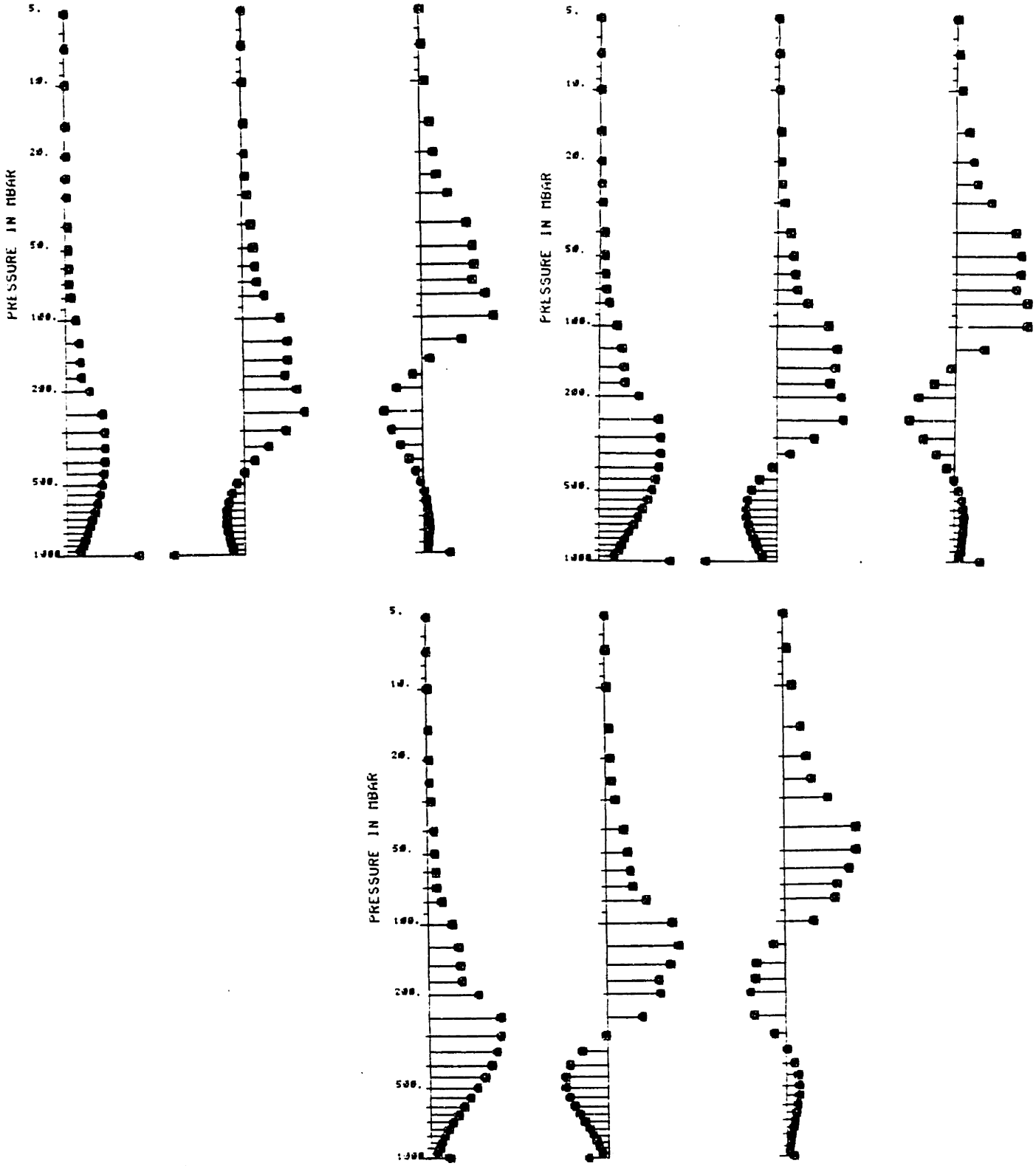


Figure V.2.3: observable components for the MSU sounder over land for three different sounding angles (nadir, 32°, and 56°)

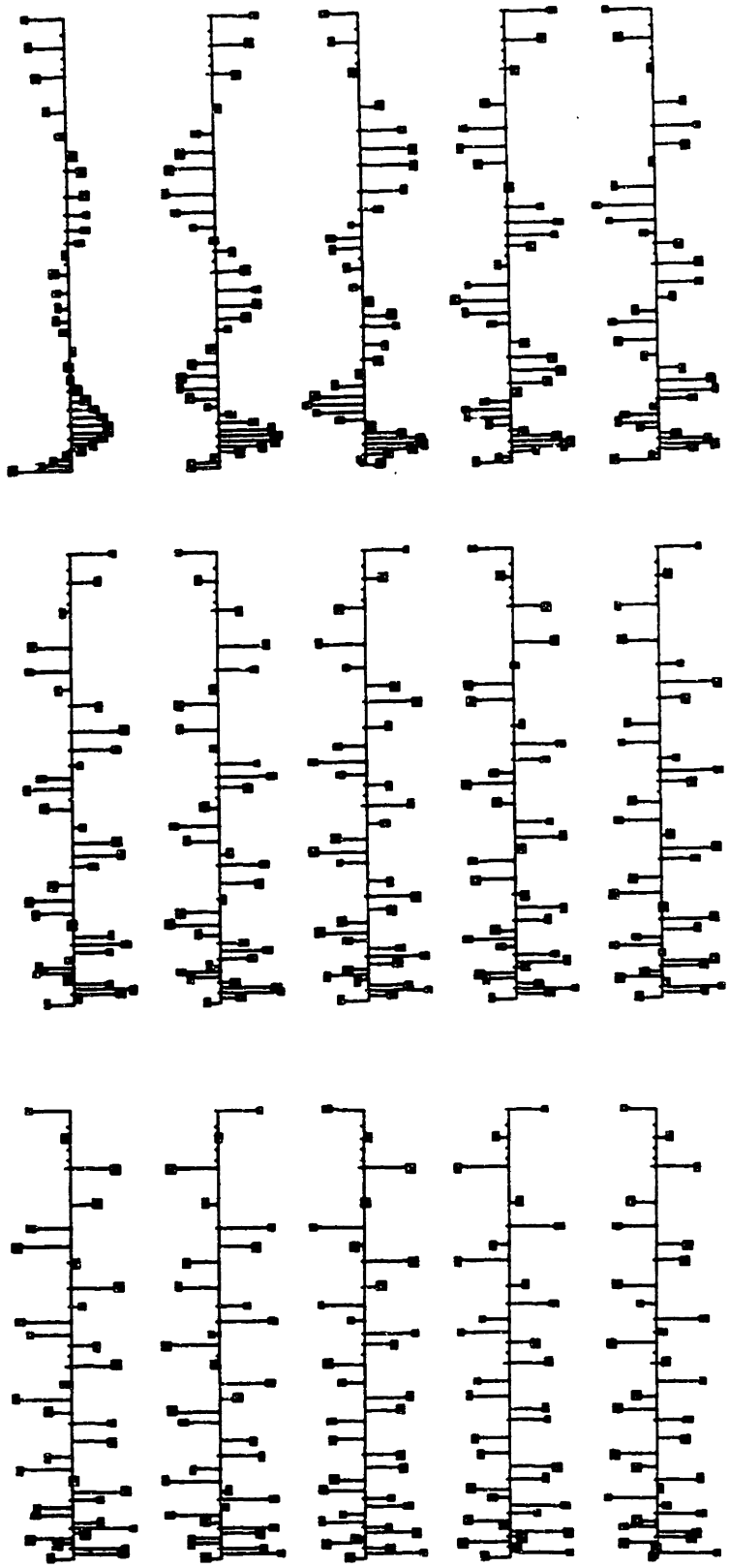


Figure V.2.4: 15 first unobservable basis vectors for the MSU sounder over land at nadir

The addition of more sounding channels can dramatically change the type

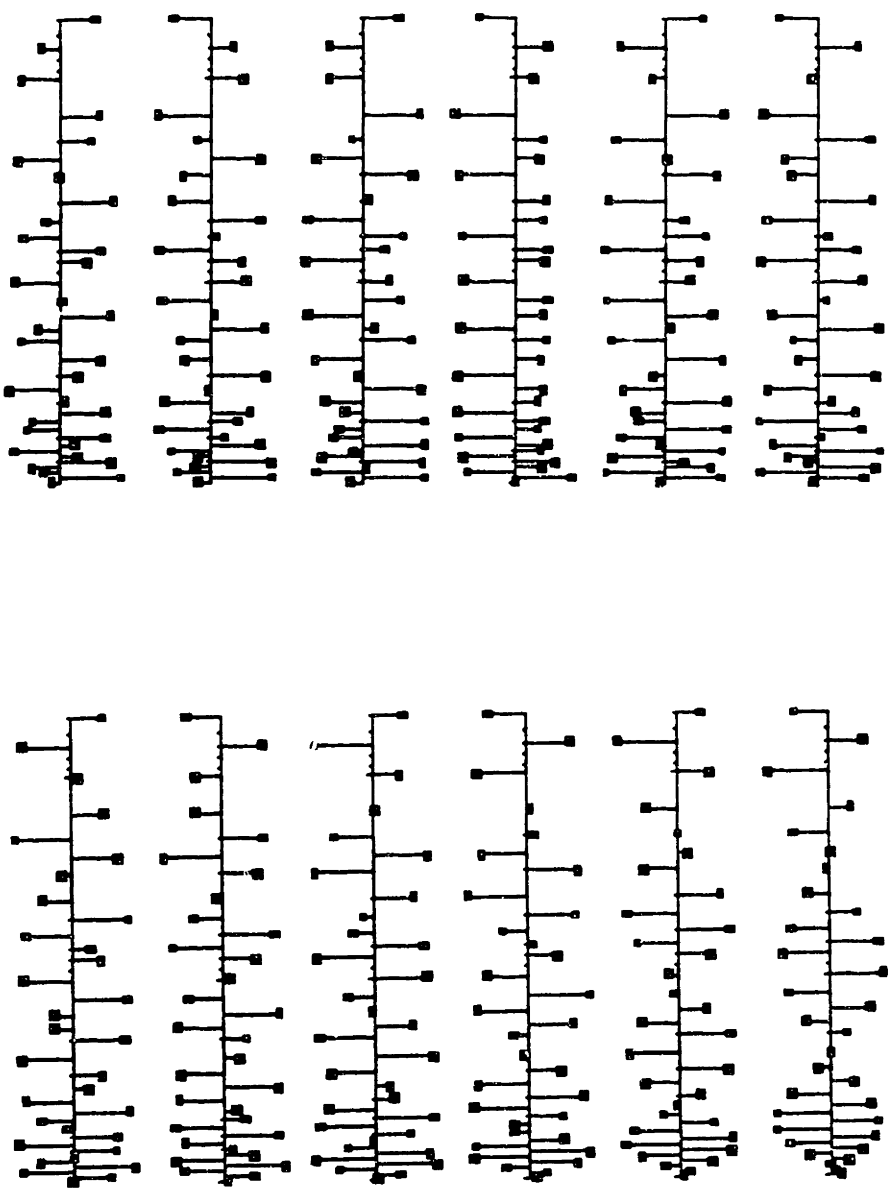


Figure V.2.5: 15 last unobservable vectors for the MSU sounder over land at nadir

of temperature profiles that belong to the observable space. To illustrate this fact,

let us consider the Advanced Microwave Sounding Unit with its 12 temperature sounding channels. Its weighting functions at nadir were presented in Figure II.4.1. Note that the highest pressure level used in the discretization of the atmosphere (5 *mbar*) is not appropriate for use with the highest probing channels of AMSU because they do probe at higher altitudes. Thus the following plot of these vectors may be misleading. The weighting functions are computed at nadir over water.

The singular values σ_i can be thought of as the gains of the weighting matrices, sometimes referred to as statistical weights (Zacks, 1981), when the channel responses are orthogonal to each other. In turn, they will determine (along with the a-priori statistics) the signal-to-noise ratio of the sounding process. Table V.2.1 presents the singular values for the MSU system as a function of the sounding angle (six different positions from nadir to 56°). The corresponding weighting functions are computed over land using the mean temperature profile over the United States over July 1979. The index of the singular value follows the rule established when implementing the Gram-Schmidt algorithm.

One can now express the LMMSE and its limiting forms by projecting the brightness temperature and the physical temperature profile into their respective SVD basis representation. The components of the temperature profile in the observable space are denoted by a subscript o whereas components in the orthogonal complement subspace (unobservable) is denoted by a u.

The determination matrix becomes

$$\begin{aligned} D \rightarrow D' &= \mathcal{W}_r D \mathcal{W}_l^t = \mathcal{W}_r^t R W^t (W R W^+ N)^{-1} \mathcal{W}_l^t \\ &= R \mathcal{W}_r \begin{pmatrix} \Sigma \\ 0 \end{pmatrix} \mathcal{W}_l^t (\mathcal{W}_l (\Sigma \quad 0) \mathcal{W}_r^t R \mathcal{W}_r \begin{pmatrix} \Sigma \\ 0 \end{pmatrix} \mathcal{W}_l^t + N)^{-1} \mathcal{W}_l^t \end{aligned} \qquad (V.2.37)$$

Using the projection of the covariance matrix onto observable and unobservable subspaces yields

$$D' = \begin{pmatrix} R_{oo}\Sigma \\ R_{uo}\Sigma \end{pmatrix} (\Sigma R_{oo} \Sigma + \mathcal{W}_l^t N \mathcal{W}_l)^{-1} \qquad (V.2.38)$$

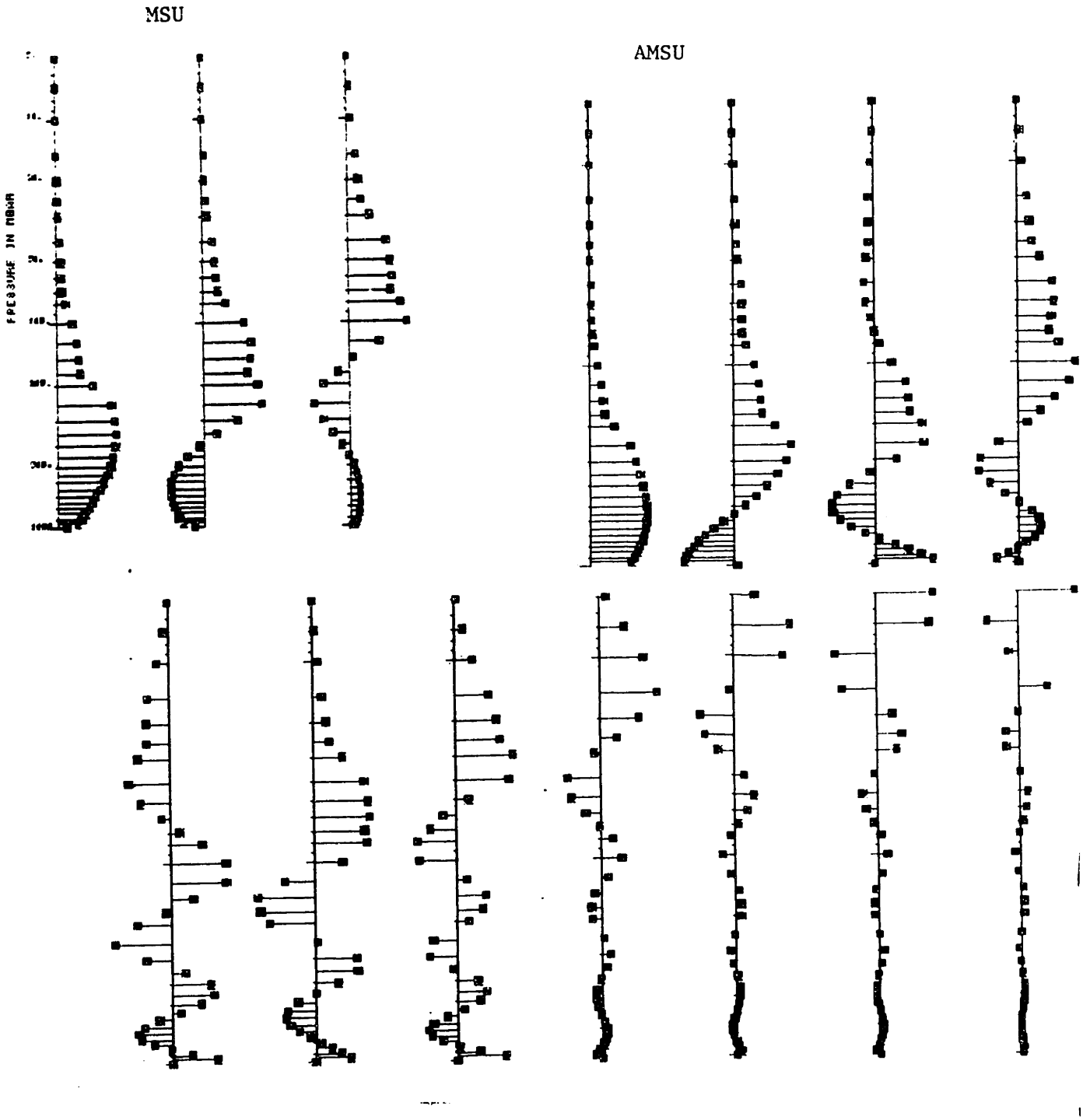


Figure V.2.6: observable components for MSU and AMSU over water at nadir

Let us rotate the sensor noise covariance matrix by the same matrix as the brightness

```
angle:  0.0
    0.3634318
    0.2460124
    0.0940057
angle: 10.7
    0.3640926
    0.2466679
    0.0947389
angle: 21.5
    0.3661347
    0.2486389
    0.0969694
angle: 32.5
    0.3694646
    0.2516959
    0.1005039
angle: 43.9
    0.3751078
    0.2565775
    0.1062978
angle: 56.2
    0.3834749
    0.2638011
    0.1148913
```

Table V.2.1: Singular Values for MSU over land for the 6 different sounding angles of the instrument

temperature, thus introducing $N' = \mathcal{W}_l^t N \mathcal{W}_l$. The D matrix takes the form

$$D' = \begin{pmatrix} R_{oo}(\Sigma R_{oo}\Sigma + N')^{-1} \\ R_{uo}(\Sigma R_{oo}\Sigma + N')^{-1} \end{pmatrix} \qquad (V.2.39)$$

This form involves the temperature and noise statistics from an observational point of view. A slightly different expression can be found that is more "signal oriented" by factoring the diagonal matrix Σ out. Hence

$$D' = \begin{pmatrix} R_{oo}(R_{oo} + \Sigma^{-1}N'\Sigma^{-1})^{-1}\Sigma^{-1} \\ R_{uo}(R_{oo} + \Sigma^{-1}N'\Sigma^{-1})^{-1}\Sigma^{-1} \end{pmatrix} \qquad (V.2.40)$$

The scaled noise covariance matrix $\Sigma^{-1}N'\Sigma^{-1}$ corresponds to the sensor's noise normalized to the temperature profile rather than to the brightness temperature.

Several keys observations can be made on the form of the optimal LMMSE (at least its linear part). The first one is that the estimation of observable and unobservable components of the temperature profile are two separate operations. The second observation is that the relevant statistical description of temperature profiles (vertical covariance kernel) does not require any knowledge of the correlation of the unobservable components with themselves (R_{uu}).

The third observation is that the solution to the remote sensing problem can be divided into a two-step estimation problem. To demonstrate this, express the estimates of the relevant projections of the temperature profile as a function of the rotated observed brightness temperature.

$$\begin{aligned}\hat{\overline{T}}_o &= R_{oo}(R_{oo}+N')^{-1}\overline{T}_b' \\ \hat{\overline{T}}_u &= R_{uo}(R_{oo}+N')^{-1}\overline{T}_b' = R_{uo}R_{oo}^{-1}\hat{\overline{T}}_o\end{aligned} \qquad (V.2.41)$$

This set of equations expresses that the optimal estimate of the unobservable components of the temperature profile based on the brightness temperature is also the optimal estimate of the unobservable components based on the observable components themselves which were optimally derived from the brightness temperature (optimality is conserved). Moreover, it shows that from a mean square error point of view, once the estimate of the observable components is found, the remainder of the estimation process has been entirely determined.

This last remark, combined with the fact that the matrix Σ is full rank, allows us to design new type of estimates based on mixed criteria for observable and unobservable components. This analysis is performed in Chapter VIII.

This separability in the estimates can also be used to correct for non-linearities in the radiative transfer function. As was shown in Chapter II, the absorption coefficients of O_2 and H_2O and the weighting matrices depend on the

complete temperature profile, hence on all its components once the temperature profile has been discretized in vector form. This implies that the non linearities and any attempt to correct for them must include observable and unobservable components. In contrast, when applying a linear estimator to retrieve the temperature profile, the estimate of the observable components determines the remainder of the retrieval process. Thus, one must characterize the impact of changes in the observable components on the weighting matrices caused by non-linearities of the radiative transfer equation with more care than those that are caused by changes in the unobservable components.

This is especially true when dealing with iterative retrieval methods such as Chahine's (1970) which iteratively solves the forward problem (estimate of absorption coefficient, weighting matrices, and brightness temperature to be matched with the observed ones.) and the inverse problem (estimate of changes in the temperature profile caused by differences between observed and computed radiances).

The limiting forms of the D matrix are the minimum information estimator (MIE) and the non-a-priori information estimator (NAPIE). The former is found by noting that a diagonal matrix with constant entries is diagonal regardless of the representational basis. Thus

$$R_{oo} = \begin{pmatrix} \infty & 0 & \cdots & 0 \\ 0 & \infty & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \infty \end{pmatrix} \qquad R_{uo} = 0 \qquad (V.2.42)$$

Moreover, the retrieval operator becomes

$$D^{n} = \begin{pmatrix} \Sigma^{-1} \\ 0 \end{pmatrix} \qquad (V.2.43)$$

The MIE reconstructs the observable components as if the sensing process has an infinite signal-to-noise ratio and sets the unobservable components to the apriori mean $\overline{T}_a$. This solution is sensitive to measurement noise. This sensitivity is large

in cases where the number of degrees of freedom in the temperature field being retrieved is so small that the deviation of the temperature profiles from their mean (power in the observable components) is small compared to the sensing noise variance.

The NAPIE goes beyond the MIE in estimating the observable components from the observation. In the SVD basis, the NAPIE retrieval operator becomes

$$D^{*\prime} = \begin{pmatrix} R_{oo}\Sigma \\ 0 \end{pmatrix} (\Sigma R_{oo}\Sigma + N')^{-1} \qquad (V.2.44)$$

The no-a-priori information solution estimates the observable components as the LMMSE and the unobservable components as the minimum information solution. It corresponds to the LMMSE solution when the cross-covariance between unobservable and observable components is zero. The no-a-priori information solution is also related, as it will be explained in Chapter VIII, to retrieval operators based upon mixed estimation theory.

The transfer function which corresponds to the retrieval process is found by multiplying the retrieval operator by the weighting matrix yielding

$$\mathcal{T} = \begin{pmatrix} R_{oo}(R_{oo} + \Sigma^{-1}N'\Sigma^{-1})^{-1} & 0 \\ R_{uo}(R_{oo} + \Sigma^{-1}N'\Sigma^{-1})^{-1} & 0 \end{pmatrix} \qquad (V.2.45)$$

Note that the eigenvalues of such a transfer function are less than one and, for $N-r$ of them equal to zero.

For a noiseless instrument the transfer function for the MIE which is equal to the one of the NAPIE is

$$\mathcal{W}_r^t W^t (WW^t)^{-1} W \mathcal{W}_r = \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} \qquad (V.2.46)$$

The transfer function for the LMMSE is

$$\mathcal{W}_r^t R W^t (WRW^t)^{-1} W \mathcal{W}_r = \begin{pmatrix} I & 0 \\ R_{uo}R_{oo}^{-1} & 0 \end{pmatrix} \qquad (V.2.47)$$

This restates the fact that the MIE reconstructs the observable components and ignores the components in the unobservable subspace while the LMMSE uses the second order statistics of the signal to reduce the error in the unobservable components.

Finally, the retrieval error is calculated, both in term of bias and rms error. The bias error is found by projecting (V.2.18) onto the SVD basis. Hence

$$\bar{b}' = \begin{pmatrix} \Sigma^{-1}N'\Sigma^{-1}(R_{oo} + \Sigma^{-1}N'\Sigma^{-1})^{-1}(\overline{T}_{au} - < \overline{T}_u >) \\ (\overline{T}_{au} - < \overline{T}_u >) - R_{uo}(R_{oo} + \Sigma^{-1}N'\Sigma^{-1})^{-1}(\overline{T}_{ao} - < \overline{T}_o >) \end{pmatrix} \quad (V.2.48)$$

The error in the observable components depends upon the matricial signal to noise ratio of the retrieval process, whereas the error in the unobservable components depends on the correlation between observable and unobservable components.

The mean square error follows the same pattern. Namely,

$$\begin{aligned} E =& trace\{\Sigma^{-1}N'\Sigma^{-1}(R_{oo} + \Sigma^{-1}N'\Sigma^{-1})^{-1}R_{oo}\} \\ &+ trace\{R_{uu} - R_{uo}(R_{oo} + \Sigma^{-1}N'\Sigma^{-1})^{-1}R_{ou}\} \end{aligned} \quad (V.2.49)$$

The error in the observable components goes to zero (perfect estimation) when the signal-to-noise ratio tends toward infinity as is expected. The error in the unobservable components follows the behavior associated with that of mean square estimation. The signal energy in the unobservable components R_{uu} is reduced by that amount which can be estimated from the observable components.

This analysis shows that the retrieval error depends upon several key factors: the number of channels and weighting functions which determine the ratio of power in the observable and unobservable components, the signal to noise ratio which determines the error in the observable components, and the degree of correlation between observable and unobservable components which controls the retrieval error for the unobservable components.

The contribution to the error caused by the sensing noise is important to quantify since it provides a mean for accessing the impact of the noise specifications of a sounding system onto the estimation error.

The error in the observable components is solely caused by imperfect observations. For the unobservable components, the impact of the sensors is calculated by subtracting the error from perfect sensors to the actual retrieval error. Hence

$$\begin{aligned} E_{noise} =& trace\{\Sigma^{-1}N'\Sigma^{-1}(R_{oo}+\Sigma^{-1}N'\Sigma^{-1})^{-1}R_{oo}\} \\ &+ trace\{R_{uo}R_{oo}^{-1}N'(R_{oo}+\Sigma^{-1}N'\Sigma^{-1})^{-1}R_{ou}\} \end{aligned} \qquad (V.2.50)$$

Because the weighting functions of any sounding system (MSU in particular) change with the sounding angle, so do the weighting functions. To quantify the changes in the SVD basis, one must introduce a measure of how parallel in the different basis functions as a function of sounding angle. The inner product of the basis vectors at one angle with the ones at another angle provides such a measure. Strictly speaking, the absolute value of the inner product is considered. Table V.2.2 presents the matrix formed of the inner products for the MSU system observing at nadir and 56°.

Finally, consider the bound on the estimation error as the observing condition become closer to ideal (large signal-to-noise ratio, large degree of correlation between components of the temperature profile). For the observable components, the retrieval error bound for a given sounding platform is

$$E_{omin} = trace\{N'\} \qquad (V.2.51)$$

which corresponds to a noise floor. For the unobservable components, the bound is

$$E_{umin} = trace\{R_{uu} - R_{uo}R_{oo}^{-1}R_{ou}\} \qquad (V.2.52)$$

This latter expression involves the sounding system's characteristics only in the form of the observable and unobservable components.

V.3. Multidimensional Retrievals

.85	.48	.08	.12	.07	.09	.02	.05	.05	.01	.01	.02	.00	.00	.00	.00	.00	.01	.01	.01	.01	.01	.01	.01	.00	.00	.00	.00	.00	.00	.00	.00	.00
.20	.63	.44	.20	.31	.29	.04	.26	.26	.00	.02	.05	.01	.03	.01	.01	.01	.03	.03	.02	.01	.01	.01	.05	.03	.00	.03	.00	.00	.00	.01	.00	.00
.06	.13	.77	.33	.07	.04	.06	.25	.31	.06	.03	.03	.13	.09	.09	.07	.06	.10	.04	.01	.07	.10	.12	.10	.06	.02	.03	.02	.01	.01	.01	.00	.00
.09	.02	.16	.82	.49	.01	.05	.13	.10	.01	.03	.03	.05	.04	.04	.01	.02	.03	.01	.00	.04	.06	.05	.03	.04	.01	.01	.01	.00	.00	.01	.00	.00
.07	.02	.21	.12	.40	.56	.13	.52	.18	.17	.10	.12	.07	.12	.07	.01	.00	.06	.03	.02	.11	.15	.10	.01	.12	.02	.06	.00	.01	.01	.02	.00	.00
.14	.18	.22	.25	.38	.01	.37	.13	.24	.20	.09	.23	.26	.15	.20	.10	.10	.21	.10	.03	.16	.23	.24	.19	.07	.00	.00	.00	.00	.00	.01	.00	.00
.02	.02	.03	.03	.08	.10	.90	.08	.17	.08	.13	.14	.14	.06	.09	.06	.08	.10	.05	.01	.07	.10	.12	.11	.01	.00	.04	.02	.01	.01	.02	.00	.01
.18	.23	.12	.11	.25	.56	.03	.47	.19	.33	.00	.23	.12	.03	.06	.06	.03	.03	.05	.08	.12	.11	.13	.14	.04	.01	.04	.00	.01	.01	.03	.00	.00
.02	.03	.20	.08	.14	.08	.02	.33	.54	.15	.07	.20	.22	.10	.01	.25	.15	.03	.11	.10	.10	.02	.16	.35	.24	.05	.24	.09	.06	.06	.02	.00	.01
.12	.17	.02	.06	.16	.42	.09	.07	.32	.76	.14	.04	.03	.03	.03	.10	.08	.03	.00	.01	.04	.07	.01	.04	.07	.04	.01	.00	.01	.01	.00	.01	.01
.00	.01	.03	.10	.24	.25	.01	.19	.29	.29	.29	.68	.02	.13	.06	.06	.04	.05	.07	.05	.00	.05	.00	.14	.20	.04	.12	.03	.03	.03	.02	.01	.01
.09	.11	.07	.05	.13	.06	.06	.01	.12	.14	.42	.08	.05	.09	.11	.12	.15	.06	.05	.18	.28	.19	.14	.36	.39	.34	.26	.04	.06	.03	.16	.00	.01
.01	.01	.00	.00	.01	.00	.02	.02	.02	.02	.10	.01	.88	.07	.13	.13	.13	.15	.06	.03	.05	.08	.16	.28	.01	.04	.15	.04	.04	.03	.01	.00	.01
.02	.02	.00	.01	.01	.01	.04	.11	.03	.10	.18	.07	.03	.78	.20	.06	.00	.18	.15	.08	.11	.24	.15	.09	.29	.05	.17	.05	.04	.04	.04	.00	.00
.02	.02	.01	.01	.03	.03	.06	.12	.06	.14	.27	.07	.06	.22	.79	.06	.10	.17	.15	.15	.11	.03	.04	.20	.19	.11	.10	.01	.02	.00	.04	.01	.00
.08	.11	.05	.05	.11	.08	.07	.12	.15	.20	.06	.24	.09	.27	.28	.71	.03	.12	.03	.03	.19	.22	.18	.02	.03	.06	.10	.01	.02	.01	.04	.00	.00
.01	.01	.01	.05	.10	.08	.05	.05	.06	.04	.05	.05	.08	.05	.15	.25	.90	.03	.04	.06	.14	.14	.11	.01	.02	.02	.04	.01	.01	.00	.01	.00	.00
.01	.02	.00	.06	.09	.02	.01	.03	.00	.02	.01	.09	.06	.07	.11	.03	.10	.90	.05	.00	.16	.21	.21	.10	.12	.01	.02	.01	.00	.00	.01	.00	.00
.05	.06	.04	.03	.04	.01	.01	.04	.08	.05	.06	.23	.03	.16	.06	.14	.01	.06	.92	.01	.06	.06	.08	.10	.05	.03	.05	.02	.02	.02	.01	.00	.00
.07	.08	.04	.00	.02	.06	.00	.03	.10	.00	.23	.17	.01	.03	.12	.00	.03	.06	.13	.91	.02	.01	.02	.08	.08	.06	.06	.01	.02	.02	.02	.00	.00
.09	.11	.05	.04	.07	.05	.00	.08	.11	.02	.30	.14	.00	.05	.13	.07	.04	.06	.09	.16	.81	.24	.21	.01	.04	.07	.11	.00	.03	.02	.05	.01	.00
.13	.16	.05	.08	.15	.05	.01	.22	.13	.05	.27	.09	.04	.24	.12	.25	.11	.05	.05	.12	.14	.71	.26	.07	.05	.09	.05	.03	.01	.00	.04	.00	.00
.18	.22	.06	.10	.19	.07	.02	.17	.18	.01	.23	.01	.07	.14	.02	.02	.03	.02	.04	.15	.15	.19	.75	.17	.08	.11	.10	.06	.01	.00	.09	.00	.00
.23	.30	.08	.12	.25	.06	.06	.13	.21	.11	.14	.02	.11	.00	.17	.28	.17	.06	.08	.05	.02	.01	.10	.64	.06	.12	.24	.05	.07	.07	.08	.01	.01
.11	.13	.00	.04	.11	.04	.04	.12	.12	.06	.34	.23	.08	.06	.16	.10	.04	.04	.03	.07	.11	.02	.04	.18	.55	.23	.50	.11	.11	.12	.06	.04	.02
.05	.05	.00	.00	.04	.03	.04	.01	.06	.03	.31	.09	.05	.05	.12	.17	.07	.04	.02	.07	.09	.10	.05	.11	.23	.83	.14	.13	.04	.03	.10	.01	.01
.01	.02	.01	.00	.00	.04	.01	.05	.01	.01	.13	.12	.00	.05	.02	.06	.05	.00	.05	.05	.06	.04	.01	.01	.10	.12	.47	.02	.32	.31	.69	.10	.04
.00	.01	.01	.00	.01	.03	.01	.01	.05	.02	.14	.10	.01	.03	.03	.04	.03	.01	.06	.08	.08	.01	.01	.01	.02	.08	.09	.96	.06	.06	.00	.04	.02
.00	.01	.02	.01	.01	.01	.03	.13	.01	.02	.01	.24	.00	.20	.04	.25	.12	.02	.15	.05	.05	.22	.05	.03	.40	.18	.41	.05	.26	.27	.48	.02	.04
.00	.00	.01	.01	.01	.01	.01	.05	.00	.02	.01	.08	.00	.08	.02	.10	.05	.01	.05	.01	.01	.09	.02	.02	.16	.08	.16	.01	.61	.69	.21	.01	.06
.00	.00	.01	.00	.01	.02	.02	.05	.03	.02	.09	.12	.01	.07	.02	.08	.06	.00	.07	.06	.06	.06	.01	.00	.13	.10	.05	.14	.62	.55	.42	.09	.16
.00	.00	.00	.00	.00	.00	.00	.00	.00	.01	.00	.01	.00	.01	.01	.02	.01	.00	.01	.00	.00	.01	.00	.01	.01	.03	.00	.02	.15	.12	.04	.90	.40
.00	.00	.00	.00	.00	.00	.00	.01	.00	.00	.01	.02	.00	.02	.00	.02	.01	.00	.01	.01	.01	.02	.00	.00	.03	.01	.04	.01	.09	.04	.02	.42	.90

Table V.2.2: inner product matrix between the observable and unobservable components of the MSU at nadir and 56°

In this section, we examine the impact of incorporating several sounding locations in the retrieval process both for the natural representation of temperature profile (user oriented) and the SVD representation (system oriented).

V.3.1 Sampling Pattern and Implications

Figure V.3.1 shows the scanning pattern of the TIROS N (MSU) sounding system which is typical of polar orbit sampling patterns. As the satellite flies along its track, its antennas sweep in the direction perpendicular to it. As the scanning or observation angle varies, so does the distance between the samples. The temperature field is convolved with a wider projected antenna pattern which in effect low pass filters the field. Also, the sounding angle used in computing the brightness temperature from the temperature field varies.

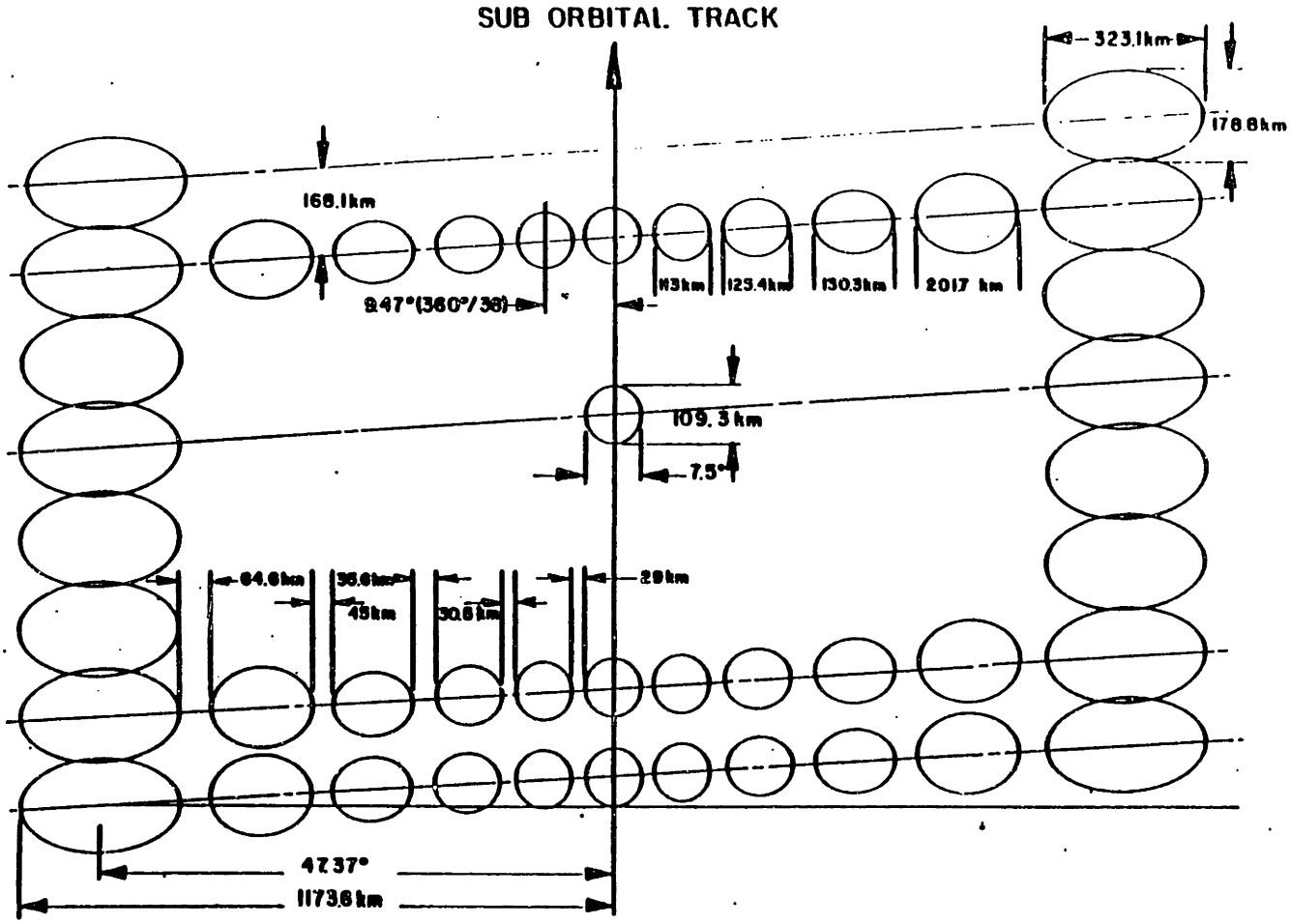
Figure V.3.2 presents the sounding geometry and the different angles associated with the computation of the radiative transfer function where α is the observation angle at the satellite, β the latitude of the sampling point (referred to nadir) and γ the sounding angle (used in the computation of the weighting functions). Let a denote the radius of the earth (6400 km) and d the altitude of the satellite (about 840 km). Simple trigonometry gives

$$\tan\alpha = \frac{a\sin\beta}{d + a(1-\cos\beta)} \qquad (V.3.1)$$

Trigonometry provides the inverse relationship between the observation and the sounding angles

$$\gamma = \sin^{-1}\left(\frac{a+d}{a}\sin\alpha\right) \qquad (V.3.2)$$

As stated previously, as a sounder's scanning angle varies so do the weighting functions associated with the different probing channels. As the scanning angle departs from nadir, the optical path associated with the observation lengthens. This



TIROS OPERATIONAL VERTICAL SOUNDER
MICROWAVE SOUNDING UNIT SCAN PATTERN PROJECTED ON EARTH
(833.4 km CIRCULAR ORBIT)

Figure V.3.1: Sounding Pattern of the Tiros N projected on earth

lengthening effectively increases the absorption of the atmosphere constituents and raises the altitude at which the weighting function peak.

These changes in the weighting matrices cause problems in the implementa-

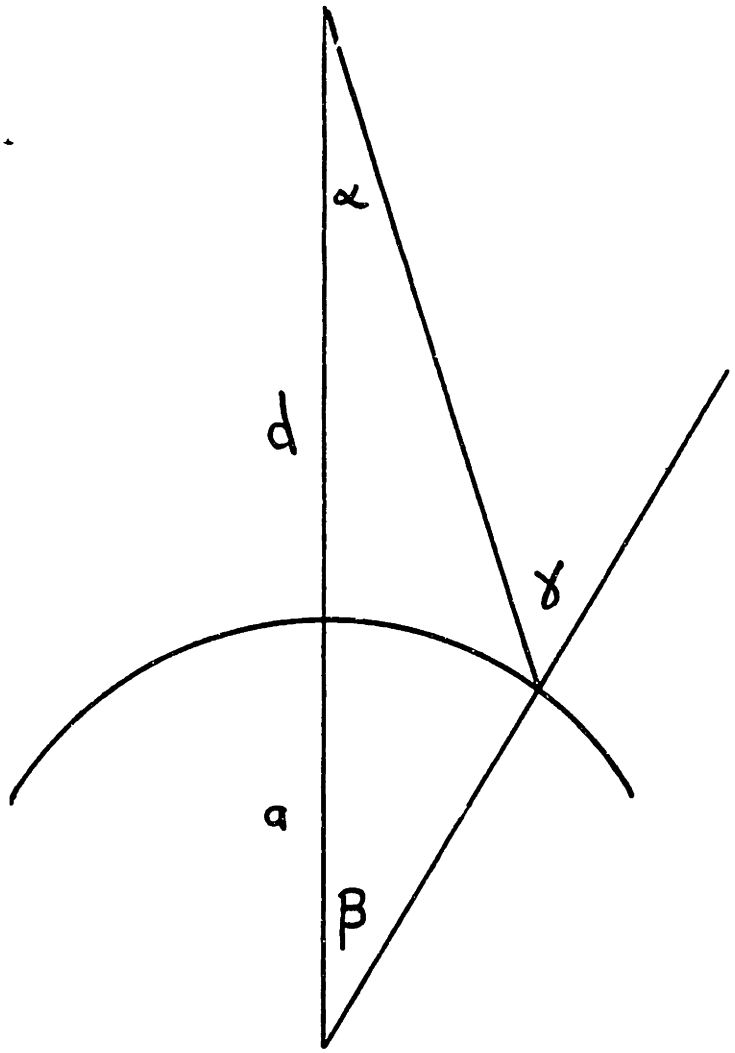


Figure V.3.2: Sounding geometry of low orbit sensing platform

tion of multidimensional recursions schemes, in that they prohibit the factorization of all the observations considered in the recursion, of the brightness temperature field using one common weighting matrix. These differences between weighting matrices Klyim that observations at nadir and at the extreme scanning angle (and any angle in between) cannot be treated as samples of the brightness temperature field (related to the physical temperature field through the radiative transfer equation), but as samples of different brightness temperature fields.

These considerations yield overinflated algorithms for the design of retrieval operators and prevent an easy understanding of the benefits of multidimensional inversion schemes.

One possible solution to this problem is known as limb correction (Westwater *et al*, 1985). The idea behind limb correction is to evaluate for scanning angles different from nadir what would have been the observed temperature, should the sounding be performed at nadir. Heretofore, the corrected brightness temperature can be thought as samples of the same brightness temperature field and the weighting functions can be assumed to be the same for all observations.

The limb correction scheme is basically a two-step linear estimator. The first step consists of the estimation of the temperature profile that gives rise to the brightness temperature at a given sensing angle. The second step consists of the estimation of the corresponding brightness temperature at nadir. To avoid relying on any a priori statistics, this estimation of the temperature profile is of a minimum information type("physical correction").

The analysis of this correction scheme is usually performed by looking at the final correction matrix between T_o the brightness temperature vector at the non nadir angle and T_n the brightness temperature at nadir. This is a 3 by 3 matrix for the case of MSU and a 12 by 12 for AMSU. The following analysis deals with the question of equivalent weighting functions.

The radiative transfer equation can be written in vectorial form for the two different sensing angles as

$$\begin{aligned} T_{bo} &= W_o T + \eta \\ T_{bn} &= W_n T + \eta \end{aligned} \qquad (V.3.3)$$

The minimum information estimate of the temperature from the non nadir angle is, omitting the mean, (Baumann, 1980)

$$\hat{T} = W_o^t (W_o W_o^t)^{-1} T_{bo} \qquad (V.3.4)$$

which yield for the estimate of the brightness temperature at nadir

$$\hat{T}_{bn} = W_n \hat{T} = W_n W_o^t (W_o W_o^t)^{-1} T_{bo} = C_{(o-n)} T_{bo} \qquad (V.3.5)$$

1.030	-0.087	0.012	
0.151	0.890	-0.024	32 to nadir
-0.079	0.136	0.945	
1.027	-0.247	-0.037	
0.517	0.572	-0.050	56 to nadir
-0.258	0.514	0.729	

Table V.3.1: Correction matrices corresponding to the limb correction for the Microwave Sounding Unit.

which defines the correction matrix from the angle o to nadir.

Table V.3.1 presents the correction matrix for nadir for each one of the sensing angles. As the sensing angle deviates from nadir, so does the correction matrix from the identity.

The equivalent weighting function $\tilde{W}_{no}$ of this correction scheme is defined by replacing the brightness temperature in equation V.3. by its equivalent temperature profile. Thus,

$$\hat{T}_{bn} = W_n W_o^t (W_o W_o^t)^{-1} W_o T + \eta = \tilde{W}_{no} T + \eta \qquad (V.3.6)$$

From the point of view of the forward problem, the closer the equivalent weighting function is to the true weighting function, the more appropriate this limb correction algorithm is. As it will soon be demonstrated in the section on the analysis of multidimensional retrieval algorithms using the singular value decomposition of weighting matrices, the reverse problem will also be used to determine when the limb correction is appropriate or not. But, let us solely consider the forward problem for the moment.

To test the limb correction, the algorithm is implemented for both the MSU and AMSU sounding systems. The weighting functions are computed at mid-latitude above land using the 33 level discretization of the atmosphere. The highest pressure level used in this representation is 5 *mbar*, a pressure level which is not truly appropriate for the high altitude probing channels of the AMSU but nonethe-

less permits to derive conclusions concerning the correction scheme.

For MSU, only the three temperature sounding channels were considered since channel one depends too strongly on the ground reflectivity. For AMSU, the 12 temperature channels of AMSU-A are used. In both cases, the extreme case of the correction was considered. That is, the correction of the weighting matrices from the extreme scanning angle (47° off nadir at the satellite, 60° at the surface of the Earth) to nadir is implemented.

Figures V.3.3 and V.3.4 illustrate this correction scheme by plotting the discrete weighting functions at nadir and the equivalent weighting functions. If the limb correction appears to be appropriate for the AMSU system (no noticeable changes in the weighting functions are present), it is not adequate for the MSU. The corrected weighting functions are broader than the real ones. Moreover, for the weighting channel probing the lowest in the atmosphere (f=53.74 GHz), the equivalent weighting function even becomes negative for certain pressure levels below the tropopause. This undesirable effect can be explained by the fact that as the weighting function rises in altitude, some information is lost forever and that the limb correction scheme cannot account for this loss solely on the basis of the measurements.

The broadening of the weighting functions results in a loss of resolution in the vertical direction. In order to quantify the broadening of the weighting function when the correction scheme is implemented, the ratio between the peak value of the weighting function to its respective absolute area τ is computed as a figure of merit. This ratio measures the sharpness of the weighting function. A "perfect" weighting functions probing at only one pressure level would have a τ equal to one, whereas a flat weighting function would have τ equal to $1/33 \approx 0.03$.

The results of this calculations are presented in Table V.3.2. As for the plots of the weighting functions, no appreciable differences are noticeable for the AMSU

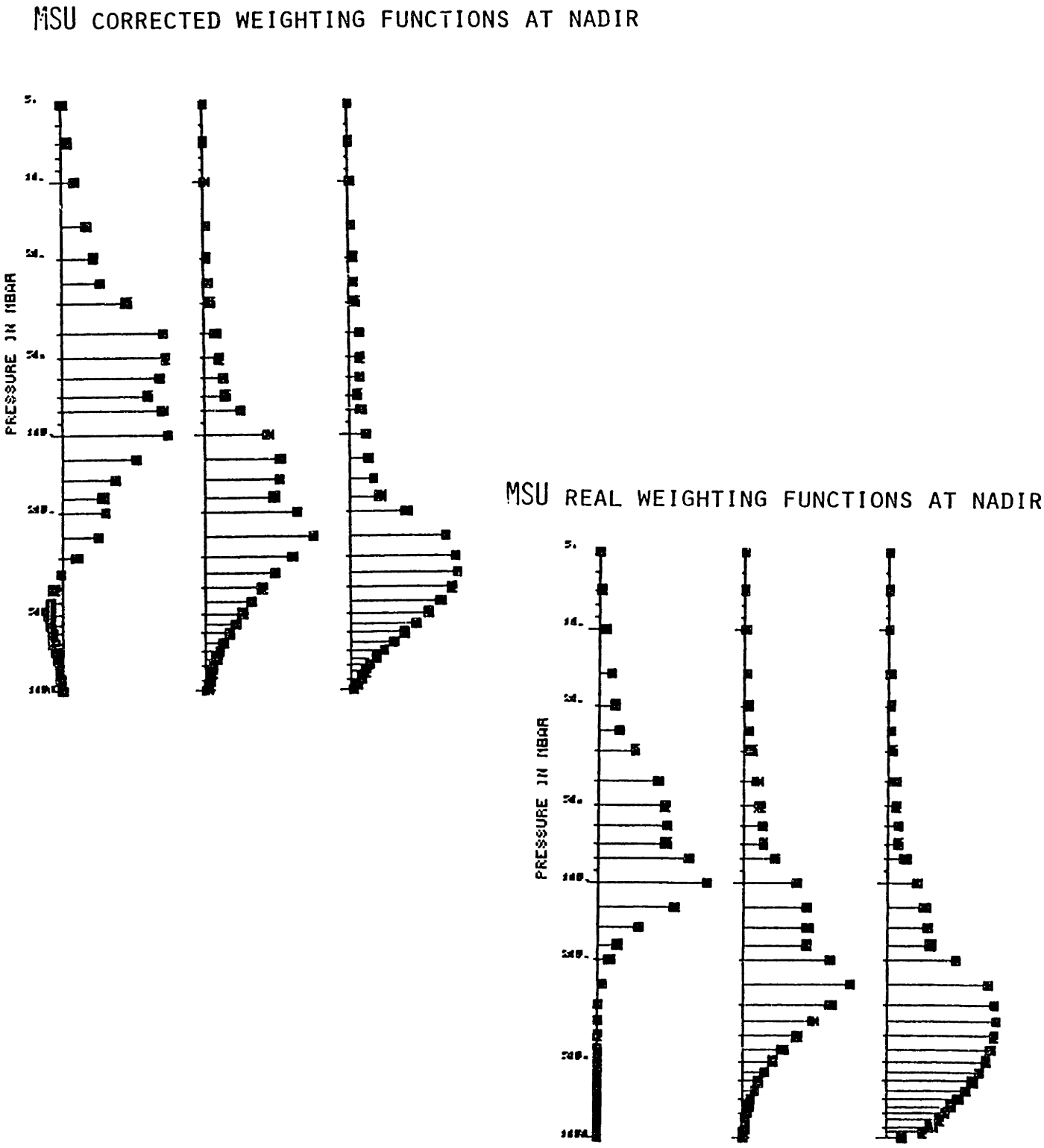


Figure V.3.3: MSU real weighting functions at nadir and equivalent weighting functions for the 47° to nadir correction.

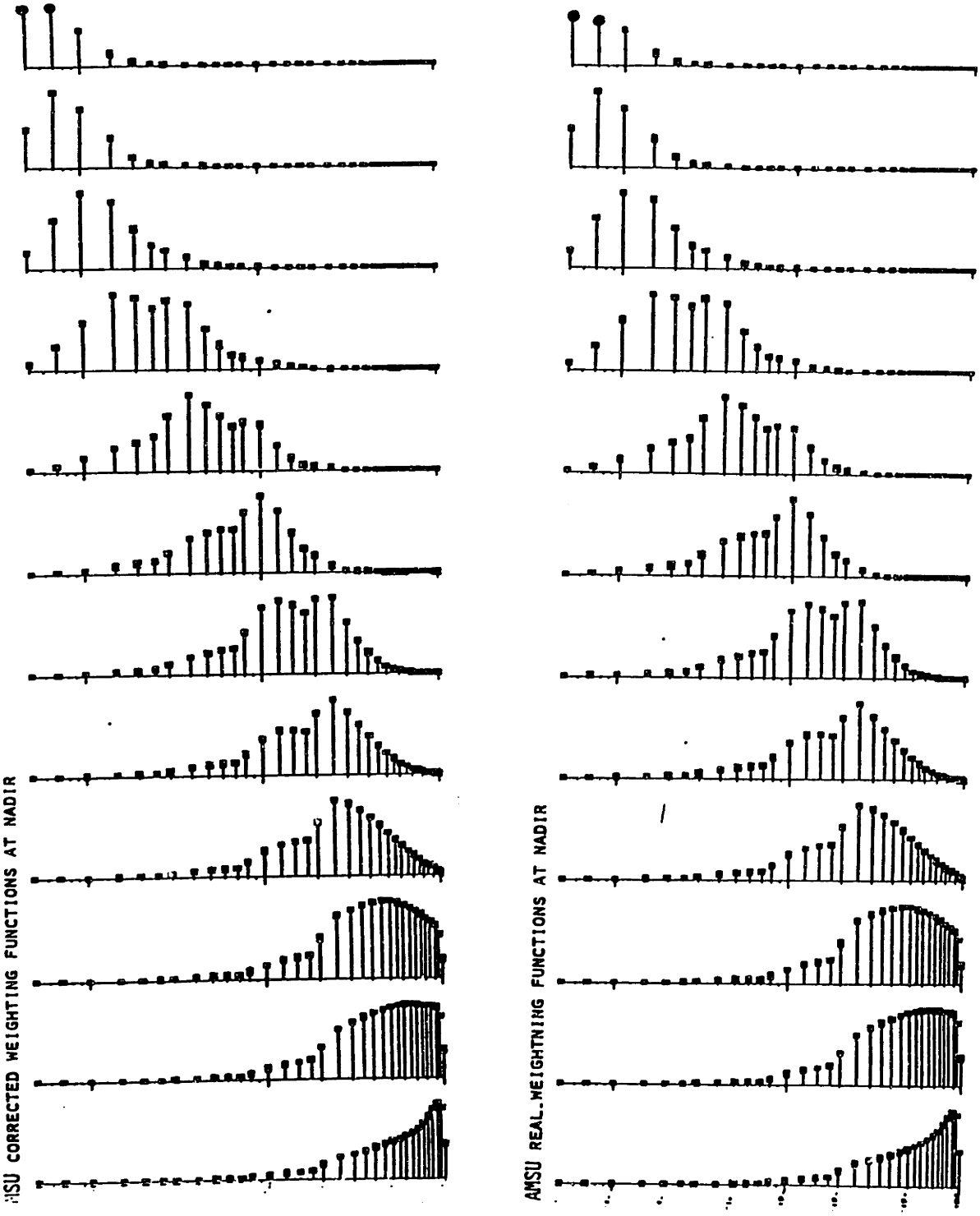


Figure V.3.4: AMSU real weighting functions at nadir and equivalent weighting functions for the 47° to nadir correction.

AMSU system

channel frequency	measure for real weight	measure for corrected weight
50.35	0.095	0.096
52.80	0.063	0.063
53.33	0.065	0.065
54.40	0.096	0.096
54.95	0.106	0.016
55.50	0.116	0.116
57.2903	0.154	0.150
57.0763	0.135	0.136
56.9201	0.128	0.140
56.9461	0.235	0.238
56.9581	0.337	0.337
56.9636	0.397	0.396

MSU system

channel frequency	measure for real weight	measure for corrected weight
53.74	0.072	0.099
54.96	0.116	0.105
57.95	0.152	0.093

Table V.3.2: Figure of merit for the sharpness of weighting function for MSU and AMSU, real vs. limb corrected weighting functions.

system. This is not true for MSU.

Based on these results, one can conclude that, from a forward problem point of view, the limb correction is not adequate for MSU. The impact of the limb correction in terms of the inverse problem is worse. This is true for two major reasons. The first is that the correction tries to estimate the brightness temperature for nadir where the influence of the surface is the most important. Thus, the impact of errors in the reflectivity and altitude of the surface on the retrieval error is amplified. The second reason is that the improvement that multidimensional retrieval schemes present over single-spot schemes depends upon changes in the observation matrices from point to point.

V.3.2 Frame and Time Series Representation

The regular spacing of the sampling pattern parallel to the track of the satellite is perfectly suited for a time series representation. This representation forms the basis for the implementation of Kalman filtering and two dimensional Wiener filters.

The sampling pattern is periodic after twenty four hours but since the temperature field evolves with time, the signal being estimated is not periodic. The time series representation does not permit to account for the correlation between the temperature profiles at points located at the same latitude.

The points sampled in the direction perpendicular to the track of the satellite do not follow a regular pattern and are not observed under the same conditions. They must be thought of as a basic unit which will hereafter be referred to as a frame. A frame is defined as the ensemble of brightness temperatures measurements (4×11 for MSU) and temperature components (33×11 for MSU) which are sampled in one swap of the satellite.

V.3.3 Analysis of the Two Point Retrieval Problem in the Natural Basis Representation

The optimal two point retrieval algorithm can readily be found by noting that nothing in equation (V.2.1) assumes that all components of the brightness temperature vector (sensing frequencies) and all components of the physical temperature vector (pressure levels) are related through the radiative transfer equation. This observation allows the treatment of the two point retrieval estimator problem as a one dimensional problem with augmented vectors (Kailath, 1976).

In the following analysis, the two points considered will be denoted as "one" and "two" (nothing too original). Point one is the point at which the estimation is being performed. Let $\overline{T}_a$ denote the concatenation of the physical tempera-

ture vectors at points one and two, namely $\overline{T}_a^t = (\overline{T}_1^t, \overline{T}_2^t)$. Let $\overline{T}_{ba}$ denote the corresponding augmented brightness temperature vector $\overline{T}_{ba}^t = (\overline{T}_{ba1}^t, \overline{T}_{ba2}^t)$. The radiative transfer equation can then be written as

$$\overline{T}_{ba} = \begin{pmatrix} \overline{T}_{ba1} \\ T_{ba2} \end{pmatrix} = \begin{pmatrix} W_1\overline{T}_1 \\ W_2\overline{T}_2 \end{pmatrix} + \begin{pmatrix} \overline{\eta}_1 \\ \overline{\eta}_2 \end{pmatrix} = \begin{pmatrix} W_1 & 0 \\ 0 & W_2 \end{pmatrix} T_a + \begin{pmatrix} \overline{\eta}_1 \\ \overline{\eta}_2 \end{pmatrix} \qquad (V.3.7)$$

thus defining a augmented weighting matrix W and noise vector $\overline{\eta}_a$.

Let us, also, define the statistics required to compute the estimator (at least the linear part of that estimator). Let $\overline{\overline{R}}$ denote the variance of the vector $\overline{T}_a$

$$\overline{\overline{R}} = \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix} \qquad (V.3.8)$$

where the submatrices R_{ij} denote the expectation of the outer product $\overline{T}_i\overline{T}_j^t$.

The optimal two point retrieval algorithm is given by

$$D^{(2)} = \overline{\overline{R}}W^t(W\overline{\overline{R}}W^t + \overline{\overline{N}})^{-1} \qquad (V.3.9)$$

The matrix $\overline{\overline{N}}$ is a diagonal matrix whose entries are the concatenation of the entries from the noise covariance matrices at point one and two. After expansion, the optimal two point retrieval operator is given by

$$\begin{pmatrix} D_{11}^{(2)} & D_{12}^{(2)} \\ D_{21}^{(2)} & D_{22}^{(2)} \end{pmatrix} = \begin{pmatrix} R_{11}W_1^t & R_{12}W_2^t \\ R_{21}W_1^t & R_{22}W_2^t \end{pmatrix} \begin{pmatrix} W_1R_{11}W_1^t + N_1 & W_1R_{12}W_2^t \\ W_2R_{21}W_1^t & W_2R_{22}W_2^t + N_2 \end{pmatrix}^{-1} \qquad (V.3.10)$$

where $D_{ij}^{(k)}$ denotes the k point retrieval operator giving the temperature at point i from the brightness temperature at point j.

To invert the covariance matrix of the augmented brightness temperature, one makes use of the matrix identity (Athans and Schweppe, 1965) giving the inverse of the block matrix

$$\begin{pmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{pmatrix}^{-1} = \begin{pmatrix} X_{11}^{-1} + X_{11}^{-1}X_{12}\Delta^{-1}X_{21}X_{11}^{-1} & -X_{11}^{-1}X_{12}\Delta^{-1} \\ -\Delta^{-1}X_{21}X_{11}^{-1} & \Delta^{-1} \end{pmatrix} \qquad (V.3.11)$$

valid if the submatrices X_{1i} and Δ admit an inverse. The matrix Δ is defined as

$$\Delta = X_{22} - X_{21}X_{11}^{-1}X_{12} \qquad (V.3.12)$$

which equates

$$W_2R_{22}W_2^t + N_2 - W_2R_{21}W_1^t(\Sigma_1R_{o1o1}\Sigma_1 + N_1)^{-1}W_1R_{12}W_2^t \qquad (V.3.13)$$

The first matrix in the above equation R_{11} will admit an inverse if the matrix $\overline{\overline{R}}$ is positive definite, that is if the vertical covariance kernel of the temperature field (or its model) is positive definite. This positive definiteness is a necessary condition. Likewise, the matrix Δ will admit an inverse if the temperature field has a positive definite three-dimensional covariance kernel. Both these conditions on the positive definiteness of the model for the three-dimensional macrostructure can be somewhat relaxed, in the sense that only the covariance in the visible space of the sounder at point one and two must be positive definite. The positive definiteness of the three-dimensional covariance kernel is not a necessary condition to the existence of an inverse and thus the definition of multidimensional retrieval operator. It simply guarantees that none of the eigenvalues of the covariance matrix in equation (V.3.8) are zero, and thus that this matrix is full rank. The condition for positive definiteness will become more relevant and important for the analysis of the retrieval error in a later portion of this section.

The "on spot" operator for point one is given by

$$\begin{aligned} D_{11}^{(2)} =& R_{11}W_1^t(\Sigma_1R_{o1o1}\Sigma_1 + N_1)^{-1} \\ &+\left(R_{11}W_1^t(\Sigma_1R_{o1o1}\Sigma_1 + N_1)^{-1} - I\right)R_{12}W_2^t\Delta^{-1}W_2R_{21}W_1^t(W_2R_{22}W_2^tN_2)^{-1} \end{aligned} \qquad (V.3.14)$$

whereas the "off spot" operator for point one is given by

$$D_{12}^{(2)} = \left(I - R_{11}W_1^t(\Sigma_1R_{o1o1}\Sigma_1 + N_1)^{-1}\right)R_{12}W_2\Delta^{-1} \qquad (V.3.15)$$

In the above equations, one can notice that two elements plays an important role in the way the two-point retrieval operator differs from the one-point retrieval

operator. The first element is, as could be expected, the cross-covariance matrix R_{12}. Should there be an minute statistical coupling between the two-points (or even a null one), the two-point estimator will become a one spot estimator. As a consequence, the error will not be reduced by the addition of additional spot in the retrieval process. The second element, which may not be as salient a priori as the one concerning R_{12} is the difference of $I - R_{11}W_1^t(\Sigma_1 R_{o1o1}\Sigma_1 + N_1)^{-1}W_1$. The second term in this equation can be recognized as the transfer function of the estimation process of the one-dimensional retrieval problem. The closer the one-dimensional retrieval is to being perfect, the smaller the difference there will be between the two spot and the one spot retrieval operators. The error as we will soon see follows the same pattern.

To summarize the form of the two-dimensional two spot retrieval operator, let us rewrite it using the transfer matrix for the one spot problem $\mathcal{T}_{11}$

$$D_{12}^{(2)} = (I - \mathcal{T}_{11})R_{12}W_2\Delta^{-1}$$

$$D_{11}^{(2)} = D_{11}^{(1)}W_2R_{12}R_{11}^{-1}\mathcal{T}_{11} \qquad (V.3.16)$$

Upon examining the error (and its reduction) of the two-point retrieval problem, the error vector is

$$\begin{aligned}\overline{e} =& \overline{T}_{11} - D_{11}^{(2)}\overline{T}_{b1} - D_{12}^{(2)}\overline{T}_{b2} \\ =& \overline{T}_1 - D_{11}^{(2)}W_1\overline{T}_1 - D_{12}^{(2)}W_2\overline{T}_2 - D_{11}^{(2)}\overline{\eta}_1 - D_{12}^{(2)}\overline{\eta}_2\end{aligned} \qquad (V.3.17)$$

and the error norm becomes the expectation

$$E = trace\{\overline{e}\overline{e}^t\} = trace\{\mathcal{E}\} \qquad (V.3.18)$$

This calculation may be simplified by noting that since the retrieval operator is optimal, the error criterion is given by

$$E = trace\{(I - D.W)R\} \qquad (V.3.19)$$

where the error pertains to the entire rms error, both at points one and two. In that case,

$$D = \begin{pmatrix} D_{11}^{(2)} & D_{12}^{(2)} \\ D_{21}^{(2)} & D_{22}^{(2)} \end{pmatrix} \qquad W = \begin{pmatrix} W_1 & 0 \\ 0 & W_2 \end{pmatrix} \qquad R = \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix} \tag{V.3.20}$$

Hence, the error is at point one

$$\begin{aligned} E =& trace\{I - D_{11}^{(1)} W_1 R_{11}\} \\ &+ trace\{D_{12}^{(2)} W_2 R_{21} W_1^t (\Sigma_1 R_{o1o1} \Sigma_1 + N_1)^{-1} W_1 R_{11}\} - trace\{i D_{12}^{(2)} W_2 R_{12}\} \end{aligned} \tag{V.3.21}$$

after regrouping,

$$\begin{aligned} E =& trace\{I - D_{11}^{(1)} W_1 R_{11}\} \\ &+ trace\{D_{12}^{(2)} W_2 (R_{21} - R_{21} W_1^t (\Sigma_1 R_{o1o1} \Sigma_1 + N_1)^{-1} W_1 R_{11})\} \end{aligned} \tag{V.3.22}$$

Let $E^{(i)}$ denote the error norm at point one using an i-point retrieval operator (in this specific case $i = 1, 2$). Let us write the error using the transfer matrix for the one spot retrieval problem. Thus,

$$\begin{aligned} E^{(2)} =& E^{(1)} - trace\{(I - \mathcal{T}_{11}) R_{12} W_2^t \Delta^{-1} W_2 R_{21} (I - \mathcal{T}_{11})\} \\ =& E^{(1)} - trace\{H \Delta^{-1} H^t\} \end{aligned} \tag{V.3.23}$$

To analyze the behavior of the error as the one-point transfer matrix and covariance matrix R_{12} vary, let us manipulate the matrix Δ

$$\begin{aligned} \Delta =& W_2 R_{22} W_2^t + N_2 - W_2 R_{21} W_1 (\Sigma_1 R_{o1o1} \Sigma_1 + N_1)^{-1} W_1 R_{12} W_2^t \\ =& W_2 R_{22} W_2^t + N_2 - W_2 (R_{21} W_1^t (\Sigma_1 R_{o1o1} \Sigma_1 + N_1)^{-1} W_1 R_{12}) W_2^t \end{aligned} \tag{V.3.24}$$

Introducing the one spot transfer matrix $\mathcal{T}_{11}$ in the definition of Δ yields

$$\begin{aligned} \Delta =& W_2 R_{22} W_2^t + N_2 - W_2 R_{21} R_{11}^{-1} R_{11} W_1 (\Sigma_1 R_{o1o1} \Sigma_1 + N_1)^{-1} W_1 R_{12} W_2^t \\ =& W_2 R_{22} W_2^t + N_2 - W_2 (R_{21} R_{11}^{-1} \mathcal{T}_{11} R_{12}) W_2^t \end{aligned} \tag{V.3.25}$$

To obtain a more symmetric formula, factor the covariance matrix at point two in the inside parentheses. Hence

$$\Delta = N_2 + W_2 (I - R_{21} R_{11}^{-1} \mathcal{T}_{11} R_{12} R_{22}^{-1}) R_{22} W_2^t \tag{V.3.26}$$

Except for the matrix $\mathcal{T}_{11}$, the matrix in parenthesis resembles the difference between the identity matrix and the square of the correlation coefficient between point one and two, should this correlation coefficient be defined in terms of matrices. In fact, the expression in parenthesis is (without the transfer matrix) the "Δ" matrix associated with the normalized two-point covariance matrix

$$\begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix} \begin{pmatrix} R_{11}^{-1} & 0 \\ 0 & R_{22}^{-1} \end{pmatrix} = \begin{pmatrix} I & R_{12}R_{22}^{-1} \\ R_{21}R_{11}^{-1} & I \end{pmatrix} \qquad (V.3.27)$$

Whenever there is a weak crosscorrelation between points one and two, the matrix Δ is close to the covariance matrix of the brightness temperature at point two. That is,

$$\Delta \approx N_2 + W_2 R_{22} W_2^t \qquad (V.3.28)$$

On the other hand, a strong crosscorrelation ($R_{12} \approx R_{22} \approx R_{21} \approx R_{11}$) yields a matrix dominated by the difference between the identity and transfer function. This difference corresponds to the normalized error for the one spot retrieval operator. Namely,

$$\Delta \approx N_2 + W_2(I - \mathcal{T}_{11})R_{22}W_2^t = N_2 + W_2 \mathcal{E}^{(1)} R_{11}^{-1} R_{22} W_2 \qquad (V.3.29)$$

where the error matrix for the spot retrieval problem $\mathcal{E}^{(1)}$ is defined as in equation (V.2.7). Moreover, if the two weighting matrices are the same (the two-points considered in the retrieval are sounding the atmosphere at the same angle), the assumption of stationarity further simplifies the expression to

$$\Delta \approx N_2 + W \mathcal{E}^{(1)} W^t \qquad (V.3.30)$$

The matrix H from equation (V.3.23) also features the error matrix for the one spot estimator

$$H = (I - \mathcal{T}_{11})R_{12}W_2^t = \mathcal{E}^{(1)} R_{11}^{-1} R_{12} W_2^t \qquad (V.3.31)$$

Plugging (V.3.31) and (V.3.25) into (V.3.23) demonstrates that the improvement from a one spot to a two spot retrieval operator is quadratic in the correlation between the two-points considered. The improvement is quadratic in the one-dimensional error $\mathcal{E}^{(1)}$ for weak correlations between the two-points (when the Δ in equation (V.3.25) is dominated by the $W_2 R_{22} W_2^t$ term). The improvement in the error is only a linear function of the one spot error for a strong correlation between points one and two. That is when Δ is dominated by $W\mathcal{E}^{(1)}W^t$. Such will be the case for sounding systems with a fine sampling grid.

The preceding computation allows us to draw important conclusions regarding the performances of different types of retrieval algorithms as applied to specific sounding systems (satellites). In terms of the error improvement as related to the implementation of multidimensional retrieval algorithms, the number of channels (sensing frequencies), distance, and footprints between adjacent points (controlling the correlation) are the parameters relevant to this analysis.

For systems with numerous sounding channels, the error using a single spot (D matrix) retrieval operator is, all thing considered, small. These systems do not profit as much from the extension of the region of support of the brightness temperature readings being operated upon as do sounding channels with fewer sounding channels.

Whenever the distance between adjacent points is large enough compared to the isobaric scale of fluctuation so that there is a small cross correlation matrix R_{21} (small compared to R_{11} and R_{22}), reducing the error using a single-spot retrieval operator is most important for it limits the possible improvements using multidimensional retrieval operators. In such cases, as for example MSU, where the resolution is poor and the distance between measurements is large compared to the decay constant of the isobaric covariance kernel, one must concentrate on reducing the one-dimensional retrieval error. This phenomenon implies a careful

determination of the vertical covariance kernel used in computing the D matrix.

The analysis of the retrieval in the natural basis representation of the temperature profile does not provide any information concerning its observable and unobservable components. In particular, one might consider if the estimates of the observable and the unobservable component are separable. The impact of different weighting matrices at points one and two is not clear. Moreover, the impact of including several input points in the retrieval process may vary from component to component. To perform this analysis, one must project this analysis in the SVD basis.

V.3.4 Analysis of the Two Point Retrieval Operator in the SVD Basis Representation.

As for the one-dimensional case, this analysis is performed using the singular value decomposition of the observation matrices at points one and two. These decompositions do not have to be identical because of the lifting of the weighting functions with the sensing angle which results with different weighting matrices at different angles.

The observation equation takes the familiar form for points one and two

$$\begin{aligned} \overline{T}_{b1} &= W_1 \overline{T}_1 + \overline{\eta}_1 \\ \overline{T}_{b2} &= W_2 \overline{T}_2 + \overline{\eta}_2 \end{aligned} \tag{V.3.32}$$

Now, introduce the two singular value decompositions corresponding to the weighting matrices at points one and two. Namely,

$$\begin{aligned} W_1 &= \mathcal{W}_{l1}^t \begin{pmatrix} \Sigma_1 & 0 \end{pmatrix} \mathcal{W}_{r1} \\ W_2 &= \mathcal{W}_{l1}^t \begin{pmatrix} \Sigma_2 & 0 \end{pmatrix} \mathcal{W}_{r1} \end{aligned} \tag{V.3.33}$$

where the matrices Σ_1 and Σ_2 are in general different and may be of unequal size. Plugging (V.3.33) into (V.3.32) yields the set of equations

$$\begin{aligned} \overline{T}_{b1} &= \mathcal{W}_{l1}^t \begin{pmatrix} \Sigma_1 & 0 \end{pmatrix} \mathcal{W}_{r1} \overline{T}_1 + \overline{\eta}_1 \\ \overline{T}_{b2} &= \mathcal{W}_{l2}^t \begin{pmatrix} \Sigma_2 & 0 \end{pmatrix} \mathcal{W}_{r2} \overline{T}_2 + \overline{\eta}_2 \end{aligned} \tag{V.3.34}$$

Let us now perform the basis changes for both brightness and physical temperatures onto the SVD basis. This yields

$$\begin{aligned}\overline{T}_1' &= \mathcal{W}_{l1}\overline{T}_1 = (\Sigma_1 \quad 0)\begin{pmatrix}\overline{T}_{1s}\\ \overline{T}_{1h}\end{pmatrix} + \mathcal{W}_{l1}\overline{\eta}_1 \\ \overline{T}_2' &= \mathcal{W}_{l1}\overline{T}_2 = (\Sigma_2 \quad 0)\begin{pmatrix}\overline{T}_{2s}\\ \overline{T}_{2h}\end{pmatrix} + \mathcal{W}_{l1}\overline{\eta}_2\end{aligned} \qquad (V.3.35)$$

The concatenation of the vectors of physical temperature can be estimated from the concatenation of the brightness temperatures by using the same formula that was previously developed for the one-dimensional case.

Let $\overline{\overline{W}}$ denote the sparse weighting matrix corresponding to the two-points, and let $\overline{\overline{N}}$ denote the covariance matrix associated with the sensor noise once projected onto the SVD basis. Finally let $\overline{\overline{R}}$ denote the covariance matrix for the temperature field. The noise covariance matrix is a block diagonal matrix

$$\overline{\overline{N}} = \begin{pmatrix} N_1 & 0 \\ 0 & N_2 \end{pmatrix} \qquad (V.3.36)$$

The noise block submatrices N_i will not in general be diagonal, since they correspond to the expectation of linear combinations of originally uncorrelated observation noises. These matrices will be diagonal if all the channels have the same noise variance. These noise covariance noise covariance matrices will be manipulated further when the brightness temperature fields are corrected to accommodate discrepancies between theoretical and experimental temperatures (see Chapter VI).

The temperature covariance matrix is a four-by-four block symmetric matrix

$$\overline{\overline{R}} = \begin{pmatrix} R_{o1o1} & R_{u1o1} & R_{o2o1} & R_{u2o1} \\ R_{o1u1} & R_{u1u1} & R_{o2u1} & R_{u2u1} \\ R_{o1o2} & R_{u1o2} & R_{o2o2} & R_{u2o2} \\ R_{o1u2} & R_{u1u2} & R_{o2u2} & R_{u2u2} \end{pmatrix} \qquad (V.3.37)$$

The two-point estimator is given by (Kailath, 1976)

$$\overline{\overline{D}} = \overline{\overline{R}}W^t(W\overline{\overline{R}}W^t + \overline{\overline{N}})^{-1} \qquad (V.3.38)$$

After several trivial manipulations, the estimator can be expressed as

$$\overline{\overline{D}} = \begin{pmatrix} R_{o1o1}\Sigma_1 & R_{o2o1}\Sigma_2 & R_{o1u1}\Sigma_1 & R_{u2o1}\Sigma_2 \\ R_{o1o2}\Sigma_1 & R_{o2o2}\Sigma_2 & R_{o1u2}\Sigma_1 & R_{o2u2}\Sigma_2 \end{pmatrix} \begin{pmatrix} \Sigma_1 R_{o1o1}\Sigma_1 + N_1 & \Sigma_1 R_{o2o1}\Sigma_2 \\ \Sigma_2 R_{o1o2}\Sigma_1 & \Sigma_2 R_{o2o2}\Sigma_2 + N_2 \end{pmatrix}^{-1} \quad (V.3.39)$$

Let us make use again of the matrix identity (Athans and Schweppe, 1965) giving the inverse of the block matrix

$$\begin{pmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{pmatrix}^{-1} = \begin{pmatrix} X_{11}^{-1} + X_{11}^{-1}X_{12}\Delta^{-1}X_{21}X_{11}^{-1} & -X_{11}^{-1}X_{12}\Delta^{-1} \\ -\Delta^{-1}X_{21}X_{11}^{-1} & \Delta^{-1} \end{pmatrix} \quad (V.3.40)$$

valid if the submatrix X_{11} and the matrix Δ admits an inverse. The matrix Δ is defined as

$$\Delta = X_{22} - X_{21}X_{11}^{-1}X_{12} \quad (V.3.41)$$

Applying the formula for the inverse yields in the case of the equation defining the two-point retrieval operator

$$\begin{aligned} \Delta &= \Sigma_2 R_{o2o2}\Sigma_2 + N_2 - \Sigma_2 R_{o1o2}\Sigma_1(\Sigma_1 R_{o1o1}\Sigma_1 + N_1)^{-1}\Sigma_1\Sigma_1 R_{o2o1}\Sigma_2 \\ &= \Sigma_2\big(R_{o2o2} + \hat{N}_2 - R_{o1o2}(R_{o1o1} + \hat{N}_1)^{-1}R_{o2o1}\big)\Sigma_2 \\ &= \Sigma_2\hat{\Delta}\Sigma_2 \end{aligned} \quad (V.3.42)$$

where the modified noise covariance matrices are defined as

$$\hat{N}_i = \Sigma_i^{-1}N_i\Sigma_i^{-1}. \quad (V.3.43)$$

The modified covariance matrices are in a way normalized to the physical temperature rather than to the brightness temperature. Note this Δ matrix is different from the previously shown Δ matrices (although usually related to them by a simple similarity transformation).

One can now derive the expressions for the two-points retrieval operator estimating the observable and unobservable components at location one from the brightness temperature measured at points one and two.

• *observable component at point one from brightness temperature at point one*

$$R_{o1o1}\Sigma_1(\Sigma_1 R_{o1o1}\Sigma_1 + N_1)^{-1}$$
$$+\big(R_{o1o1}\Sigma_1(\Sigma_1 R_{o1o1}\Sigma_1 + N_1)^{-1}\Sigma_1 R_{o2o1}\Sigma_2 - R_{o2o1}\Sigma_2\big)\Delta^{-1}\Sigma_2 R_{o1o2}\Sigma_1(\Sigma_1 R_{o1o1}\Sigma_1 + N_1)^{-1} \quad (V.3.44)$$

after manipulations

$$R_{o1o1}(R_{o1o1} + \hat{N}_1)^{-1}\Sigma_1^{-1} - \hat{N}_1(R_{o1o1} + \hat{N}_1)^{-1}R_{o2o1}\hat{\Delta}^{-1}R_{o1o2}(R_{o1o1} + \hat{N}_1)^{-1}\Sigma_1^{-1} \quad (V.3.45)$$

• *observable component at point one from brightness temperature at point two*

$$-R_{o1o1}\Sigma_1(\Sigma_1 R_{o1o1}\Sigma_1 + N_1)^{-1}\Sigma_1 R_{o2o1}\Sigma_2\Delta^{-1} + R_{o2o1}\Sigma_2\Delta^{-1} \quad (V.3.46)$$

which simplifies to

$$\hat{N}_1(R_{o1o1} + \hat{N}_1)^{-1}R_{o2o1}\Sigma_2^{-1}\Delta^{-1} \quad (V.3.47)$$

• *unobservable component at point one from brightness temperature at point one*

$$R_{o1u1}\Sigma_1(\Sigma_1 R_{o1o1}\Sigma_1 + N_1)^{-1}$$
$$+\big(R_{o1u1}\Sigma_1(\Sigma_1 R_{o1o1}\Sigma_1 + N_1)^{-1}\Sigma_1 R_{o2o1}\Sigma_2 - R_{o1u2}\Sigma_2\big)\Delta^{-1}\Sigma_2 R_{o1o2}\Sigma_1(\Sigma_1 R_{o1o1}\Sigma_1 + N_1)^{-1} \quad (V.3.48)$$

which yields

$$R_{o1u1}(R_{o1o1} + \hat{N}_1)^{-1}\Sigma_1^{-1}$$
$$+\big(R_{o1u1}(R_{o1o1} + \hat{N}_1)^{-1} + R_{o1u2}R_{o2o1}^{-1}\big)R_{o2o1}\Sigma_2^{-1}\Delta^{-1}\Sigma_2 R_{o1o2}(R_{o1o1} + \hat{N}_1)^{-1}\Sigma_1^{-1} \quad (V.3.49)$$

provided that the covariance matrix R_{o1o2} admits an inverse.

• *unobservable component at point one from brightness temperature at point two*

$$\begin{aligned} &- R_{o1u1}\Sigma_1(\Sigma_1 R_{o1o1}\Sigma_1 + N_1)^{-1}\Sigma_1 R_{o2o1}\Sigma_2\Delta^{-1} + R_{o2u1}\Sigma_2\Delta^{-1} \\ &= \big(R_{o2u1}R_{o2o1}^{-1} - R_{o1u1}(R_{o1o1} + \hat{N}_1)^{-1}\big)R_{o2o1}\Sigma_2\Delta^{-1} \end{aligned} \quad (V.3.50)$$

In order to interpret the preceding equations, one must introduce new matrices. Let $D^{(k)}_{oiTbj}$ denote the k-point retrieval operator that estimates the observable component (a u would signify the unobservable component) from the brightness temperature measured at point j $(i,j,k = 1,2)$. Let $G_{(1-2)}$ denote the gain matrix defined as

$$G_{(1-2)} = \Sigma_2 R_{o1o2}(R_{o1o1} + \hat{N}_1)^{-1}\Sigma_1^{-1} \qquad (V.3.51)$$

Let $C_{o(1-2)}$ denote the coupling matrix for the observable components defined as

$$C_{o(1-2)} = \hat{N}_1(R_{o1o1} + \hat{N}_1)^{-1} \qquad (V.3.52)$$

Likewise the coupling matrix for the unobservable components is defined as

$$C_{u(1-2)} = -R_{o1u1}(R_{o1o1} + \hat{N}_1)^{-1} + R_{o1u2}R_{o2o1}^{-1} \qquad (V.3.53)$$

Using the newly defined matrices, the equations defining the two-point retrieval operator can be written as

$$\begin{aligned} D^{(2)}_{o1Tb1} &= D^{(1)}_{o1Tb1} - D^{(2)}_{o1Tb2}G_{(1-2)} \\ D^{(2)}_{o1tb2} &= C_{o(1-2)}R_{o2o1}\Sigma_2\Delta^{-1} \\ D^{(2)}_{u1Tb1} &= D^{(1)}_{u1Tb1} - D^{(2)}_{u1Tb2}G_{(1-2)} \\ D^{(2)}_{u1tb2} &= C_{u(1-2)}R_{o2o1}\Sigma_2\Delta^{-1} \end{aligned} \qquad (V.3.54)$$

Both retrieval operators for observable and unobservable components follow the same type of recursion when growing from one-point to two-points.

- The core of the transformation common to the two operators is the matrix product $R_{o2o1}\Sigma_2\Delta^{-1}$ which depends solely on the correlation between the brightness temperatures and the signal-to-noise ratio of the sounding at point two.

- The off-spot retrieval operator is the product of the core matrix by a coupling matrix $C_{.(1-2)}$ which differs for observable and unobservable components.
- The on spot operator is corrected from the one spot operator by the product of the off spot operator times a gain matrix $G_{(1-2)}$. This gain matrix measures the cross-covariance between the observable components at point one and two normalized to the covariance of the measured observable components plus noise.

To complement the analysis of the two-point retrieval operator in the SVD basis, let us consider the evolution of the error. The error analysis follows closely the derivation of the error norm for the estimation expressed in the regular basis (where each component of the physical temperature vector is a specific pressure level). The error norm is given by

$$E = trace\{(I - D.W)R\} \qquad (V.3.55)$$

where the different matrices are expressed in the SVD basis.

After separating the error norm at points one and two, one expresses the retrieval error in the observable and unobservable subspaces of the weighting matrix at point one.

- *retrieval error for the observable components at point one*

$$\begin{aligned} E_{oo} =& trace\{R_{o1o1} - D^{(1)}_{o1Tb1}\Sigma_1 R_{o1o1}\} \\ &+ trace\{D^{(2)}_{o1Tb2}G_{(1-2)}\Sigma_1 R_{o1o1}\} - trace\{D^{(2)}_{o1Tb2}\Sigma_2 R_{o1o2}\} \end{aligned} \qquad (V.3.56)$$

After several algebraic manipulations,

$$\begin{aligned} E_o =& trace\{R_{o1o1} - D^{(1)}_{o1Tb1}\Sigma_1 R_{o1o1}\} \\ &+ trace\{D^{(2)}_{o1Tb2}\Sigma_2 R_{o1o2}\big((\Sigma_1 R_{o1o1}\Sigma_1 + N_1)^{-1}R_{o1o1} - I\big)\} \\ =& trace\{R_{o1o1} - D^{(1)}_{o1Tb1}\Sigma_1 R_{o1o1}\} \\ &- trace\{\hat{N}_1(\Sigma_1 R_{o1o1}\Sigma_1 + N_1)^{-1}R_{o2o1}\Sigma_2\Delta^{-1}\Sigma_2 R_{o1o2}(\Sigma_1 R_{o1o1}\Sigma_1 + N_1)^{-1}\hat{N}_1\} \end{aligned} \qquad (V.3.57)$$

which yields after further manipulations

$$E_o = trace\{R_{o1o1} - D^{(1)}_{o1Tb1}\Sigma_1 R_{o1o1}\} + trace\{C_{o(1-2)}R_{o2o1}\Sigma_2\Delta^{-1}\Sigma_2 R_{o1o2}C^t_{o(1-2)}\} \quad (V.3.58)$$

- *retrieval error for the unobservable components at point one*

$$\begin{aligned} E_u =& trace\{R_{u1u1} - D^{(1)}_{u1Tb1}R_{u1o1}\} \\ &+ trace\{D^{(2)}_{u1Tb2}G_{(1-2)}R_{o1o1} - D_{u1Tb2}R_{u1o2}\} \end{aligned} \quad (V.3.59)$$

$$\begin{aligned} E_u =& trace\{R_{u1u1} - D^{(1)}_{u1Tb1}R_{u1o1}\} \\ &+ trace\{C_{u(1-2)}R_{o2o1}\Sigma_2\Delta^{-1}\Sigma_2 R_{o1o2}(\Sigma_1 R_{o1o1}\Sigma_1 + N_1)^{-1}R_{u1o1}\} \\ &- trace\{C_{h(1-2)}R_{o2o1}\Sigma_2\Delta^{-1}\Sigma_2 R_{u1o2}\} \end{aligned} \quad (V.3.60)$$

which yields

$$E_u = trace\{R_{u1u1} - D^{(1)}_{u1Tb1}R_{u1o1}\} + trace\{C_{u(1-2)}R_{o2o1}\Sigma_2\Delta^{-1}\Sigma_2 R_{o1o2}C^t_{u(1-2)}\} \quad (V.3.61)$$

As for the analysis it is apparent that the improvement in the retrieval error depends on the cross covariance matrix between points one and two. As it could have been expected from intuitive arguments, only the cross covariance between the components of the temperature profile in the observable part of the observation matrices at point one and two matters. This covariance is within a similarity transformation and a scaling the cross-covariance of the brightness temperature.

The analysis of the error in the SVD basis provides results which are not apparent in any other basis representation of temperature profiles. These results pertain to the error improvement for observable and the unobservable components and to the fact that they are different and controlled by two different aspects of the retrieval problem.

For both components the reduction in error by incorporating an extra point in the retrieval process can be written as $trace\{CK_{core}C^t\}$ where K_{core} is the core

of the improvement characterizing the quality of the second spot information *vis a vis* the retrieval problem

$$K_{core} = R_{o2o1}\left(R_{o2o2} + \hat{N}_2 - R_{o1o2}(\Sigma_1 R_{o1o1}\Sigma_1 + N_1)^{-1}R_{o2o1}\right)^{-1}R_{o1o2} \quad (V.3.62)$$

and C the appropriate coupling matrix described by equations (V.3.52) and (V.3.53).

Before analyzing the impact of the coupling matrices on the retrieval error, let us analyze the core of the improvement K_{core}. As for the analysis in the natural basis, let us manipulate that matrix to make the normalized correlation matrices appear. Namely,

$$K_{core} = R_{o2o1}R_{o2o1}^{-1}\left(N_2R_{o2o2}^{-1} + I - R_{o1o2}R_{o1o1}^{1}D_{o1Tb1}^{(1)}R_{o2o1}R_{o2o2}^{-1}\right)^{-1}R_{o1o2} \quad (V.3.63)$$

Several limiting cases are possible depending on the strength of the correlation between the observable components over the retrieval input region of support (ROS).

As expected, should the correlation R_{o2o1} go to zero so would the improvement. In such a situation, there is no point in doing multidimensional retrievals. When there is a weak correlation, the matrix K_{core} can be approximated to

$$\begin{aligned} K &\approx R_{o2o1}R_{o2o2}^{-1}(I + N_2R_{o2o2}^{-1})^{-1}R_{o1o2} \\ &\approx R_{o2o1}R_{o2o2}^{-1}R_{o1o2} \end{aligned} \quad (V.3.64)$$

for typical sounding systems with good SNRs. This improvement is a quadratic function of the cross-covariance between observable components and is independent of the retrieval error of the one-spot operator.

To relate this quadratic dependence to the model for the isobaric covariance kernel, let us assume a case where the decay constant dominates the oscillatory term in the description of the isobaric covariance kernel (Equation III.4.4). In addition, assume that the decay constant does not change abruptly over the main

contribution of the weighting functions. That is, the isobaric decay constant associated with each one of the observable components is fairly equal. This assumption becomes increasingly restrictive as the number of sounder channels increases. This is true, because it implies a complete separability between the pressure and isobaric covariance kernels as is the case for a purely two-dimensional flow.

Under these working assumptions, the cross-covariance between the observable components can be modeled as a first order Markovian process

$$R_{o1o2}(d_{12}) = R_{o1o1}\exp(-\alpha d_{12}) \qquad (V.3.65)$$

where d_{12} denotes the distance between the two-points. Now, consider a sounder with twice the sounding resolution. The corresponding cross-covariance becomes

$$R_{o1o2}(\frac{d_{12}}{2}) = R_{o1o1}\exp(-\alpha\frac{d_{12}}{2}) = R_{o1o1}\sqrt{\exp(-\alpha d_{12})} \qquad (V.3.66)$$

The improvement in the retrieval error obtained by incorporating the second point will be $\exp(\alpha d_{12})$ larger for the fine sounder than for the coarse sounder. For a sounder with a distance of 150 km between samples over typical winter atmospheres, changing the distance to 75 km will improve the reduction in error by a factor 1.2.

When there is a strong correlation, the core matrix K_{core} becomes

$$K_{core} \approx R_{o2o1}R_{o2o2}^{-1}(I - D_{o1Tb1}^{(1)})R_{o1o2} = R_{o2o1}R_{o2o2}^{-1}\mathcal{E}_s^{(1)-1}R_{o1o2} \qquad (V.3.67)$$

which displays the same quadratic in the cross-covariance and varies as the inverse of the one-spot error for the observable components.

The estimation of the observable component for both the one-spot and the multispot retrievals solely relies on the a-priori statistics and the sensing noise characteristics. The coupling matrix for the observable component $C_{o(1-2)}$ is a matricial signal-to-noise ratio. It is purely determined by the characteristics of the sounding at the point one and does not depend on the cross-covariance between points one and two.

In all cases the coupling matrix tends to zero as the signal-to-noise ratio for the sounding at point one improves. Then, the observable components of the temperature profile are computed directly from the brightness temperature through similarity and scaling. When this is the case, the expected retrieval error becomes

$$R_{o1o1} - R_{o1o1}\Sigma_1(\Sigma_1 R_{o1o1}\Sigma_1)^{-1}\Sigma_1 R_{o1o1} = 0 \qquad (V.3.68)$$

In the cases of high SNR, the estimator for the observable part is primarily a single-spot estimator centered at the point that performs the actual measurements. This characteristic implies that once projected along the basis vectors of the observable vectorial subspace, the retrieval operators will be narrow in the space domain. In the context of the time series representation, this point spread response corresponds to a "wide" Fourier transform. That is, the retrieval operator is an all-pass filter. When the sensor noise is important, the retrieval operator presents smooth variations in its impulse response, and becomes a low-pass filter that improves retrieval error by filtering sensor noise out.

In summary, under reasonable circumstances the improvement in the retrieval error when implementing a multidimensional retrieval scheme does not happen from the low-pass smoothing of the observable components.

The other case, of less interest, is that of a poor SNR at point one. Under these circumstances, the entries of the one-point retrieval operator are small, i.e. the a priori mean is guessed. This is true whenever the noise specifications of the satellite are loose, or when the number of degrees of freedom used in fitting the vertical covariance kernel are few and thus the a priori power of the signal is small. In such cases, the coupling matrix approaches identity and the improvement (if any) from one- to two-point retrieval is solely controlled by the core matrix.

Note that the relative signal-to-noise ratio (at each of the common eigenvectors) is solely controlled, once the satellite is orbiting, by the vertical covariance kernel. Should a good adaption scheme to changes in statistics be implemented,

the observable retrieval operator will be less all pass than it would have been if operating without changes in the climatology. This tendency to low pass filter the observable components would also occur had the a priori mean been updated regularly from an NMC numerical weather forecast. This is true, because less deviation of the profiles from the mean can be expected. Thus the power in the signal will be less. The resulting signal-to-noise ratio will be less than if larger deviations from the mean were expected. This is true independently of the "form" of the correlation matrix – by form is meant the normalized correlations $r_{ij}/\sqrt{r_{ii}r_{jj}}$ –.

The coupling matrix $C_{u(1-2)}$ for the unobservable components depends on the cross-covariance between points one and two for both observable and unobservable components.

Let us locate the elements (submatrices) of the coupling matrix in reference to the four-by-four covariance matrix of the two profiles at points one and two (Equation V.3.37), adding the normalized covariance matrix for the noise to the temperature covariances.

$$\begin{pmatrix} R_{o1o1} + \hat{N}_1 & \bullet & R_{o2o1} & \bullet \\ R_{o1u1} & \bullet & R_{o2u1} & \bullet \\ \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet \end{pmatrix} \qquad (V.3.69)$$

The coupling matrix computation is as follows: The matrices of the second row of the above equation are multiplied on the right by the inverse of the corresponding matrix on the first row, then the difference between these two products is calculated. The first element of this difference is the optimal estimator of the unobservable components at point one given the brightness temperature at point one. This estimator is similar to the mapping matrices encountered in the analysis of the one-spot estimator (Section V.2). The second element of the difference is more difficult to characterize. Should R_{o1o2} be $R_{o1o1} + \hat{N}_1$, it would be the optimal estimator of the unobservable component at point two from the brightness temperature at point one. The second element will close to this entity if the signal-to-noise

ratio is large and the temperature strongly correlated in the horizontal direction. In a certain sense, the coupling matrix for the unobservable components measure the non uniformity of the isobaric correlation.

Let us relate this coupling matrix to the three-dimensional covariance of temperature fields in order to better interpret its variations. In order to simplify the analysis, take the case where the signal-to-noise ratio at point one is good, that is where all eigenvalues of N_1 are small as compared to the ones of R_{o1o1}. Let $K_{TT}(\sigma, p, p')$ denote the three-dimensional covariance field and $\overline{K}_{TT}(\sigma)$ denote its vectorial representation (that is the expectation of the outer product $T_1 T_2^t$ when the two physical temperature vectors are separated by σ).

Project the covariance onto the right basis functions to obtain the different matrices involved in (V.3.69). Let P_{oi} and P_{uj} denote the projection operators onto the observable and unobservable components at points i and j. These two operators, when expressed in the SVD basis, are diagonal matrices with ones and zeros as entries. Using the rotation operators of equation (V.3.33) and the projections, the coupling matrix is

$$\begin{aligned} C_{u(1-2)} = {} & P_{u1} \mathcal{W}_{r1} K_{TT}(\sigma) \mathcal{W}_{r2}^t P_{o2} (P_{o1} \mathcal{W}_{r1} K_{TT}(\sigma) \mathcal{W}_{r2}^t P_{o2})^{-1} \\ & - P_{u1} \mathcal{W}_{r1} K_{TT}(0) \mathcal{W}_{r1}^t P_{o1} (P_{o1} \mathcal{W}_{r1} K_{TT}(0) \mathcal{W}_{r1}^t P_{o1})^{-1} \end{aligned} \qquad (V.3.70)$$

Two elements contribute to the coupling matrix: The three-dimensional covariance field and the differences between the rotation matrices $\mathcal{W}_{r1}$ and $\mathcal{W}_{r2}$. In order to characterize the impact of these terms, consider some limiting cases.

Contrary to the core matrix K_{core}, the coupling matrix $C_{u(1-2)}$ does not go to zero if the correlation between the temperature fields at the two-points goes to zero. In order to find its limiting form as the correlation length of the temperature field decreases, one must invoke a matricial L'hospital rule which does not provide any interesting results or interpretations.

Now, consider the case where the three-dimensional covariance kernel is as-

sumed to be exactly barotropic with the two-dimensional flow of the atmosphere presenting the same isobaric characterization. That is,

$$K_{TT}(\sigma) = K^{p}_{TT}(0) k^{i}_{TT}(\sigma) \qquad (V.3.71)$$

where the isobaric term k^{i}_{TT} is a scalar (common to all pressure levels). Under these assumptions, one can factor out that isobaric behavior and the coupling matrix becomes

$$\begin{aligned} C_{u(1-2)} = {} & P_{u1} \mathcal{W}_{r1} K^{p}_{TT}(0) \mathcal{W}^{t}_{r2} P_{o2} (P_{o1} \mathcal{W}_{r1} K^{p}_{TT}(0) \mathcal{W}^{t}_{r2} P_{o2})^{-1} \\ & - P_{u1} \mathcal{W}_{r1} K^{p}_{TT}(0) \mathcal{W}^{t}_{r1} P_{o1} (P_{o1} \mathcal{W}_{r1} K^{p}_{TT}(0) \mathcal{W}^{t}_{r1} P_{o1})^{-1} \end{aligned} \qquad (V.3.72)$$

This matrix is identically null if the two matrices $\mathcal{W}_{r1}$ and $\mathcal{W}_{r2}$ are identical. This is true, because under these assumptions the second points observable components do not carry any new information about the first point's observable components that could not being determined from the first point's observable components. In that sense, there is no innovations brought by the new spot. Thus, the coupling matrix and the improvement in the retrieval error tends toward zero.

The sounding frequencies are typically unchanged once the sounder is in orbit. Thus, the changes in the weighting matrices between the two-points in the retrieval input region of support (ROS) can only be caused by the lifting of the weighting functions with changes in the sounding angle. These liftings occur perpendicular to the satellite's direction.

This analysis shows that when performing soundings above an atmosphere characterized by a barotropic flow with similar characteristics at each pressure level, two-dimensional filters perpendicular to the track of the satellite will outperform the ones parallel the satellite. The penalty paid for this benefit is in the design of these filters which cannot be recursively computed.

This argument also demonstrates the fact that for sounders with a small number of sounding frequencies (where the energy in the unobservable components

is substantial as compared to the total a-priori energy), the effect of the limb correction of brightness temperature to "equivalent" nadir values can be disastrous because it prohibits the use of changes in the singular value decomposition basis. This reemphasizes the conclusions of section V.3.1.

Since the atmospheric flow is more barotropic in summer than it is in winter (Boer and Sheperd, 1983), the error improvement using multidimensional retrieval operators along the track of the satellite instead of using single-spot operators, will be larger in winter than it is in summer. This result agrees with previous experimental analyses (Toldalagi, 1980, Nathan, 1983).

For the single-spot retrieval problem, setting the vertical covariance kernel to a diagonal matrix with infinite entries allows us to take the limiting case of the LMMSE called the MIE, which is the optimal solution to the retrieval problem when the a priori statistics are unreliable.

One may use the same argument to analyze the limiting behavior of the coupling matrix whenever the vertical covariance is infinitely diagonal and the isobaric correlation is the same at each pressure level. This is a variation of the minimum information solution where the statistical description allows for an optimal multispot solution. The true minimum solution to the retrieval problem has no correlation between adjacent points of the retrieval operator input ROS, thus no coupling matrix can be designed, and the optimal estimator is single-spot for its support.

The proposed analysis finds the optimal solution when the isobaric behavior of the temperature field is known (say from physics) and the vertical statistics are unreliable. Under these assumptions only the isobaric part of the covariance kernel is assumed to be known. The coupling matrix becomes

$$P_{u1} W_{r1}^t K_{TT}^i(\sigma) W_{r2} P_{o2} (P_{o1} W_{r1}^t K_{TT}^i(\sigma) W_{r2} P_{o2})^{-1} \qquad (V.3.73)$$

Should the isobaric behavior be the same at all pressure levels, the matrix further

simplifies to

$$P_{u1}\mathcal{W}_{r1}^{t}\mathcal{W}_{r2}P_{o2}(P_{o1}\mathcal{W}_{r1}^{t}\mathcal{W}_{r2}P_{o2})^{-1} \qquad (V.3.74)$$

The term $P_{u1}\mathcal{W}_{r1}^{t}\mathcal{W}_{r2}P_{o2}$ corresponds to the matrix of the inner products between the SVD basis vectors of the unobservable subspace at point one and the unobservable subspace of two. If this inner product is zero (this would be the case when the two-points are parallel to the track of the satellite), the coupling matrix for the unobservable space becomes null and no improvement occurs when going from one to two-point retrieval operators. The improvement is quadratic in the inner product.

In summary, the improvement in the retrieval error using a two-point rather than a single-spot retrieval operator can be written as an expression of the form $trace\{CKC^{t}\}$ where K is a matrix that is solely a function of the cross-covariance of observed components, and C is a coupling matrix which differs for observable and unobservable components.

The matrix K is a quadratic function of the cross-covariance between points, independent of the single-spot retrieval error for cases of weak correlation between points, and varies inversely proportionally to that error for strongly correlated fields. This analysis shows that for systems with fine sampling grids (such as AMSU, or airborne systems), the reduction of the error for the single-spot retrieval (D matrix) is essential to further improvements. In such cases, if the one-spot estimate is bad the improvement will be minimal.

The coupling matrix for the observable components is a matrix signal to noise ratio for the sounding process at point one and goes to zero in cases of perfect measurements.

The coupling matrix for the unobservable components depends upon the barotropicity of the atmospheric field being estimated and the inner product between SVD basis vectors. It goes to zero for pure two-dimensional flow with char-

acteristics that do not vary in the vertical direction.

For sounding systems with good signal-to-noise ratios, the retrieval operator is narrow in the observable space and does not low pass filter the temperature profile components. For the unobservable components, the behavior of the retrieval operator is not easy to characterize, but is in general wider in the space domain than is its observable counterpart. Thus, it tends to low pass filter the unobservable components of the temperature field to reduce the retrieval error. Moreover, it extends more perpendicularly to the satellite's track than in the parallel direction.

V.3.5 Analysis of the N-Point Retrieval Problem

The analysis of the N-point retrieval problem is a straight forward extension of the 2-point problem and the conclusions presented in the previous section extend directly.

The improvement in error using a $N+1$-point operator over a N point operator can still be written in the form $trace\{CK_{core}C^t\}$. However this improvement decreases, past a certain point, with the number of point used. This is true for two separate reasons: the first one is that as more points are added to the retrieval process, their respective distance increases, the horizontal correlation decreases, and the matrix K_{core} decreases. The second reason is that as the number of points increases, more and more of the weighting functions will be the same. This is true because there is only a finite number of scan positions. Moreover, as the distance between points increases, the covariance become more two-dimensional (Herring, 1980). These two facts combined show that the coupling matrices will also tend towards zero.

V.4 Design Equation for Multidimensional Retrieval Algorithms.

In this section, the equation defining the multiple frame retrieval examined

with more emphasis on the specific case of symmetric retrieval operator.

V.4.1 General Case.

The retrieval operation is written as a convolution of frames along the track of the satellite (time series axis). The size of a retrieval operator (also called input region of support by analogy with image processing) is $M + N - 1$ where $M - 1$ is the number of frames ahead of the estimation and $N - 1$ is the number of frames behind the estimation. This nomenclature is explained by the convolution equation

$$\hat{T}(m) = \sum_{n=-N+1}^{M-1} D(n) T_b(m-n) \qquad (V.41)$$

The situation $M = N = 0$ corresponds to the purely two-dimensional case perpendicular to the satellite's track. When M is equal to zero the retrieval operator operates only on "past" values of the brightness temperature and in that sense is causal. In the opposite case, when N equals zero, the operator uses only "future" values of the brightness temperature and is thus anticausal.

In the above expression, both temperature estimate and brightness temperature are frames. In the case of MSU (with 11 scan angles), these are matrices of size 11×33 for the temperature and 11×3 for the brightness temperature. D is a five index tensor, regardless of the sounding system.

The retrieval error at frame n is given by

$$e(n) = \hat{T}(n) - T(n) = \sum_{n=-N+1}^{M-1} D(n) W T_b(m-n) - T(n) + \sum_{n=-N+1}^{M-1} D(n) \eta(m-n) \qquad (V.4.2)$$

The error criterion to be optimized is found by taking the expected value of the error vector Euclidian norm. After taking all the necessary expectations, the optimization

criterion becomes

$$E = trace\{\sum_{m=-N+1}^{M-1}\sum_{m'=-N+1}^{M-1} D(m)WR_{TT}W^t(m-m')D^t(m') + R(0)\} + trace\{\sum_{m=-N+1}^{M-1} D(m)R_{\eta\eta}(0)D(m) - 2\sum_{m=-N+1}^{M-1} D(m)WR_{TT}(m)\} \quad (V.4.3)$$

where the covariance R_{TT} is a five index tensor and W the three index weighting tensor. The sensing noise is assumed uncorrelated from frame to frame.

The optimal retrieval tensor is found by setting the appropriate derivatives to zero. This yields the set of equations

$$WR_{TT}(m) = \sum_{m'=-N+1}^{M-1} D(m')\left(WR_{TT}(m-m')W^t + R_{\eta\eta}(m-m')\right) \quad (V.4.4)$$

for $m = -N+1, \ldots, M-1$. These equations are the Yule-Walker or normal equations in block form which are common to Autoregressive (AR) models and linear prediction (Oppenheim and Shaffer, 1976).

These equations can be rewritten using the cross-covariance between brightness temperature and temperature profiles. With obvious notations, (V.4.) becomes

$$R_{TTb}(n) = \sum_{m=-N+1}^{M-1} R_{TTb}(m)D(n-m) \quad (V.4.5)$$

This equation expresses the principal of orthogonality between error and observed brightness temperatures (Kailath, 1981, Wiggins and Robinson, 1965). Expressed in matrix form, the Yule-Walker equations become a linear system of equation where the kernel is a block Toeplitz matrix. Namely

$$\begin{pmatrix} R_{TTb}(-N+1) \\ R_{TTb}(-N+2) \\ \vdots \\ R_{TTb}(M-1) \end{pmatrix} =$$

$$\begin{pmatrix} R_{TbTb}(0) & R_{TbTb}(1) & \cdots & R_{TbTb}(N+M-1) \\ R_{TbTb}(-1) & R_{TbTb}(0) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \cdots \\ R_{TbTb}(1-N-M) & \cdots & R_{TbTb}(-1) & R_{TbTb}(0) \end{pmatrix} \begin{pmatrix} D(-N+1) \\ D(-N+2) \\ \vdots \\ D(M-1) \end{pmatrix} \tag{V.4.6}$$

Because R_{TT} is an autocorrelation tensor, when written in matrix form (this is performed by reformatting a frame as a vector) it is symmetric, As shown by

$$R_{TT}(j) = R_{TT}(-j) \tag{V.4.7}$$

but is not block symmetric (Robinson and Treitel, 1978).

As the distance between points increases the covariance between the corresponding temperatures decreases to zero (except for a slight negative overshoot in some cases). As seen in section V.3.2, the error improvement provided by the multidimensional retrieval decreases as the cross-covariance decreases. Thus it is better from a retrieval point of view to avoid incorporating points with little or small correlation with the frame at which the retrieval is being performed. Thus, one must choose the retrieval operator as symmetric as possible with respect to the location where the retrieval is being performed.

Consider the special case of symmetric retrieval operators where $N = M$. These operators have an odd number of frames in their input ROS. Symmetry considerations show that the coefficient of the retrieval operator should be symmetric with respect to the center frame, that is using the time series representation $D(m) = D(-m)$. Introducing the causal part of the retrieval operator (this is related to its Hilbert transform) yields

$$\begin{aligned} D(m) &= D^+(m) && m \geq 0 \\ &= 2D^+(0) && m = 0 \\ &= D^+(-m) && m \leq 0 \end{aligned} \tag{V.4.8}$$

The set of normal equations becomes

$$\begin{aligned} R_{TTb}(n) &= \sum_{m=1}^{N-1} D^{+}(m) R_{TbTb}(n-m) + \sum_{m=-N+1}^{-1} D^{+}(m) R_{TbTb}(n+m) + \\ &\quad 2D^{+}(0) R_{TbTb}(n) \\ &= \sum_{m=1}^{N-1} D^{+}(m) \big(R_{TbTb}(n+m) + R_{TbTb}(n-m) \big) \end{aligned} \tag{V.4.9}$$

In matricial form, this set of equations corresponds to a Toeplitz plus Hankel matrix. This type of linear set of equations can be recursively solved with efficiency. The recursion for inverting this type of matrix is based upon a matricial variation of the Levinson algorithm. It will be pursued in more detail in Chapter VI.

A different interpretation of the solution of the design equation can be found by considering a similarity transformation of the covariance matrix of the brightness temperature over the retrieval operator input ROS. The design equation can be written (with simplified notation) as

$$R_{TTb} = D R_{TbTb} \tag{V.4.10}$$

Consider a similarity transform of the covariance matrix R_{TbTb}. Namely

$$R_{TbTb} = F^{t} \tilde{R}_{TbTb} F \tag{V.4.11}$$

Plugging (V.4.12) into (V.4.11) yields

$$\begin{aligned} R_{TTb} &= F^{t} \tilde{R}_{TbTb} F \\ R_{TTb} F^{t} &= \tilde{R}_{TTb} = (D F^{t}) \tilde{R}_{TbTb} = \tilde{D} \tilde{R}_{TbTb} \end{aligned} \tag{V.4.12}$$

One can represent the different matrices in a different basis and still keep the same type of design equation. This is not a surprising result in view of the observable-unobservable component analysis. The operator F performs a rotation of the brightness temperature and does not affect the estimation process.

From a design point of view this transformation can be helpful if it facilitates the inversion of $R_{T_bT_b}$. Such is the case if the matrix F corresponds to the Karhunen Loeve transformation of the matrix $R_{T_bT_b}$ and thereby diagonalizes it. After transformation, the optimal D matrix is given as

$$D = R_{TT_b}F^t(FR_{T_bT_b}F^t)^{-1}F \qquad (V.4.13)$$

A suboptimal transform F (from a inversion of $R_{T_bT_b}$ point of view) is a "block KL" transform that transforms $R_{T_bT_b}$ into a block diagonal matrix whose blocks have size the product of number of sounding frequencies by the number of locations per frame. This is the size of a "frame". The inverse of this block diagonal matrix is the block diagonal matrix composed of these submatrices' inverses. There is not real incentive for further reduction of the size of these submatrices past that of the size of the covariance of the brightness temperatures of a frame since they are not of any special structure.

Furthermore, it is easier to obtain the eigenvectors of a block KL transform than to obtain the eigenvectors of the Kl transform, thus possibly eliminating suboptimality. This block KL transform can be found using fast algorithms developed in recent years for the inversion of Toeplitz matrices (Kapouptsidis *et al.*, 1983).

Let Φ denote a vector being operated on by $R_{T_bT_b}$. Φ has $N \times M$ components where $2N-1$ is the number of frames of the retrieval operator and M the number of sensors per frame. The number of sensors is defined as the product the number of sensing frequency times the number of scanning position. Each component p of that vector ($p = 0, 1, \ldots, N \times M - 1$) can be uniquely represented by the doublet n, m where $p = nM + m$ ($n = 0, 1, \ldots, N$; $m = 0, 1, \ldots, M$).

The vector Φ will be an eigenvector of $R_{T_bT_b}$ if

$$\lambda\Phi(p) = \sum_{p=0}^{NM-1} R_{T_bT_b}(p,q)\Phi(q) \qquad (V.4.14)$$

for all p. Both indexes p and q can be represented as pairs (n,m) and (l,o) where the first indexes in these pairs pertain to the frame and the second to the sensor within the frame.

The summation defining the KL transform becomes

$$\lambda\Phi(l,o) = \sum_{n=0}^{N-1}\sum_{m=0}^{M-1} R_{TbTb}(l,n,o,m)\Phi(n,m) \qquad (V.4.15)$$

for all (l,o) where

$$R_{TbTb}(l,n,o,m) = R_{TbmTbo}(n-l) + R_{TbmTbo}(n+l) \qquad (V.4.16)$$

The vector Φ will be a block eigenvector of R_{TbTb} if

$$\sum_{o=0}^{M-1}\mu(o)\Phi(l,o) = \sum_{n=0}^{N-1}\sum_{m=0}^{M-1} R_{TbTb}(l,o,n,m)\Phi(n,m) \qquad (V.4.17)$$

The multiplication by the scalar λ has been replaced by the multiplication by a matrix μ.

This submatrix μ is not unique for each one of the blocks. To show this, one constructs a series of block matrices obtained by considering the diagonalized matrix Λ_{TbTb} obtained by projecting by R_{TbTb} onto its eigenvectors, then by performing a rotation along a few of the basis vectors to create a block matrix. To illustrate this, consider the simple case of a four-by-four matrix (with eigenvalues $\lambda_1,\lambda_2,\lambda_3,\lambda_4$) Performing a rotation in the plane of the first two eigenvectors yields

$$\begin{pmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \lambda_3 & 0 \\ 0 & 0 & 0 & \lambda_4 \end{pmatrix} \rightarrow \begin{pmatrix} \nu & -\varsigma & 0 & 0 \\ \varsigma & \chi & 0 & 0 \\ 0 & 0 & \lambda_3 & 0 \\ 0 & 0 & 0 & \lambda_4 \end{pmatrix} \qquad (V.4.18)$$

Both ν,ς,χ depend on the eigenvalues λ_1, λ_2, and the angle of rotation. Since the latter can be chosen arbitrarily, different block diagonal representations are possible and result in different block eigenvectors that can be derived from R_{TbTb}.

To easily determine such block eigenvectors, let us extract a series of $N \times N$ matrices $R_{m,o}$ out of R_{TbTb} by extracting the same element in each block

$$R_{m,o}(l,n) = R_{TbTb}(l,n,o,m) \qquad (V.4.19)$$

Let $\tilde{\Phi}$ denote the matrix whose columns are the N eigenvectors of such an extracted matrix

$$\lambda'\tilde{\Phi}_{m,o}(n) = \sum_{l=0}^{N-1} R_{o,m}(l,n)\tilde{\Phi}_{m,o}(l) = \sum_{l=0}^{N-1} R_{TbTb}(l,n,o,m)\tilde{\Phi}_{m,o}(l) \qquad (V.4.20)$$

The extracted matrix is the sum of a Toeplitz and Hankel matrix whose KL transform can be obtained efficiently (Jain, 1976).

Now compose the vector φ by stacking the M vectors $\tilde{\Phi}_{m,o}$ element by element, block by block according to

$$\varphi(l,o) = \tilde{\Phi}_{m,o}(l) \qquad (V.4.21)$$

They are $N \times M$ vectors (N vectors for each index m and M index m). Plugging (V.4.21) into (V.4.20) yields

$$\sum_{n=0}^{N-1}\sum_{m=0}^{M-1} R_{TbTb}(l,n,o,m)\varphi(n,m) \sum_{m=0}^{M-1}\sum_{n=0}^{N-1} R_{TbTb}(l,n,o,m)\tilde{\Phi}_{o,m}(n) \qquad (V.4.22)$$

Since the vector $\varphi(l,o)$ is an eigenvector of the extracted matrix, the sum becomes

$$\sum_{m=0}^{M-1} \xi(o,m)\tilde{\Phi}_{m,o}(l) = \sum_{m=0}^{M-1} \xi(o,m)\varphi(l,o) \qquad (V.4.23)$$

which is the equation defining the block KL transform.

This method of computation of the block KL transform consists of extracting the same cross-covariance out of every frame used in the computation of R_{TbTb},and packing these values in accordance to the frame location. Then, the eigenvectors of these extracted matrices are computed and packed accordingly.

These KL transforms are taken along the satellite axis for each cross-covariance between sensors at different frames located at the same sensor (frequency, sensing angle, and side of satellite) for each frame. This methods allows us to save the Toeplitz plus Hankel structure of the matrix which facilitates computation. This computation as well as its implications will be discussed in Chapter VI.

V.4.2 Limiting Case: Long Support Retrieval Operators

Let us examine the asymptotic form of the solution of the block diagonalization of the design equation to find the solution to the retrieval operator for long and infinite observation ROS parallel to the track of the satellite.

The equation defining the block KL transform of the design equation is

$$\lambda\Phi(n) = \sum_{m=0}^{N-1} \big(R_{ij}(n-m) + R_{ij}(n+m)\big)\Phi(m) \qquad (V.4.24)$$

for all n where $R_{ij}(p)$ is the crosscorrelation between the i^{th} and the j^{th} brightness temperature of the frame when separated by p frames. As previously discussed, the time series constructed from such correlation does not have to be symmetric with respect to the origin except when the two brightness temperature measurements are done using the same channels (frequency and location within frame). – For the purpose of simplifying the discussion, the index ij is dropped in the following derivation.

Since the brightness temperature is homogeneous, it is stationary along the direction of the satellite. The time series $R(n)$ admits a Fourier transform (Oppenheim and Shaffer, 1975), thus

$$R(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} S(\omega)\exp(j\omega n)d\omega \qquad (V.4.25)$$

The spectrum does not have to be real since the series of covariances is not an even series. However, it has to be Hermitian because of the the series is real.

Using the spectral representation, the equation defining the eigenvectors becomes

$$\Phi(n) = \sum_{m=0}^{N-1} \frac{1}{2\pi} \int_{-\pi}^{\pi} S(\omega) \exp(j\omega(n+m)) + \exp(j\omega(n-m)) d\omega \Phi(m) \qquad (V.4.26)$$

which after invoking Fubini's theorem (inverting the summation) yields

$$\Phi(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} S(\omega) \exp(j\omega n) \Big(\sum_{m=0}^{N-1} (\exp(j\omega m) + \exp(-j\omega m)) \Phi(m) \Big) d\omega \qquad (V.4.27)$$

Let us take for the eigenvector $\Phi(m) = \exp(j\omega_o m)$ in the RHS of (V.4.28) and compute the resulting $\tilde{\Phi}$ in the LHS.

$$\tilde{\Phi}(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} S(\omega) \exp(j\omega n) \Big(\sum_{m=0}^{N-1} \exp(j(\omega - \omega_o)m + \exp(j(\omega + \omega_o))) \Big) d\omega \qquad (V.4.28)$$

The sum in parenthesis can be computed as

$$\Sigma = \exp(j(\omega_o - \omega)) \frac{1 - \exp(j(\omega_o - \omega)N)}{1 - \exp(j(\omega_o - \omega))} + \exp(j(\omega_o + \omega)) \frac{1 - \exp(j(\omega_o + \omega)N)}{1 - \exp(j(\omega_o + \omega))} \qquad (V.4.29)$$

$$\Sigma = \exp(j(\omega_o - \omega)) \frac{N+1}{2} \frac{\sin \frac{N}{2}(\omega_o - \omega)}{\sin \frac{1}{2}(\omega_o - \omega)} + \exp(j(\omega_o + \omega)) \frac{N+1}{2} \frac{\sin \frac{N}{2}(\omega_o + \omega)}{\sin \frac{1}{2}(\omega_o + \omega)} \qquad (V.4.30)$$

The ratio of sine functions can be recognized as modified Dirichlet functions which are often encountered in antenna pattern synthesis (sometimes referred to as the Fejer kernel).

Figure V.4.1 presents the typical shape of such functions. The width of the function, defined as the 3 dB bandwidth of the main lobe is equal to $\Delta\omega = 2\pi/n$. The function repeats itself with a period of 2π. As the number of elements along the satellite track increases (N), the modified Dirichlet function becomes narrower. For N large, equation (V.4.30) becomes

$$\Sigma \approx \frac{2\pi}{N} \delta(\omega + \omega_o) + \frac{2\pi}{N} \delta(\omega - \omega_o) \qquad (V.4.31)$$

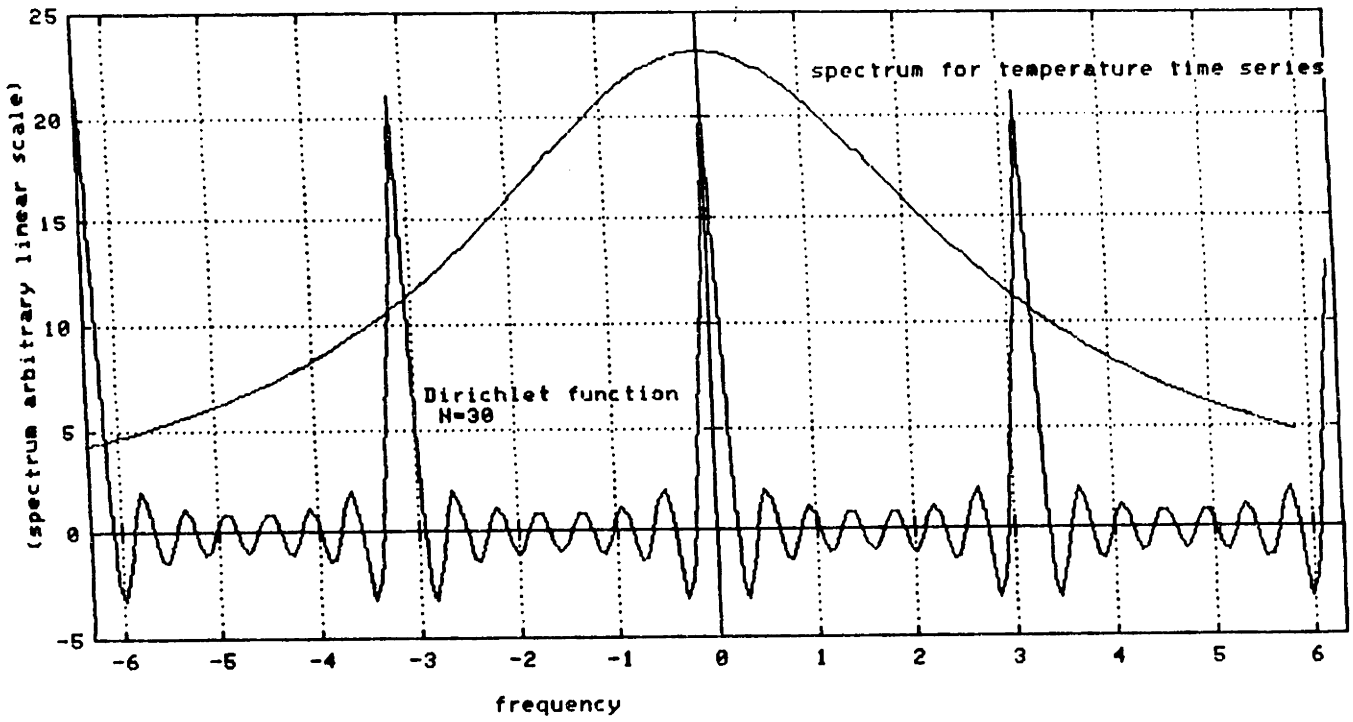


Figure V.4.1: Modified Dirichlet function superposed spectrum of temperature along the track.

Plugging (V.4.31) into (V.4.28) yields

$$\begin{aligned}\tilde{\Phi}(n) =& \frac{1}{2\pi}\int_{-\pi}^{\pi} S(\omega)\left(\frac{2\pi}{N}\delta(\omega-\omega_o)+\frac{2\pi}{N}\delta(\omega+\omega_o)\right)\exp j\omega n d\omega \\ =& \left(\frac{1}{N}S(\omega_o)+\frac{1}{N}S(-\omega_o)\right)\Phi(n)\end{aligned} \qquad (V.4.32)$$

Thus, as the size of the input ROS of the retrieval operator increases, the block KL transform of the kernel of the design equation tends to its Fourier transform.

This is an extension of the result known as SPLOT (Stationary Process Long Observation Time) to discrete retrieval operators (Van Trees, 1968). The eigenvectors which block diagonalize the covariance of the brightness temperature become complex exponentials as the length of the retrieval operator increases. The corresponding eigenvalues are samples of the real part of the spectra of the different

cross-covariances (using the hermitivity of the spectrum).

This approximation improves when the spectrum varies slowly across the length of the Dirichlet impulse. Such is the case if the spectrum is flat across the entire axis. This flat spectrum occurs if the correlation decreases quickly (all things considered) with respect to the interframe distance.

The solution of the infinite (or large) retrieval operator design equation using a Fourier transform of the correlation function is nothing else but the multidimensional Wiener filter which has been proposed first by Rosenkranz (1978) and successfully implemented by Nathan (Nathan, 1983, Nathan et al., 1985).

The approach of the infinite support operator from the finite size retrieval operator shows that finite and infinite size estimators can be derived from the same basis. It also demonstrates that the truncated Wiener filter is not the optimal solution to the finite size retrieval problem.

For infinite size filters the design equation is transformed using a Fourier transform along the satellite axis to yield

$$S_{TT^b}(u) = D(u) S_{T^bT^b}(u) \qquad (V.4.33)$$

where u is a spatial frequency variable corresponding to the distance between frames.

Another limiting case to the design equation where both the Fourier transform and the Karhunen Loeve transform become equivalent is the case of strongly correlated field modeled as a first order vectorial Markovian processes. This special case is relevant to sounding systems with a fine sampling grid and is performed in appendix V.B.

V.5 Conclusions

In this Chapter, we analysed the impact of vertical and horizontal statistics on the retrieval process in the context of mean square error criterion. This analysis

was performed in the natural representation and in the SVD basis representation. Analysing the MSE in the SVD basis demonstrated that performance bounds for observable and unobservable subspaces are different as the signal to noise ratio related to the sounding equation varies Likewise the impact of horizontal correlation was found to differ for observable and unobservable components: In the former case, SNR considerations dictate the impact of horizontal correlation, and the improvement added by multi-spots comes from smoothing the sensor noise. In the later case, the horizontal correlation as well as the vertical variations of this corrclation determine the impact of multidimensional retrievals.

In the later sections and the appendices of this Chapter, the design equation for multidimensional retrieval operators, as well as efficient methods to solve them, are presented. These operators will be implemented in Chapter VI.

Appendix V.A: Kalman Filtering and SVD Decomposition.

The remote sensing of temperature fields based on Kalman filtering has been successfully implemented as evidenced by the reduction of the retrieval error over single spot retrieval techniques (Ledsham 1976, Toldalagi, 1981, Bassili *et al*, 1981). The first two authors implemented the Kalman filtering from a satellite (SCAMS) using the spatial correlation between adjacent points whereas Bassili *et al* considered a ground-based radiometer where the correlation used is temporal and is based on a time series with a sampling period on the order of minutes. Both techniques are related using the Taylor assumption (1935) which relates temporal and spatial correlations.

Kalman filtering is a natural candidate for the remote sensing of temperature by satellites. This is true due to the regular sampling pattern parallel to the satellite, and since it allows the incorporation of horizontal (isobaric) statistics about the satellite without any assumption about the stationary of the underlying atmospheric flow. It is also an efficient method from a computational point of view. Its major drawback, if any, is its assumption that the temperature fields are first order Markovian processes (Gelb, 1974) described by a state space equation. The state is thought of as the temperature field in vectorized form. This description of the field is one dimensional in nature and accounts for the three dimensional aspect of the satellite sampling grid.

This purpose of this appendix is to interpret the implementation of the spatial Kalman filter from satellites for the purpose of remote sensing by analyzing its observable and unobservable components and develop some new insight in the problem. Also, the specific implementation of Ledsham will be discussed in detail.

First, begin by a brief overview of the Kalman filter theory where only the basic equations are presented. The interested reader may refer to the extensive literature (papers aforementioned and references therein) on the subject for possible

interpretations, implementations, and modifications of this filter.

Two important elements in the framework of Kalman filtering included the optimization criterion (mean square error) and the state space description of the dynamic system under analysis (Markovian modeling). The latter allows for the derivation of increasingly more accurate estimates as the filter is used without resorting to extensive computations of those filters.

The model (state space) description of the temperature field along the track of the satellite is given by

$$T_k = \Phi T_{k-1} + w_k \qquad (V.A.1)$$

where the matrix Φ is called the transmission matrix that controls the correlation structure from point to point and w_K is a additive plant noise assumed normal with covariance matrix Q and independent of the temperature. The temperature field is considered to be a stochastic process. The plant noise controls the amount of uncertainty in the new temperature profile T_k. In Ledsham and Toldalagi's works, the temperature profiles are represented by using only 14 pressure levels from the ground up to 10 *mbar* thereby coinciding with the NMC analysis field taken as the ground truth.

The observation equation is the regular radiative transfer equation discretized to the same pressure levels as the ones of the model or, depending on the applications the NMC grid. To simplify the analysis, the sounding locations used in the retrieval are parallel to the satellite's track, such that the weighting matrices remain the same for the entire course of experiments, namely

$$T_{bk} = WT_k + \eta_k \qquad (V.A.2)$$

The observation noise η_k is assumed to be Gaussian $N(0, N)$ and uncorrelated with the plant noise w_k.

The temperature field is stationary (this is not a necessary condition for good

performance of the Kalman filter) if the transition matrix allows a non-infinite solution to the equation

$$P =< T_k T_k^t >= \Phi < T_{k-1} T_{k-1}^t > \Phi^t + Q = \Phi P \Phi^t + Q \qquad (V.A.3)$$

Stationarity puts a upper bound of one on the magnitude of the eigenvectors of the transition matrix Φ.

The Kalman filter is initialized at point zero by taking the a priori mean and covariance of the temperature profile as estimate and error covariance matrix, respectively. Namely,

$$x_0(+) =< T >, \qquad P_o(+) =< (T- < T >)(T- < T >)^t > \qquad (V.A.4)$$

This a priori error covariance matrix equals the matrix P of equation (V.A.3) for stationary fields.

The recursion defining the discrete time consists of two different steps. The first step is the extrapolation phase where the future value of the temperature profile is guessed from previous values(this is denoted by a minus sign). The second step is the updating step where the new measurement of brightness temperature is incorporated in the estimation process (the estimate and error covariance matrices are denoted by a plus sign). To further simplify this analysis, all relevant inverses are assumed to exist (no degenerate cases).

The extrapolation equation for the temperature profile estimate is

$$\hat{T}_k(-) = \Phi \hat{T}_{k-1}(+) \qquad (V.A.5)$$

and for the associated estimation error covariance matrix is

$$P_k(-) = \Phi P_{k-1} \Phi^t + Q \qquad (V.A.6)$$

When the temperature field is stationary the extrapolation error can be written as

$$P_k(-) - P = \Phi (P_{k-1}(+) - P) \Phi^t \qquad (V.A.7)$$

In addition, when the field is stationary, the eigenvalues of Φ are less than or equal to one in magnitude and show that $P_K(-1)$ is closer to P than $P_{k-1}(+)$ independent of the statistics. In such a case, the error in the extrapolation is worse than at the updating step which is expected. The update equations for the estimate are

$$\hat{T}_k(+) = \hat{T}_k(-) + K_k(T_{bk} - W\hat{T}_k(-)) \qquad (V.A.8)$$

where the Kalman gain is

$$K_k = P_k(-)W^t(W P_k(-)W^t + N)^{-1} \qquad (V.A.9)$$

and the associated error covariance matrix is

$$P_k(+) = (I - K_k W)P_k(-) \qquad (V.A.10)$$

The above mentioned equations were expressed in the natural representation of temperature profiles where each component of a profile is the temperature at a specific pressure level. Two new and different representations of the observation process and of the state space description of the temperature profiles can be developed to understand the improvement of the Kalman filter over single spot techniques. One representation is the singular value representation associated with the weighting matrix and the other is the modal description of the state space description. Both concepts illustrate one aspect of the question of observability of the Kalman filter which is at the bases of the analysis of its steady state performance. The observability issue will first be examined as a analysis of the statistical content provided by the model and the observations, then will be related to the more abstract mathematical analysis of control theory.

As it will soon appear in this discussion, the SVD based representation is most suited to study of the propagation of the estimation error through the updating step whereas the modal analysis is more relevant to the analysis through the extrapolation step.

First consider the singular value decomposition of the weighting matrix or observable-unobservable analysis. Now perform a rotation of both temperature profile and brightness temperature to express the Kalman filter's expressions in the observable and unobservable subspaces. These rotations are defined by the SVD of the observation matrix.

$$W = \mathcal{W}_l^t \begin{pmatrix} \Sigma & 0 \end{pmatrix} \mathcal{W}_r \qquad (V.A.11)$$

The temperature profile (state) becomes

$$T \rightarrow T' = \begin{pmatrix} T'_o \\ T'_u \end{pmatrix} = \mathcal{W}_r T \qquad (V.A.12)$$

and the brightness temperature

$$T_b \rightarrow T'_b = \mathcal{W}_l T_b \qquad (V.A.13)$$

In the new basis, the state space description becomes

$$T'_k = \Phi' T'_{k-1} + w'_k \qquad (V.A.14)$$

where the new transition matrix

$$\Phi' = \mathcal{W}_r \Phi \mathcal{W}_r^t, \qquad w'_k \sim N(0, \underbrace{\mathcal{W}_r Q \mathcal{W}_r^t}_{Q'}) \qquad (V.A.15)$$

and the observation equation

$$T'_{bk} = \Sigma T'_{k-1o} + \eta'_k \qquad (V.A.16)$$

where

$$\eta'_k \sim N(0, \underbrace{\mathcal{W}_l N \mathcal{W}_l^t}_{N'}) \qquad (V.A.17)$$

The extrapolation equations become

$$\hat{T}'_k(-) = \Phi' \hat{T}'_{k-1}(+) \qquad (V.A.18)$$

$$P'_k(-) = \Phi' P'_{k-1} \Phi'^t + Q' \qquad (V.A.19)$$

The evolution in the error covariance matrix for the extrapolation process is dictated by the correlation structure of the atmospheric temperature though the matrix Φ' Its relevance will be detailed in a future section on modal representation.

The update equation for the temperature estimate is

$$\hat{T}'_k(+) = \hat{T}'_k(-) + K'_k(T'_{bk} - \Sigma \hat{T}_{ok}(-)) \qquad (V.A.20)$$

and the Kalman gain becomes

$$K'_k = \begin{pmatrix} P'_{uok}(-)\Sigma\big(\Sigma P'_{ok}(-)\Sigma + N'\big)^{-1} \\ P'_{uok}(-)\Sigma\big(\Sigma P'_{ok}(-)\Sigma + N'\big)^{-1} \end{pmatrix} \qquad (V.A.21)$$

Another expression can be obtained by normalizing the sensor noise with respect to temperature rather than to brightness temperature. Hence

$$K'_k = \begin{pmatrix} P'_{uok}(-)\big(P'_{ok}(-) + \tilde{N}'\big)^{-1}\Sigma^{-1} \\ P'_{uok}(-)\big(P'_{ok}(-) + \tilde{N}'\big)^{-1}\Sigma^{1} \end{pmatrix} \qquad (V.A.22)$$

where $\tilde{N}' = \Sigma^{-1} N' \Sigma^{-1}$ is the newly defined sensor equivalent covariance matrix. The error covariance matrix for the update estimate is given by

$$P'_k(+) = \begin{pmatrix} \tilde{N}'\big(P'_{ook}(-) + \tilde{N}'\big)^{-1} P'_{ook}(-) & \tilde{N}'\big(P'_{ook}(-) + \tilde{N}'\big)^{-1} P'_{ouk}(-) \\ P'_{uok}(-)\big(P'_{ook}(-) + \tilde{N}'\big)^{-1}\tilde{N}' & P'_{uuk}(-) - P'_{uok}(-)\big(P'_{ook}(-) + \tilde{N}'\big)^{-1} P'_{ouk}(-) \end{pmatrix} \qquad (V.A.23)$$

The error norm of the estimate is the trace of the above matrix which can be decomposed into the mean square error in the observable and unobservable components. Namely,

$$\begin{aligned} E =& trace\{\tilde{N}'\big(P'_{ook}(-) + \tilde{N}'\big)^{-1} P'_{ook}(-),\} \\ &+ trace\{P'_{uuk}(-) - P'_{uok}(-)\big(P'_{ook}(-) + \tilde{N}'\big)^{-1} P'_{ouk}(-).\} \end{aligned} \qquad (V.4.24)$$

Finally, the estimate itself becomes

$$\hat{T}'_k(+) = \hat{T}'_k(-) + K_k(T'_{bk} - \Sigma \hat{T}'_{ok}(-)) \qquad (V.4.25)$$

As might have been expected from previous experience with the SVD representation, the error in the update estimate (the one involving measurements) behaves differently in the observable and unobservable subspaces.

A matrix-update-error-to-noise ratio determines the error in the observable components. The two errors (extrapolation and measurements) add in parallel (like resistors placed in parallel). Thus, the resulting error is controlled by the smaller of these two errors.

If the sensor noise is small (good SNR), the Kalman filter fully incorporates the measurements and discards the state space description. In such a case, the normalized noise covariance $\tilde{N}' = \Sigma^{-1} N' \Sigma^{-1}$ corresponds to a noise floor in the estimation process. In the case of bad SNR the measurements are discarded and the updating error is the same one as at extrapolation. The Kalman filter now operates in open loop.

The error in the unobservable components depends on the statistics of the cross-covariance between observable and unobservable components. The error in the unobservable subspace is reduced during the updating step by an amount predictable from the measurements. If there is no correlation between the observable and unobservable components, the error remains the same through the integration of new measurements, because no prediction of the unobservable components can be performed. Namely,

$$P_{ook}(+) \approx P_{ook}(-) \qquad (V.A.26)$$

The other representation of temperature profiles useful in understanding the performance of the Kalman filters is modal decomposition. This representation directly relates to the state space description of the isobaric correlation.

The basis for the modal description is the diagonalization of the transition

matrix Φ. Projecting the state space equation onto the eigenvectors of Φ yields

$$T"_{ki} = \lambda_i T"_{(k-1)i} + \eta"_{ki} \qquad (V.A.27)$$

All projections along this bases are denoted by the symbol ". What was a matrix multiplication becomes a series of scalar multiplications by the eigenvalues λ_i of the transition matrix. Each component of the temperature profile (mode) only depends on the same component at the previous point.

Unfortunately, the expression of the different elements of the Kalman filtering equations in the modal basis functions does not provide any new insight into the problem. The exception is the extrapolation step which becomes

$$P"_{kij}(-) = \lambda_i \lambda_j P"_{k-1ij}(+) + Q"_{ij} \qquad (V.A.28)$$

The covariance of the error for the i^{th} and j^{th} modes depends on the same covariance at the previous spot and also on the plant noise covariance. The state space equation that decouples the modes of the temperature field also decouples the error in extrapolating the profiles.

The retrieval error is reduced by the Kalman filter over single spot techniques only if the modes of the temperature field and the observable components are not orthogonal to each other as the filter evolves. This condition can be mathematically expressed using the observability concept of control theory and will be addressed at the end of this appendix. Now, consider a specific implementation of Kalman filtering and analyze how real modeling and system identification problems force some modifications on this idealistic mathematical analysis.

One of the advantages of the Kalman filter is that its elements can be computed in advance (off line). Furthermore, assumptions concerning the statistics of the variable being estimated can be altered to change the behavior of these terms. As for the single spot LMMSE, one can tune the filter to the a-priori knowledge and

confidence one has in the statistics or in the measurements through appropriate choices of $\Phi, Q, N, P(0)(+)$.

Since one is prone to error in the compilation of statistics and the evaluation of the radiative transfer equation, it may be more advantageous from an implementation point of view to consider a conservative description of the atmospheric temperature field in order to avoid excessive reliance on the statistics and its information content.

When using a Kalman filter, one may make the conservative assumption of no correlation between adjacent spots. In this degenerate case, the Kalman filter becomes a single-spot retrieval operator. A slightly less conservative assumption is that of a diagonal transition matrix Φ with constant entries. This is a conservative assumption because any set of orthonormal vectors diagonalizes such a matrix. Thus any set of orthonormal vectors is a modal description of the atmosphere statistics. With such a description, the observable and unobservable components do not interact (mix) with each other from point to point and the incorporation of statistics through the state equation is minimal in that respect. The reduction of retrieval error in the unobservable components by virtue of the isobaric correlation is almost null.

A constant diagonal transition matrix corresponds to a barotropic flow with identical behavior at each pressure level. The analysis of the two point retrieval problem in the SVD basis (section V.3.4) showed that in such a case, the inclusion of horizontal statistics almost does not reduce the retrieval error.

In an extremely conservative approach, Ledsham (1976) took for his transition matrix the identity matrix, thus forcing the variance of the temperature field being estimated to increase as the filter is utilized.

Although the theory of Kalman filters does not consider the following case, there is nothing to prohibit the implementation of the Kalman filter using a different

transition matrix and plant noise covariance than the ones used to derive the series of Kalman gains. This implementation is by no means optimal from a mean-square point of view but may be useful from a least trouble point of view when dealing with temperature fields that present large statistical changes in the vertical direction (tropopause inversion). This is, after all, applied engineering where mathematical elegance must sometimes be sacrificed for the purpose of results (and eventually graduation).

When implementing the filter using real data, Ledsham and Toldalagi choose to model Φ as a tridiagonal symmetric Toeplitz matrix with the sum of the elements along each row equal to or less that one (constant except for the very first one and last one). Such a matrix presents several obvious advantages. It is that it is the closest thing to a diagonal matrix one can achieve. Furthermore, it corresponds to a correlation only between immediately neighboring pressure levels. In that respect, it is the most conservative deviation from minimum-information type correlation (diagonal with identical entries) one can obtain. Finally, it avoids the modeling of the strong anticorrelation across the tropopause inversion by setting the correlation of elements across the tropopause to zero (when they are away from it).

This family of tridiagonal matrices Φ also presents several mathematical properties which makes it a good candidate to model the isobaric cross-covariance. The eigenvalues of such matrices are somewhat evenly distributed between their maximum and minimum values. Let α denote the entry on the main diagonal of Φ, and β denote the term on the adjacent diagonals. The eigenvalues of such a matrix are (Jain, 1979)

$$\alpha + 2\beta \cos \frac{\pi i}{N+1}, \qquad i = 1, \ldots, N \qquad (V.A.29)$$

where N is the size of the matrix(14). With α=0.5 and β=0.25 (Ledsham's values), the eigenvalues are distributed between zero and one which allows for a stationary temperature field.

The eigenvectors of Φ which determine the modes of the temperature field are sinusoidal vertical waves whose frequencies increase with the index of the corresponding eigenvalues (Jain, 1979). Namely,

$$\Phi_i(n) = \sin \frac{\pi i n}{N+1} \qquad (V.A.30)$$

The fact that the eigenvectors increasingly vary as their index increases is a phenomenon common to all covariance matrices.

To fully interpret the impact of approximating the transition matrix as a tridiagonal matrix, compute and analyze the transition matrix based on the model of the three-dimensional covariance of temperature fields.

To obtain Φ, let us consider the crosscovariance between the temperature profiles at two successive points. The state space description is

$$T_k = \Phi T_{k-1} + \eta_k \qquad (V.A.31)$$

This yields for the crosscovariance

$$< T_k T_{k-1}^t >= \Phi < T_{k-1} T_{k-1}^t > \qquad (V.A.32)$$

or its equivalent

$$\Phi =< T_k T_{k-1}^t >< T_{k-1} T_{k-1}^t >^{-1} \qquad (V.A.33)$$

Recalling the model for the three dimensional model

$$K_{TT}(\sigma, p, p') = K_{TT}^p(p, p') K_{TT}^i(\sigma, p, p') \qquad (V.A.34)$$

The distance σ is the distance between the points along the satellite's track, it is the interframe distance. The transition matrix can be recognized as the matrix formed by sampling the normalized isobaric covariance kernel $K_{TT}^i(d, p, p')$ which varies as a function of pressure levels. Table V.A.1 presents the 14×14 transition matrix obtained by fitting the model to summer data over the United States (for

.50	.74	.74	.66	.72	.73	.60	.72	.74	.80	.84	.83	.83	.52
.74	.88	.88	.84	.87	.88	.80	.87	.88	.91	.93	.93	.92	.75
.74	.88	.87	.83	.86	.87	.80	.87	.88	.91	.92	.92	.92	.75
.66	.84	.83	.78	.82	.83	.74	.82	.84	.87	.90	.90	.89	.67
.72	.87	.86	.82	.85	.86	.78	.86	.87	.90	.92	.92	.91	.73
.73	.88	.87	.83	.86	.87	.80	.86	.88	.91	.92	.92	.92	.74
.60	.80	.80	.74	.78	.80	.69	.79	.80	.85	.88	.88	.87	.61
.72	.87	.87	.82	.86	.86	.79	.86	.87	.90	.92	.92	.92	.73
.74	.88	.88	.84	.87	.88	.80	.87	.88	.91	.93	.93	.92	.75
.80	.91	.91	.87	.90	.91	.85	.90	.91	.93	.94	.94	.94	.81
.84	.93	.92	.90	.92	.92	.88	.92	.93	.94	.96	.95	.95	.84
.83	.93	.92	.90	.92	.92	.88	.92	.93	.94	.95	.95	.95	.84
.83	.92	.92	.89	.91	.92	.87	.92	.92	.94	.95	.95	.95	.84
.52	.75	.75	.67	.73	.74	.61	.73	.75	.81	.84	.84	.84	.53

Table V.A.1: Transition Matrix for the Kalman filter based on the Model for the Three dimensional for the Covariance of Temperature Field fitted to Summer over the United States.

1	-0.3629011189638245
2	-4.3662837330649280E-05
3	-1.5677252754023870E-07
4	-1.0203582123794520E-07
5	-6.4488330706680950E-08
6	-3.6640178494876050E-08
7	1.1081944989361550E-08
8	3.7695175641655390E-08
9	8.0363076713509100E-08
10	1.2184365761142290E-07
11	2.6877718414331510E-07
12	8.6262697570081350E-06
13	7.0113235238333600E-03
14	11.85917285710274

Table V.A.2: Eigenvalues of transition matrix fitted to summer over the United States.

more details on this fit, check chapter IV). Table V.A.2 lists the eigenvalues of that matrix.

The transition matrix fitted to the statistics is a full matrix which indicates that the estimates at different pressure levels will be strongly mixed through the state space description if a transition matrix fitted to observed temperature fields is used. Hence, the improvement in error at each of these levels will be difficult to track. Furthermore, any error in the statistical description of the temperature field is likely to be disastrous in terms of the performance of the Kalman filter. This is true, because the behavior of the estimates along the track of the satellite will be dramatically different from the behavior of the real temperature field. More of this analysis of the sensitivity to modeling error is performed in chapter VII. In addition, the largest eigenvalue of Φ is of the order of 10, which implies that the temperature modeled by such a state space equation cannot be stationary and thus tends to have an infinite power as the filter is run. When this occurs, the Kalman filter overestimates the signal-to-noise ratio of the measurements. Any error in the brightness temperature (caused by changes in emissivity, non-linearities in the radiative transfer equation or orographic effects) will result in devastating repercussions.

Let us come back to the simplified description and note how it avoids the problems just mentioned. The modes of the temperature field, as described by the tridiagonal transition matrix, are sine waves in the vertical direction. To analyze the impact of the Kalman filter, one can examine how these modes of the temperature field differ from the basis vectors of the observable and unobservable subspace of the SCAMS observation matrix. The weighting functions for SCAMS are essential those of MSU The SVD basis vectors for MSU and SCAMS are similar in form. Recall that when constructing these vectors, the first three basis vectors are chosen to be the discretized weighting functions and the remainders are sine waves with increasing frequency. These vectors are then orthogonalized using the Gram-Schmidt algorithm. As was previously reported, as the orthonormalization algorithm progresses through the a-priori guessed vectors, it leaves the latter guessed vectors

(rapidly varying sine waves) almost unaltered.

Consider now the different basis vectors that describe the modes of atmospheric temperature. The ones with low frequencies (the ones which carry the most energy) have non zero inner products with the observable basis vectors as well as with the low order unobservable basis vectors. This inner product is the measure of the coupling between those vectors. As the oscillation frequency increases, the modal basis vectors become closer sine waves, resembling the high order unobservable basis vectors. Thus the inner products of the high index modal basis vectors with the observable basis vectors will become negligible. No coupling occurs. Table (V.A.1) presents the normalized magnitude of the inner products (a number between zero and one) between the modal and the observable unobservable basis vectors.

In summary, the observable and low-index unobservable components are mixed with each other as the temperature field evolves according to its state equation. Meanwhile, the unobservable components of high order do not mix with the other components. Thus the rms error in the observable and low-index unobservable components is altered (and hopefully reduced) by the Kalman filtering whereas the error in the high-index unobservable components is left to the vertical statistical description which is insensitive to errors in the isobaric description.

Allowing the rapidly varying (in the vertical direction) modes to remain unaffected by the Kalman filtering parallels the retrieval method of Smith and Woolf (1976) which discriminates between eigenvectors on the basis of their corresponding eigenvalues and does not try to estimate the ones with the smallest eigenvalues (the ones with the fastest vertical oscillations).

The tridiagonal matrix appears now to be a judicious choice for several reasons: First, it yields stable statistics (unlike the Φ derived from the statistics), and second it mixes the low-index unobservable components with the observable ones to reduce the estimation error. Moreover it discards the rapidly varying who do not contribute

modal vector ——————>														
SVD vector														
	0.599	0.022	0.055	0.356	0.065	0.233	0.131	0.247	0.029	0.105	0.180	0.003	0.349	0.156
	0.527	0.669	0.041	0.007	0.066	0.347	0.002	0.130	0.064	0.013	0.027	0.102	0.252	0.129
v	0.445	0.316	0.062	0.252	0.497	0.099	0.465	0.256	0.109	0.190	0.078	0.052	0.077	0.181

Table V.A.3: normalized magnitude of the inner product between modal basis vectors and SVD basis vectors for SCAMS at nadir

to the overall temperature profile (small component magnitude). The tridiagonal description of the transition matrix low-index filters the modes of the atmospheric flow to capture most of the improvement in retrieval error and avoids dealing with

potentially unstable high-index modes.

Combined with Kalman gains computed from a pessimistic description of the atmospheric temperature field, this approach results in a conservative implementation of multidimensional filtering which improves the retrieval error over single spot retrieval methods and is robust to modeling error. This analysis explain why the somewhat ad-hoc implementation of Kalman filtering by Ledsham was bound to succeed. However, because the isobaric correlation is only introduced in the direction parallel to the satellite where all weighting matrices are identical, the improvement will not be as large as for retrieval operators perpendicular to that track.

Before closing this appendix, let us discuss the issue of observability from a more mathematical point of view and further argue the importance of the SVD basis representation.

The issue of observability of the entire temperature profile as the Kalman filters evolves is a mathematical statement of the condition of the weighting function and its observable space, and of the transition matrix and its modal decomposition so that the errors of all components of the profile tend towards steady-state values. Combined with the issue of controllability, the steady state behavior of the Kalman filter can be analyzed though the Ricatti equation (Kailath, 1976).

The system (described by the state space and observation equations) is observable if the following matrix is full rank

$$rank \begin{pmatrix} W \\ W\Phi \\ \vdots \\ W\Phi^{N-1} \end{pmatrix} = N \qquad (V.A.35)$$

where N is the dimension of the vectorial space of the temperature profiles. The different powers of Φ correspond to the mixing of components of the temperature field through the state space description.

The rank of the above mentioned composite matrix is not altered by multiplication by a full rank matrix. Hence, one can express the relationship required to ensure that all components of the temperature profile end up observed in either the SVD or modal basis.

In the SVD basis, the equation becomes

$$rank \begin{pmatrix} (\Sigma \quad 0) \\ (\Sigma \quad 0)\,\Phi' \\ \vdots \\ (\Sigma \quad 0)\,\Phi'^{N-1} \end{pmatrix} = N \qquad (V.A.36)$$

where the symbol $'$ denotes the usage of the SVD basis. The system is fully observable if when the transition matrix is operated upon itself, the diverse modes of the temperature field are pushed in front of the submatrix Σ to be sensed.

The dual explanation is found by using the modal analysis where the equation becomes

$$rank \begin{pmatrix} & W" & \\ W" \begin{pmatrix} \lambda_1 & & 0 \\ & \cdots & \\ 0 & & \lambda_N \end{pmatrix} \\ \vdots \\ W" \begin{pmatrix} \lambda_1^N & & 0 \\ & \cdots & \\ 0 & & \lambda_N^N \end{pmatrix} \end{pmatrix} = N \qquad (V.A.37)$$

where " denotes the modal basis representation.

The eigenvalues λ_i of the transition matrix change the rows of the matrix $W"$ by multiplying each components of these rows by a different power of the eigenvalues. This multiplication corresponds to the propagation of the observable space of the weighting matrix to future sensing locations. Equation (V.A.) clearly shows that if all the eigenvalues of Φ are equal (two dimensional constant flow), the new rows of the matrix $W"$ are scaled versions of the original ones. Thus the matrix will not be full rank and the system will not be observable.

To conclude, the issue of observability (as well as other aspects of the Kalman filtering) can be studied and expressed in three different representations: the natural representation (user oriented) where each component of the temperature profile is the temperature at a specific pressure level. This is a practical representation to produce results but does not permit a good analysis of the Kalman filtering. The modal representation (signal oriented) tends to explain the improvement from the perspective of atmospheric flow and the resulting statistics. Finally, the SVD representation (sounding system oriented) tends to look at things from an observation point of view.

Appendix V.B: Fast Design of Suboptimal Retrieval Operators for a Special Case of Strongly Correlated Temperature Fields.

Fast algorithms to design finite size retrieval operators are possible in instances when the KL transformation required along the track of the satellite (Equations V.4.14 to V.4.23) is easy to compute. This is the case if the support of the retrieval operator is much larger than the correlation lengths related to the isobaric covariance.

Another instance, discussed in this appendix, is the case where the correlation can be modeled as first order Markovian process. It further simplifies if the correlation coefficient associated with such a model is close to one. The proposed design is strongly connected to the transform coding of images (Pratt, 1978).

Before developing the required theory for the computation of such retrieval operators, let us examine the conditions for which such an approximation is valid.

As mentioned, the basis for the fast computation of the KL transform along the track of the satellite is a correlation of the form

$$r(k) = r(0)\rho^{|k|} \qquad (V.B.1)$$

where ρ is the interframe correlation coefficient, k the frame difference and $r(0)$ the value of the correlation for no displacement.

As described in chapter III, the correlation between the temperatures at two locations separated by σ at pressures p and p' is the form

$$K_{pp'} \exp(-\sqrt{\xi(p)\xi(p')}\sigma) \qquad (V.B.2)$$

where ξ are complex poles associated with each pressure levels. These poles have typically larger real part (α) than imaginary part (ω) and can thus be modeled as real. The correlation obeys

$$K_{pp'} \exp(-\alpha_{pp'}\sigma) \qquad (V.B.3)$$

where α denotes the geometric mean of the two α. This is not a restrictive hypothesis if the displacement between the two points is small compared to the scales of fluctuation (introduced in chapter IV) associated with the two pressure levels. Namely

$$\begin{aligned} \sigma &\leq \frac{\alpha(p)}{\alpha(p)^2 + \omega(p)^2} \\ &\leq \frac{\alpha(p')}{\alpha(p')^2 + \omega(p')^2} \end{aligned} \qquad (V.B.4)$$

The distance between two points in the sounding raster located at the i^{th} and the j^{th} sensors separated by k frames is

$$\sigma = \sqrt{\left(pos(i) - pos(j)\right)^2 + \left(d_{frame}k\right)^2} \qquad (V.B.5)$$

where $pos(i)$ is the signed distance from the center of the frame to the i^{th} position an d_{frame} the distance between adjacent frames.

Plugging the distance measure into the decaying exponential and extracting the distance on the interframe displacement yields

$$K_{pp'} \exp(-\alpha_{pp'} d_{frame} k \sqrt{1 + \left(\frac{pos(i) - pos(j)}{d_{frame}}\right)^2 \frac{1}{k^2}}) \qquad (V.B.6)$$

This function behaves like a scalar first order Markovian process along the satellite's track when the perpendicular distance $pos(i) - pos(j)$ is small compared to the interframe distance d_{frame}. This condition is meet when $pos(i) = pos(j)$, that is when the correlation between adjacent points is only considered along the track of the satellite. The proposed method is hence appropriate to design two-dimensional finite size filters parallel to the track of the satellite similar to those implemented by Nathan (1983).

The interframe correlation $\rho_{pp'} = \exp(-\alpha_{pp'} d_{frame})$ will be close to one if the sampling distance is small compared to the fluctuation length. Such a fine raster can be obtained for airborne systems such as MTS and to a lesser extend for AMSU.

Let us reexamine the design equation for the finite size retrieval operator in the case of two dimensional retrieval operators found by means of block KL transform. For each sensing position with respect to the track (corresponding to a sensing angle and specific weighting matrix), the optimal operator is given by

$$D = R_{TTb}F^t(FR_{TbTb}F^t)^{-1}F \qquad (V.B.7)$$

where F is the matrix whose rows are block eigenvalues of the matrix R_{TbTb} developed in section V.4. Because the filter is two dimensional, each matrix $R_{TbTb}(i)$ in equation V.4.6 has for size the number of sounding channels for the instrument (3 for MSU, 12 for AMSU).

To compute the block eigenvectors, one must extract the same covariance (pair of sensors) in each of those submatrices. Assuming that the Markovian assumption on the temperature fields extend to the brightness temperature covariance (this assumption will be examined momentarily), the extracted submatrix has the form

$$R^{(i,j)}_{TbTb} = \begin{pmatrix} R^{(i,j)}_{TbTb}(0) & R^{(i,j)}_{TbTb}(1) & \dots & RTbTb^{(i,j)}(N+M-1) \\ R^{(i,j)}_{TbTb}(-1) & R^{(i,j)}_{TbTb}(0) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ R^{(i,j)}_{TbTb}(1-N-M) & \dots & R^{(i,j)}_{TbTb}(-1) & R^{(i,j)}_{TbTb}(0) \end{pmatrix}$$

$$= R^{(i,j)}_{TbTb}(0) \begin{pmatrix} 1 & \rho_{ij} & \dots & \rho_{ij}^{N+M-1} \\ \rho_{ij} & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \rho_{ij}^{M+N-1} & \dots & \rho_{ij} & 1 \end{pmatrix} \qquad (V.B.8)$$

where i and j pertain to the channels considered at each frame and $R^{(i,j)}_{TbTb}(0)$ the covariance when the two channels are the same location. ρ_{ij} is the interframe correlation coefficient. One recognizes the Toeplitz matrix form so characteristic of filtering problems of scalar markovian processes.

The one step correlation coefficient can be readily computed from the model

by solving the radiative transfer equation in connection with the three-dimensional model to the covariance. Let $W(i,p)$ denote the discrete weight for the i^{th} channel at pressure p. The one frame correlation is given by

$$R^{(i,j)}_{TbTb}(1) = \sum_{p,p'} W(i,p) K_{pp'} \exp(-\alpha_{pp'} d_{frame}) W(j,p') \qquad (V.B.9)$$

If the average isobaric pole $\alpha_{pp'}$ varies slowly with height compared to $W(i,p)$, the equation above can be simplified

$$R^{(i,j)}_{TbTb}(0)\rho_{ij} = R^{-(i,j)}_{TbTb}(0) \exp(-\alpha_{pp'} d_{frame}) \qquad (V.B.10)$$

and the interframe correlation coefficient ρ_{ij} becomes

$$\rho_{ij} \approx \exp(-\alpha d_{frame}) \qquad (V.B.11)$$

A better approximation for the correlation of the brightness temperature can be obtained using a Taylor expansion of the exponential term. It is

$$\rho_{ij} \approx \exp(-\sum_{p,p'} W(i,p) \alpha_{pp'} W(j,p) d_{frame}) \qquad (V.B.12)$$

At this point, one might question the validity of this approximation on two grounds: The first one is in terms of the differences between the true covariance of brightness temperature along the track of the satellite and the one derived from the Markovian representation of temperature fields and the solving of the radiative transfer equation. The second one is in terms of the accuracy of such an approximation and its impact on the future computation of retrieval operators.

Figure V.B.1 (from Nathan, 1983) presents the observed covariance for channel 3 of MSU (mid height) at nadir as function of the frame displacement along the track. For small displacement ($\sigma \leq 0.5Mm$), the exponential model is well suited. Hence, the approximation is appropriate. It should be noted that the curve plotted was compiled using a training set covering the northern hemisphere during early

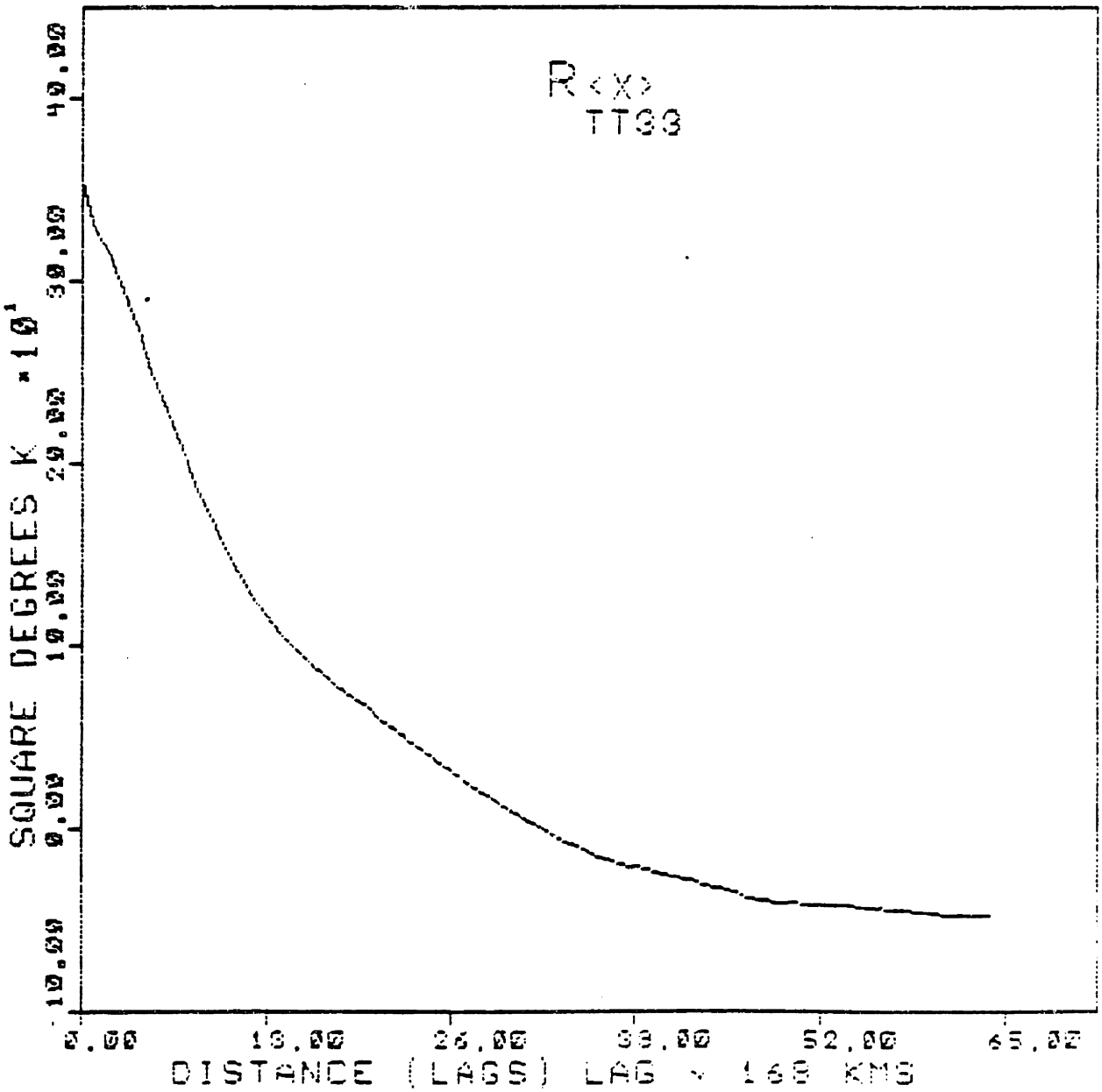


Figure V.B.1: Empirical covariance of the brightness temperature of channel 9 of MSU (from Nathan, 1989)

July 1979. As discussed in chapter IV and chapter V, the variations of the isobaric poles ($\xi = \alpha + j\omega$) during the summer (and the summer of 1979 in particular) as a function of height are smooth. Hence the factorization of the exponential term in

equation V.B.10 is justified.

To answer the second criticism, one must analyze the performance of the KL transform when the correlation coefficient is mispecified. Clarke (1983) performed such an analysis and quantified the decorrelation efficiency as a function of assumed and real correlation coefficient. This efficiency is defined in terms of the off diagonal terms of the original and transformed R_{TbTb} matrices. The efficiency is defined as one minus the ratio between the off diagonal taxicab norm of the final matrix to the one of the original matrix. It equals one if the KL transform is derived using the correct correlation coefficient ρ_{ij} and zero if the KL transform is the identity (no transformation). Over the range $0.9 \leq \rho \leq 1.0$, the efficiency of the transformation is always greater than 98 % when designed with a correlation coefficient somewhere in the same range.

Hence the modeling of the brightness temperature covariance matrix as a Toeplitz matrix is appropriate.

Noe that the ground work has been performed and the validity of the time series Markovian established, let us compute the optimal retrieval operator in the case of high correlation coefficient. This computation involves the inversion of the inverse covariance tensor $FR_{TbTb}F^t$ which is a block diagonal matrix. Its inverse is found by taking the matrix formed of in the inverse of each block. The computation also involves the inverse transform of such a matrix. The block-eigenvectors of F are found by concatenating the eigenvectors of the all R_{TbTb} matrices described in V.4.21.

These eigenvectors take a simple form when the size of the retrieval operator P ($P = N+M-1$ with the notation of section V.4.1) is even. This case corresponds to non symmetric filters. To reduce the retrieval error as much as possible, these operators are made one frame longer in the causal or in the anticausal part than in the other one. The n^{th} eigenvector associated with the extracted submatrix

characterized by $R^{(i,j)}_{TbTb}(0)$, ρ_{ij} is (Ray and Driver, 1970, Ahmed and Flickner, 1982)

$$F_n = \begin{pmatrix} \sqrt{\frac{2}{P+\lambda_{nij}}} \sin\left\{\omega_{nij}[0 - (\frac{P-1}{2})] + (n+1)\frac{\pi}{2}\right\} \\ \vdots \\ \sqrt{\frac{2}{P+\lambda_{nij}}} \sin\left\{\omega_{nij}[k - (\frac{P-1}{2})] + (n+1)\frac{\pi}{2}\right\} \\ \vdots \\ \sqrt{\frac{2}{P+\lambda_{nij}}} \sin\left\{\omega_{nij}[P-1 - (\frac{P-1}{2})] + (n+1)\frac{\pi}{2}\right\} \end{pmatrix} \qquad (V.B.13)$$

$\lambda_{nij} R^{(i,j)}_{TbTb}(0)$ is the corresponding eigenvalue where

$$\lambda_{nij} = \frac{1-\rho^2_{ij}}{1 - 2\rho_{ij}\cos\omega_{nij} + \rho^2_{ij}} \qquad (V.B.14)$$

and the frequency ω_{nij} is the solution of the transcendental equation

$$\tan(P\omega_{nij}) = \frac{(\rho^2_{ij} - 1)\sin\omega_{nij}}{\rho^2_{ij}\cos\omega_{nij}} \qquad (V.B.15)$$

Once the correlation coefficient and the DC value has been found for a pair of channels, the corresponding eigenvalue is readily computable (or looked up in a table) and the matrix $FR_{TbTb}F^t$ is computable with few multiplications. It inversions requires a series of $P = N + M - 1$ inversions of $p \times p$ matrices for a sounding system with p channels.

The inverse transformation is more time (cpu time) consuming. This is the place where the high correlation from point to points enters into account. As the correlation ρ_{ij} approaches unity, the KL transform can be approximated by the Discrete Cosine Transform (DCT) (Clarke, 1981).

In terms of energy compaction, the DCT is not as optimal as other transform based on Fourier series (Unser, 1984) but it presents the advantage to be readily computable using a FFT (Oppenheim and Shaffer, 1975). This DCT is applied to the submatrices extracted from the transformed R_{TbTb} similarly to the forward extraction.

Finally, the D tensor is computed by multiplying R_{TTb} by the matrix just computed.

To clarify this procedure, let us summarize its principal steps

- For each pair of channels (i,j), determine $R^{(i,j)}_{TbTb}(0)$ and ρ_{ij}.
- Compute the block matrix $FR_{TbTb}F^t$ is computed by evaluating a set of λ_{Nij} according to (V.B.).
- The block matrix is inverted block by block (DPPFA, DPPDI in the LINPACK software package).
- The inverse transformation is calculated by a series of DCT computed using FFT's.
- The matrix R_{TTb} is multiplied by the inverse matrix.

This method of computation of suboptimal filters may not be as efficient as those methods based on Levinson type recursions (Kalouptsidis *et al.*, 1983, Levi *et al.* but it is much easier to code. Moreover, it demonstrates that computation of retrieval operator based on Fourier transform are possible for two different limiting cases. The form is when the support of the retrieval operator is much longer than any scale of fluctuation associated with the temperature field and the discrete Wiener filter can be used. The latter is when the sampling distance is much smaller than the scales of fluctuations, the correlation of brightness temperatures obeys an approximate Markovian description, and the DCT can be used.

CHAPTER VI
COMPUTATION AND IMPLEMENTATION OF MULTIDIMENSIONAL RETRIEVAL OPERATORS

-We are 106 miles from Chicago, we have a full tank of gas and half a pack of cigarettes.

-Hit it!

Dan Aykroyd, John Belushi, "The Blues Brothers"

VI.1 Introduction

In this chapter, multidimensional retrieval operators based on the model for the three-dimensional covariance kernel are computed, then implemented. The ground truth used to assess the performance of the retrieval operators is the National Meteorological Center (NMC) analysis field. It is unreliable ground truth; it is not available at all pressure levels and thus does not permit to characterize the retrieval process in the SVD basis, which is, in view of chapter V's results, the appropriate method to characterize multidimensional retrieval operators. Moreover, because of possible errors in the weighting functions caused by changes in ground reflectivity and altitude and by non-linearities in the radiative transfer equation, the forward problem is not solved accurately and the observation equation may be flawed. Hence, it is difficult to truly assess the impact of multidimensional retrieval operators relying solely on experimental data. The dissertation's implementation will be performed first on synthesized data (using a short-cut) and then on measured data (after appropriate corrections). For the synthesized data, the error characterization associated with different experiments will be estimated according to the formulas developed in Chapter V and analyzed accordingly. Since atmospheric temperature is a non-stationary stochastics process, conclusions about the impact of multidimensional operators cannot be characterized analytically and

require implementation on real case studies. The retrieval error will hence vary from experiment to experiment, depending on the number of degrees of freedom associated with each experiments. For these reasons, several error criterion associated with the rms and bias error will be introduced that render the error characterization less set dependent. To quantify the impact of season, both expected and experimental errors were considered for summer and winter climatologies. In all cases, the improvement provided by the introduction of vertical then three-dimensional statistics will be quantified and its variations analyzed.

VI.2 Retrieval Operators

VI.2.1 Computation

In the derivation presented, all elements are composed out of matrices whose sizes are determined by the number of points and sounding frequency used within a frame. To simplify this presentation, scalar matrices are denoted by small letters, block vectors are overlined and block matrices are overlined twice. The block transposition is denoted by a capital T whereas the traditional element transposition is denoted by a small t. For each increase in the size of centered retrieval operators the recursion is run twice.

As discussed in Chapter V, the design equation for three-dimensional retrieval operators are the Yule-Walker equations

$$\overline{\overline{DK}}_{TbTb} = \overline{K}_{TTb} \qquad (VI.2.1)$$

where the retrieval operator is defined as

$$\overline{D} = (D(-N+1) | D(-N+2) | \ldots | D(M-1)) \qquad (VI.2.2)$$

where the indexes $N-1, \ldots, M-1$ pertain to the frame position, and the relevant statistics about the brightness temperatures and physical temperatures are (using

notations introduced in Chapter V)

$$\overline{K}_{TTb} = (K_{TTb}(-N+1) | K_{TTb}(-N+2) | \dots K_{TTb}(M-1)) \qquad (VI.2.3)$$

and

$$\overline{\overline{K}}_{TbTb} = \begin{pmatrix} K_{TbTb}(0) & K_{TbTb}(1) & \dots & K_{TbTb}(N+M-1) \\ K_{TbTb}(-1) & K_{TbTb}(0) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \dots \\ K_{TbTb}(1-N-M) & \dots & K_{TbTb}(-1) & K_{TbTb}(0) \end{pmatrix} \qquad (VI.2.4)$$

For sounding systems with frames with P sounding positions and C channels, The design of an N frame operator requires the inversion of the $PCN \times PCN$ matrix $\overline{\overline{K}}_{TbTb}$. This inversion requires $O(P^3C^3N^3)$ operations (multiplications and additions). Using the block Toeplitz nature of the covariance matrix in (VI.2.4) permits to invert this matrix in $O(P^3C^3N^2)$ operations (Wiggins and Robinson, 1965, Frielander and Morf, 1980). This presentation follows the one of Kaloupsidis *et al.* (1983). It is an extension to the multichannel case of the Levinson-based recursion of scalar signal processing.

The entire retrieval operator (of length m) can be written as

$$\overline{D}^T_{m+1} = (\overline{D}^T_m | 0) + k^d_{m+1} (\overline{v}^T_m | i) \qquad (VI.2.5)$$

where i denotes the identity matrix, and the innovation operator $\overline{v}_m$ is found using the two elements recursion

$$\begin{aligned} \overline{v}^T_{m+1} \overline{\overline{J}} &= (v^T_m \overline{\overline{J}} | 0) + k^v_{m+1} (\overline{g}^T_m | i) \\ \overline{g}^T_{m+1} \overline{\overline{J}} &= (g^T_m \overline{\overline{J}} | 0) + k^g_{m+1} (\overline{v}^T_m | I) \end{aligned} \qquad (VI.2.6)$$

The innovation operator correspond to the coefficients of the new operator than cannot be predicted from previous operators. The updating gains k (similar to the Kalman gains in Kalman-Bucy filtering) which controls the changes between

successive operators obey

$$\begin{aligned} k^d_{m+1} &= -\alpha^d_m {\beta^d_m}^{-1} \\ k^v_{m+1} &= -\alpha^d_g {\beta^v_m}^{-1} \\ k^g_{m+1} &= -\alpha^d_g {\beta^d_m}^{-1} \end{aligned} \qquad (VI.2.7)$$

The intermediate gains α and β directly depend on the covariance matrices between brightness temperature and cross-covariance brightness temperature physical temperature as a function of the frame difference. Namely

$$\begin{aligned} \alpha^g_m &= \overline{g}^T_m \overline{\overline{J}}\, \overline{K}_{TbTbm} + K_{TT}(m) W^t \\ \alpha^d_m &= \overline{d}^T_m \overline{K}_{TbTbm} + K_{TT}(m) W^t \end{aligned} \qquad (VI.2.8)$$

and

$$\begin{aligned} \beta^d_{m+1} &= \beta^d_m + k^v_{m+1} \beta^g_m \\ \beta^v_{m+1} &= \beta^v_m + k^g_{m+1} \beta^g_m \end{aligned} \qquad (VI.2.9)$$

where the covariance vector $\overline{K}_{TbTbm}$ is defined as the concatenation of covariances between brightness temperatures across several frames according to

$$\begin{aligned} \overline{K}_{TbTbm} = &(K_{TbTb}(m), K_{TbTb}(m-1), \ldots, K_{TbTb}(0)) \\ &(W K_{TT}(m) W^t, W K_{TT}(m-1) W^t, \ldots, W K_{TT}(0) W^t) \end{aligned} \qquad (VI.2.10)$$

The block matrix $\overline{\overline{J}}$ is the counter-identity block matrix whose entries are the identity matrices correspond to the frame size (PC)

$$\overline{\overline{J}} = \begin{pmatrix} 0 & \ldots & \ldots & 0 & i \\ 0 & & & & 0 \\ \vdots & & i & & \vdots \\ 0 & & & & \vdots \\ i & 0 & \ldots & \ldots & 0 \end{pmatrix} \qquad (VI.2.11)$$

Every odd iteration, the retrieval operator is symmetric and can be expressed as a function of its causal part according to equation V.3.

Note that, for any pressure level, at any point in the recursion, the gain k^d_m goes to zero for an entire row, the retrieval operator will remain unchanged at that iteration and at all successive iterations.

VI.2.2 Analysis and Characterization

In this section, the form of the multidimensional retrieval operators is examined with regards to Chapter V's analysis. This characterization will not be exhaustive for such an analysis would be too lengthy (retrieval operators are five-index tensors) and tedious.

As noted in Chapter V, multidimensional retrieval operator tend to present a peak at the sounding location. This is true because the "on-spot" coefficients are larger than "off-spot" coefficients. This peak is all the more visible whenever the proportion of signal in the observable space of the instrument is large and/or when the sounding SNR is large. Table VI.2.1 presents the coefficients of a single-spot retrieval operator operating at nadir for the 12 pressure levels of the NMC analysis grid as well as the corresponding single-frame (two dimensional perpendicular to the satellite's track) operator. The operators are derived from summer statistics and uses the three temperature sounding channels of MSU. This summer statistics as well as all statistics used in the remainder of this thesis is found by fitting the three-dimensional covariance kernel to NMC observational data which is obtained without any corrections of the measurements, except for a possible extrapolation of the profile close to the ground according to a subadiabatic lapse rate when sea-level pressure temperature is missing. This statistics is finer than the NMC field used for comparison purposes. Since the retrieval is performed at nadir which is at the center of the sampling frame, only six out the eleven positions of the MSU sampling grid need to be presented for the two-dimensional operator. The remaining five positions are found by symmetry with respect to the center position. Each line of the table represents the coefficient by which to multiply the brightness temperature measured at a given frequency (the sensing frequency is given on the left-most side of the line) to estimate the physical temperature at nadir for the indicated pressure level. The coefficients are listed according to their positions in the sampling grid. That is, 56°

is on the left-most position whereas nadir on the next to right-most position. Note that the single-spot retrieval operator coefficients (listed on the extreme right) are greatly altered by the introduction of horizontal correlation.

The coefficients listed are those of the causal part of the retrieval operator (see equation (V.4.8)). The largest coefficients for each pressure levels are usually found at nadir. Note that because of the implementation, each coefficient of the center frame, which is the only frame for single spot and single frame retrieval operators, will be multiplied by two before multiplying it by the corresponding brightness temperature.

Grouping the coefficients associated with the same channels allows us to generate the three-dimensional plots of Figures VI.2.1 and VI.2.2. In these plots, the coefficients are plotted versus position in the sounding frame and versus estimation pressure. In the first case, the estimation is performed at nadir and the plots are symmetric with respect to the satellite's track. In the second plot, the estimation is performed at the edge of the sounding grid. Note the largest entries for the three channels correspond to the altitude where the weights peak (at nadir, 600 *mbar* for 53.73 GHz, 270 for 54.96, and 80 for 57.95).

Tables VI.2.2 and VI.2.3 presents the three-dimensional three frame retrieval operator associated with the same statistics. The center frame and the off center frame are presented. Figure VI.2.3 presents the same retrieval operator as a three-dimensional plot. The operator presented in the plot is a three-frame operator estimating the temperature at 500 *mbar* at nadir. The retrieval operator is spiky in nature.

The differences in the retrieval operators in the observable and unobservable components are most apparent when performing the retrieval in those bases. Table VI.2.4 presents the single spot and single frame retrieval operators described above for the observable components, the first three and last three unobservable compo-

						1000 mbar	
53.74 GHz	-0.124	-0.203	-0.170	-0.093	0.062	1.989	1.713
54.96 GHz	0.156	0.269	0.194	0.035	-0.303	-1.301	-1.497
57.95 GHz	-0.122	-0.171	-0.080	0.063	0.285	0.107	0.568
						850 mbar	
53.74 GHz	0.002	0.034	0.043	0.040	0.082	0.834	0.953
54.96 GHz	0.026	-0.025	-0.046	-0.045	-0.112	-0.440	-0.612
57.95 GHz	-0.053	-0.017	-0.019	-0.042	-0.013	0.464	0.160
						700 mbar	
53.74 GHz	0.040	0.071	0.062	0.044	0.047	0.513	0.692
54.96 GHz	-0.049	-0.096	-0.079	-0.046	-0.033	-0.020	-0.182
57.95 GHz	0.008	0.046	0.022	-0.014	-0.029	0.109	-0.019
						500 mbar	
53.74 GHz	-0.025	0.024	0.034	0.030	0.009	0.257	0.393
54.96 GHz	0.047	-0.037	-0.038	-0.018	0.028	0.202	0.207
57.95 GHz	0.023	0.074	0.078	0.066	0.006	-0.622	-0.239
						400 mbar	
53.74 GHz	0.056	0.060	0.038	0.016	-0.009	0.214	0.332
54.96 GHz	-0.074	-0.082	-0.037	0.004	0.064	0.323	0.353
57.95 GHz	0.081	0.074	0.035	-0.001	-0.054	-0.355	-0.299
						300 mbar	
53.74 GHz	0.084	0.054	0.023	-0.002	-0.036	0.158	0.219
54.96 GHz	-0.151	-0.078	-0.022	0.022	0.113	0.610	0.697
57.95 GHz	0.129	0.059	0.005	-0.036	-0.090	-0.294	-0.396
						250 mbar	
53.74 GHz	-0.040	-0.015	0.001	0.006	-0.051	-0.091	-0.065
54.96 GHz	-0.039	-0.035	-0.048	-0.044	0.103	1.079	1.012
57.95 GHz	0.060	0.057	0.073	0.084	0.013	-0.840	-0.415
						200 mbar	
53.74 GHz	-0.009	-0.085	-0.082	-0.070	-0.086	-0.191	-0.547
54.96 GHz	-0.017	0.141	0.118	0.096	0.185	0.968	1.458
57.95 GHz	0.001	-0.135	-0.123	-0.098	-0.068	0.313	-0.203
						150 mbar	
53.74 GHz	-0.093	-0.069	-0.034	-0.013	0.014	-0.076	-0.287
54.96 GHz	0.178	0.108	0.043	0.012	-0.007	-0.005	0.165
57.95 GHz	-0.227	-0.129	-0.061	-0.019	0.064	0.810	0.508
						100 mbar	
53.74 GHz	-0.052	0.000	0.008	0.011	0.033	-0.033	-0.017
54.96 GHz	0.177	0.014	0.001	-0.004	-0.051	-0.323	-0.283
57.95 GHz	-0.150	0.012	0.032	0.040	0.086	0.501	0.716
						70 mbar	
53.74 GHz	0.043	0.019	0.006	0.001	0.001	-0.052	0.034
54.96 GHz	-0.106	-0.033	-0.006	0.004	0.016	0.094	0.051
57.95 GHz	0.110	0.035	0.010	0.004	0.023	0.310	0.486
						50 mbar	
53.74 GHz	0.052	0.019	0.008	0.005	0.008	0.040	0.139
54.96 GHz	-0.150	-0.034	-0.014	-0.007	0.005	0.098	0.001
57.95 GHz	0.141	0.021	-0.002	-0.009	0.009	0.331	0.417

Table VI.2.1: Single-spot and single-frame retrieval operators for MSU at nadir over summer statistics

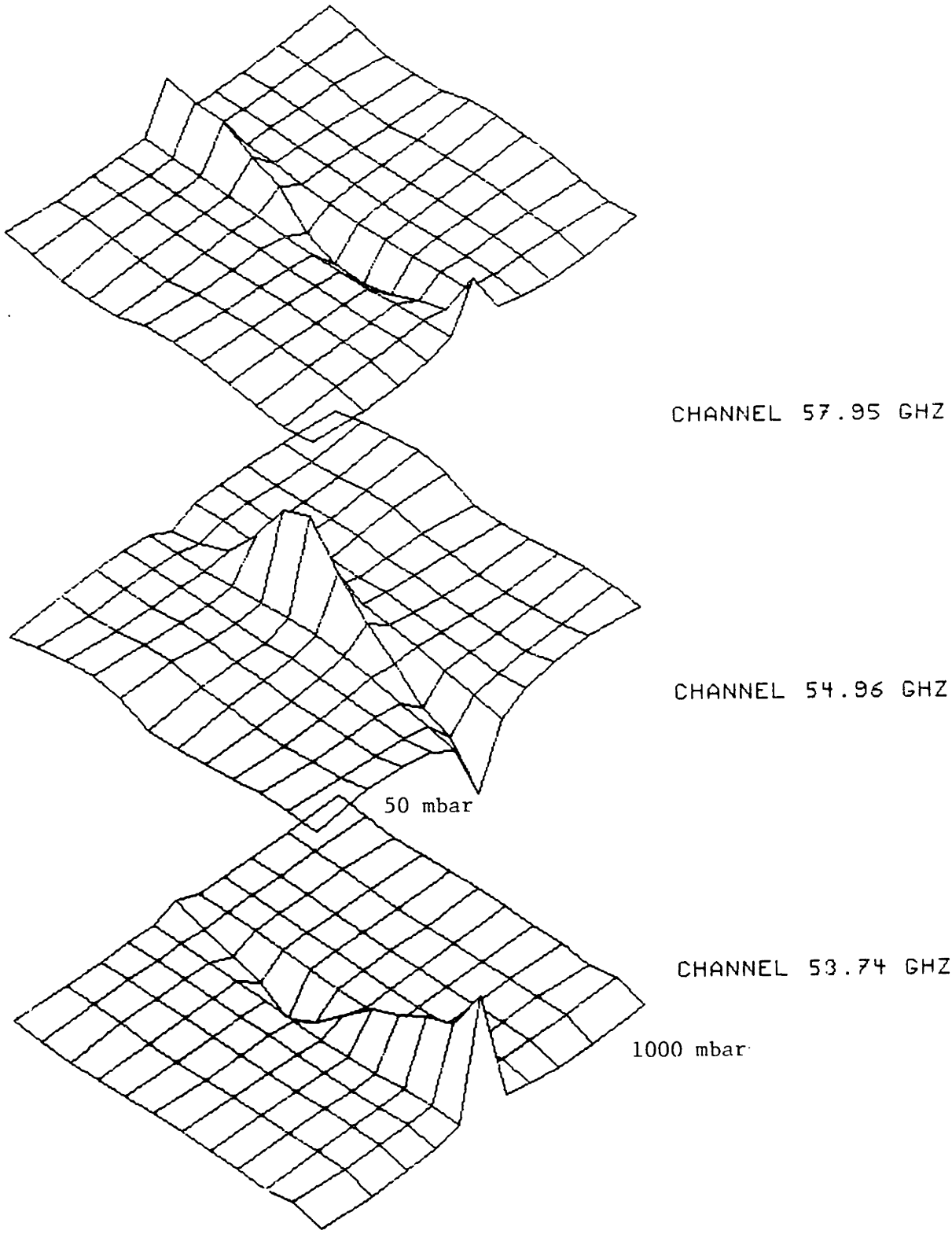


Figure VI.2.1: single-frame retrieval operator as a function of estimated pressure level for MSU at nadir

nents (the first three bases exhibit slow variations in the vertical direction, whereas the last three ones have fast variations).

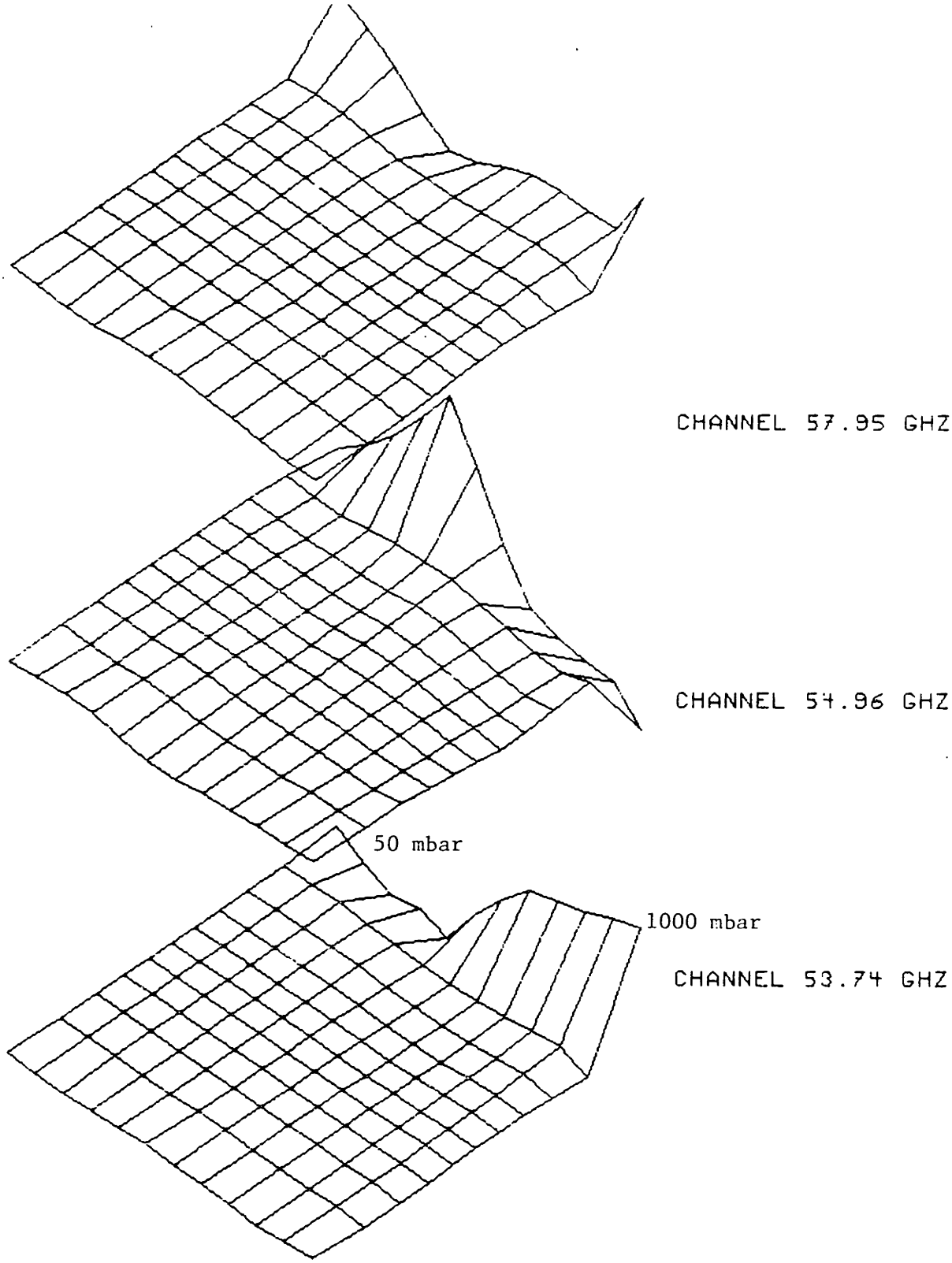


Figure VI.2.2: single-frame retrieval operator as a function of estimated pressure level for MSU for the extreme angle

Figure VI.2.4 present the same operator than Figure VI.2.1 except that the pressure axis is replaced by an index axis along the SVD basis vectors. The first three lines correspond to the observable components whereas the six last lines corre-

						1000 mbar
53.74 GHz	-0.010	-0.019	-0.029	-0.034	0.078	2.251
54.96 GHz	0.019	0.041	0.056	0.045	-0.205	-1.433
57.95 GHz	-0.011	-0.040	-0.046	-0.042	0.077	0.684
						850 mbar
53.74 GHz	-0.008	-0.012	0.002	0.017	0.055	0.540
54.96 GHz	0.018	0.024	-0.001	-0.027	-0.085	-0.209
57.95 GHz	-0.015	-0.009	0.007	0.023	0.051	0.040
						700 mbar
53.74 GHz	0.004	0.007	0.010	0.013	0.023	0.376
54.96 GHz	-0.006	-0.015	-0.018	-0.020	-0.022	0.069
57.95 GHz	0.002	0.018	0.017	0.018	0.016	-0.134
						500 mbar
53.74 GHz	-0.001	0.013	0.010	0.007	-0.003	0.320
54.96 GHz	0.000	-0.023	-0.018	-0.010	0.027	0.309
57.95 GHz	0.001	0.008	0.007	0.003	-0.025	-0.404
						400 mbar
53.74 GHz	0.009	0.011	0.008	0.005	-0.007	0.219
54.96 GHz	-0.020	-0.025	-0.015	-0.006	0.037	0.338
57.95 GHz	0.014	0.017	0.011	0.006	-0.023	-0.369
						300 mbar
53.74 GHz	0.015	0.007	0.004	0.001	-0.019	0.160
54.96 GHz	-0.037	-0.018	-0.009	0.000	0.064	0.513
57.95 GHz	0.028	0.016	0.009	0.004	-0.031	-0.380
						250 mbar
53.74 GHz	0.002	0.007	0.004	-0.005	-0.051	0.053
54.96 GHz	-0.018	-0.009	-0.005	0.008	0.129	1.016
57.95 GHz	0.018	-0.001	-0.003	-0.011	-0.065	-0.467
						200 mbar
53.74 GHz	-0.001	-0.027	-0.019	-0.015	-0.041	-0.190
54.96 GHz	-0.001	0.052	0.034	0.027	0.103	0.737
57.95 GHz	0.008	-0.033	-0.020	-0.014	-0.016	0.187
						150 mbar
53.74 GHz	-0.017	-0.012	-0.006	-0.003	0.006	-0.139
54.96 GHz	0.047	0.024	0.011	0.003	-0.010	-0.027
57.95 GHz	-0.041	-0.012	-0.004	0.000	0.043	0.703
						100 mbar
53.74 GHz	-0.017	0.006	0.002	0.003	0.016	-0.026
54.96 GHz	0.042	-0.007	-0.003	-0.004	-0.029	-0.143
57.95 GHz	-0.042	0.005	0.000	-0.001	0.037	0.589
						70 mbar
53.74 GHz	0.008	0.003	0.000	0.000	0.002	-0.011
54.96 GHz	-0.020	-0.009	-0.002	0.000	0.003	0.047
57.95 GHz	0.018	0.003	-0.001	-0.003	0.014	0.321
						50 mbar
53.74 GHz	0.014	-0.003	0.000	0.001	0.005	0.025
54.96 GHz	-0.035	0.002	-0.001	-0.003	-0.002	0.045
57.95 GHz	0.034	-0.002	0.002	0.003	0.018	0.273

Table VI.2.2: Center frame for three frames retrieval operators for MSU at nadir over summer statistics

spond to the 4,8,11,14,18, and 21^{st} unobservable components as defined in Chapter III. In this plot, the retrieval components have been raised to the $7/10^{th}$ power to enhance the off-spot coefficients.

As noted in Chapter V, for a given region of support, the retrieval operator extends more for the unobservable components than it does for the observable components. This is visible in Figure VI.2.4 where the sidelobes are present for the unobservable components and almost non-existant in comparison for the observable components. This result is also visible when comparing, for the nadir estimation retrieval operator, the coefficients the 56^o brightness temperatures measurements (the measurements the further away from the point of estimation) are multiplied with. All entries for the observable components are equal to zero, whereas for the unobservable components, these coefficients are comparable to the coefficients multiplying the nadir brightness temperatures.

The variations of the coefficients along each row of the table differ also for observable and unobservable components. At each sensing frequency, the coefficients' amplitudes monotonically decrease as the distance between the location where the sounding is being performed (nadir on the right of the table) and the measurements' location increases. Also, the amount of negative or positive overshoot, if any, is small. This behavior is characteristic of smoothing filters encountered in scalar signal processing where a low-pass signal is observed corrupted by an additive white noise. The retrieval operator is basically a low-pass filter for the observable components that improves retrieval error by smoothing the sensor noise out.

For the unobservable components, such behavior is not present. The entries' magnitudes at 56^o can be the largest off-spot entries (see for instance the last unobservable component row for 54.96 GHz). Likewise, the amount of negative or positive overshoot can be substantial which indicates that the retrieval operator is not low-pass in nature. The unobservable components retrieval operator is all-pass.

						1000 mbar
53.74 GHz	-0.098	-0.137	-0.159	-0.164	-0.097	0.008
54.96 GHz	0.116	0.167	0.181	0.159	-0.021	-0.246
57.95 GHz	-0.017	-0.095	-0.106	-0.096	-0.009	0.098
						850 mbar
53.74 GHz	0.001	0.004	0.045	0.083	0.126	0.158
54.96 GHz	0.042	0.052	-0.019	-0.083	-0.153	-0.200
57.95 GHz	-0.083	-0.060	-0.007	0.037	0.077	0.101
						700 mbar
53.74 GHz	0.038	0.052	0.055	0.054	0.056	0.061
54.96 GHz	-0.043	-0.062	-0.061	-0.052	-0.043	-0.037
57.95 GHz	-0.011	0.030	0.030	0.026	0.020	0.015
						500 mbar
53.74 GHz	-0.021	0.019	0.017	0.008	-0.008	-0.018
54.96 GHz	0.004	-0.063	-0.054	-0.031	0.010	0.047
57.95 GHz	0.038	0.055	0.046	0.030	0.002	-0.025
						400 mbar
53.74 GHz	0.035	0.041	0.027	0.013	-0.005	-0.018
54.96 GHz	-0.069	-0.075	-0.040	-0.007	0.040	0.081
57.95 GHz	0.054	0.060	0.035	0.014	-0.018	-0.046
						300 mbar
53.74 GHz	0.063	0.038	0.019	0.004	-0.018	-0.038
54.96 GHz	-0.121	-0.058	-0.014	0.017	0.073	0.131
57.95 GHz	0.091	0.053	0.019	-0.004	-0.039	-0.072
						250 mbar
53.74 GHz	-0.030	-0.005	-0.003	-0.010	-0.043	-0.081
54.96 GHz	-0.056	-0.059	-0.065	-0.058	0.017	0.116
57.95 GHz	0.103	0.054	0.040	0.022	-0.023	-0.068
						200 mbar
53.74 GHz	0.003	-0.070	-0.054	-0.034	-0.034	-0.051
54.96 GHz	0.014	0.157	0.114	0.067	0.080	0.135
57.95 GHz	0.000	-0.118	-0.087	-0.054	-0.034	-0.027
						150 mbar
53.74 GHz	-0.042	-0.037	-0.018	-0.003	0.012	0.022
54.96 GHz	0.144	0.088	0.042	0.011	-0.015	-0.030
57.95 GHz	-0.179	-0.094	-0.048	-0.019	0.023	0.066
						100 mbar
53.74 GHz	-0.068	-0.004	0.003	0.007	0.019	0.031
54.96 GHz	0.129	-0.008	-0.019	-0.025	-0.046	-0.069
57.95 GHz	-0.135	0.006	0.015	0.018	0.041	0.074
						70 mbar
53.74 GHz	0.028	0.013	0.000	-0.007	-0.010	-0.009
54.96 GHz	-0.059	-0.026	0.001	0.014	0.022	0.026
57.95 GHz	0.080	0.034	0.010	-0.003	-0.002	0.011
						50 mbar
53.74 GHz	0.056	0.013	0.009	0.007	0.007	0.009
54.96 GHz	-0.109	-0.009	-0.001	0.002	0.005	0.007
57.95 GHz	0.117	0.010	0.001	-0.004	-0.001	0.010

Table VI.2.3: Off center frame for three frames retrieval operators for MSU at nadir over summer statistics

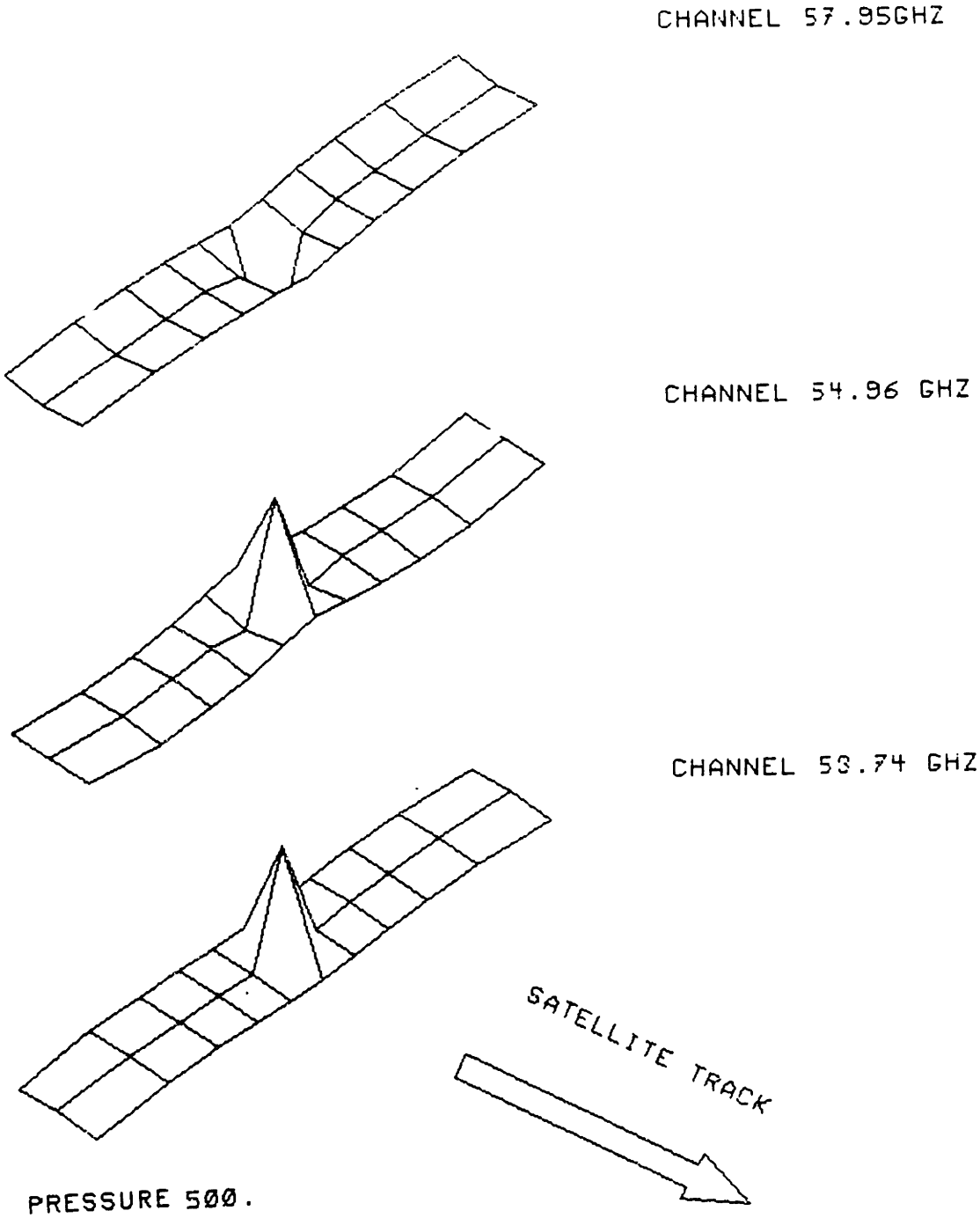


Figure VI.2.3: Three frame three-dimensional retrieval operators for MSU at nadir at 500 mbar

Hence, its coefficients (the equivalent of its impulse response) will present slow and fast variations along the frame. For the unobservable components, the lifting of the weighting functions with angle reinforces this all-pass tendency by changing the

				observable components			
53.74 GHz	0.000	0.001	0.002	0.006	0.058	1.988	2.072
54.96 GHz	0.000	-0.001	-0.003	-0.013	-0.058	0.163	0.080
57.95 GHz	0.000	0.001	0.002	0.009	0.057	-0.140	-0.035
53.74 GHz	0.000	-0.001	-0.005	-0.022	-0.129	-1.630	-1.822
54.96 GHz	0.000	0.003	0.012	0.060	0.343	1.773	2.334
57.95 GHz	0.000	-0.002	-0.005	-0.023	-0.091	0.247	0.109
53.74 GHz	0.000	0.001	0.003	0.016	0.097	0.429	0.587
54.96 GHz	0.000	-0.001	-0.007	-0.037	-0.195	-0.952	-1.264
57.95 GHz	-0.001	0.000	0.003	0.022	0.169	1.491	1.802
				first three unobservable components			
53.74 GHz	-0.209	-0.261	-0.196	-0.116	0.028	1.205	0.549
54.96 GHz	0.324	0.312	0.146	-0.033	-0.287	-0.734	-0.978
57.95 GHz	-0.345	-0.333	-0.238	-0.145	0.086	1.773	0.808
53.74 GHz	0.098	0.016	-0.020	-0.026	-0.043	-0.126	0.025
54.96 GHz	-0.282	-0.014	0.074	0.090	0.081	-0.038	-0.076
57.95 GHz	0.317	0.101	0.062	0.080	0.050	-0.960	-0.172
53.74 GHz	-0.032	0.002	-0.004	-0.007	-0.040	-0.473	-0.474
54.96 GHz	0.146	-0.022	-0.008	0.001	0.039	0.081	0.150
57.95 GHz	-0.019	0.085	0.073	0.061	-0.011	-0.690	-0.275
				last three unobservable components			
53.74 GHz	0.029	0.012	-0.002	-0.016	-0.022	0.128	0.055
54.96 GHz	-0.079	-0.034	-0.027	-0.023	0.005	0.290	0.172
57.95 GHz	0.042	-0.016	-0.056	-0.100	-0.112	0.498	-0.035
53.74 GHz	0.039	0.026	0.019	0.006	-0.008	0.188	0.179
54.96 GHz	-0.133	-0.090	-0.086	-0.064	0.015	0.538	0.284
57.95 GHz	0.069	-0.011	-0.056	-0.112	-0.135	0.634	-0.020
53.74 GHz	0.060	0.050	0.032	0.002	-0.041	0.271	0.248
54.96 GHz	-0.224	-0.178	-0.177	-0.153	0.004	1.189	0.589
57.95 GHz	0.148	0.019	-0.074	-0.188	-0.260	0.990	-0.066

Table VI.2.4: Single-spot and single-frame retrieval operator for MSU at nadir in the SVD basis representation

proportion of signals in the observable and unobservable subspaces.

VI.3 Predicted Retrieval Errors for MSU

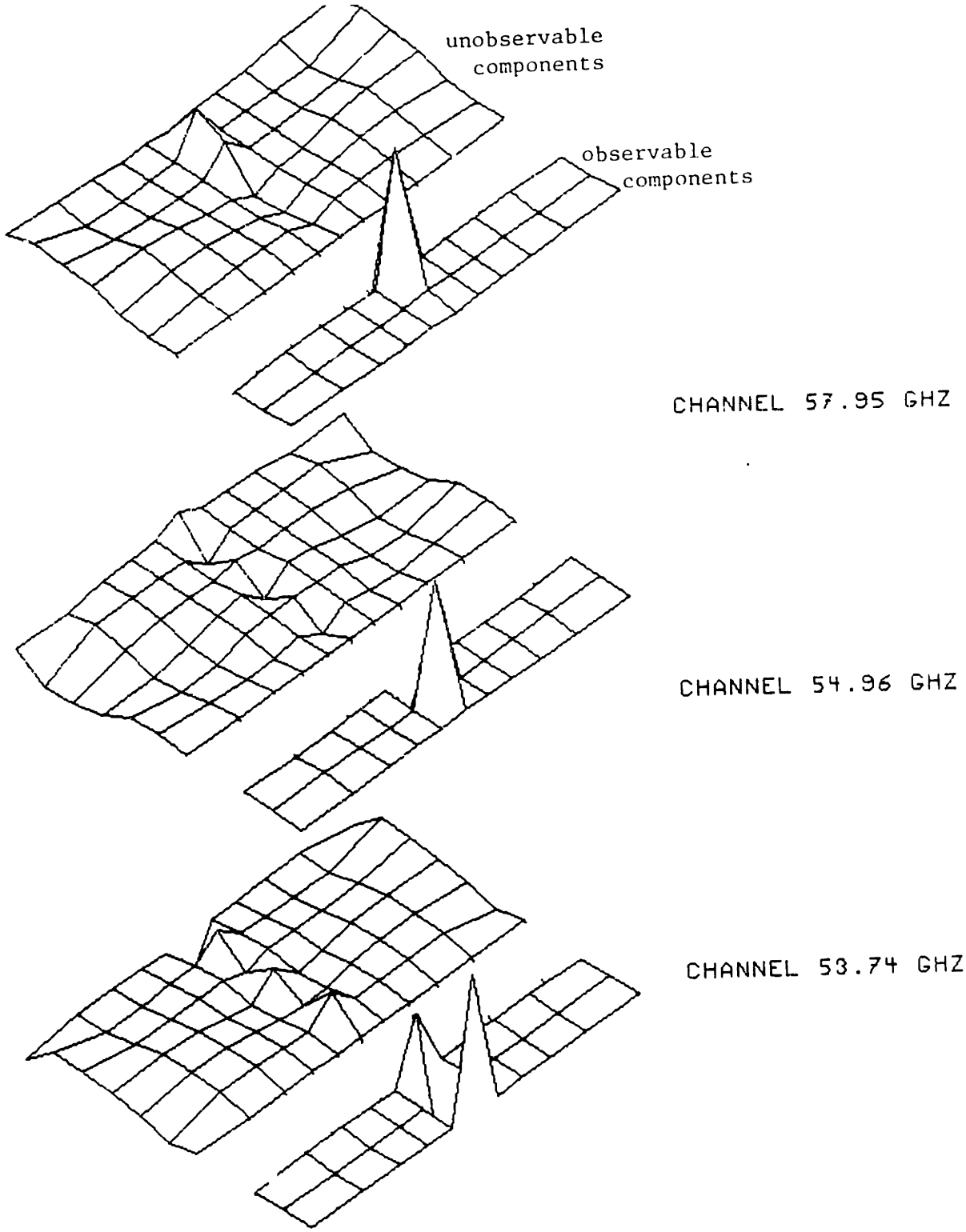


Figure VI.2.4: Single frame retrieval operator for MSU at nadir in the SVD representation

The purpose of this section is to quantify the impact of statistics (vertical and three-dimensional) on the retrieval process as well as the impact of the different parameters controlling the statistics and the measurements: Extent of correlation

in the vertical direction, in the horizontal direction, variations of the horizontal correlation with altitude as well as impact of the weighting functions and their peak locations. Using the model for the three-dimensional macrostructure in conjunction with the radiative transfer equation permits to compute and compare the retrieval errors of different retrieval experiments as well as the different parameters associated with them according to equation V.2.16. This computation can be performed directly and does not require the computation of the brightness temperatures from the temperature profiles. In a later section of this Chapter the retrieval operators considered will be implemented on measured brightness temperatures (after correction). This analysis can be performed in three different representation bases:

(1) The natural representation (user oriented) where each component of the temperature profile is the temperature at a specific pressure level. The pressure levels used in this dissertation are given in section III.4.1.

(2) The SVD (singular value decomposition) representation (instrument oriented) where the profiles are projected along the observable and unobservable subspaces of the sounding instrument as described in section V.2.3. This representation depends on the instrument and sounding angle, hence does not permit one to analyze the impact of sounding angle.

(3) The KL (Karhunen-Loeve) representation where the profiles are expanded along the eigenvectors (natural modes) of the vertical covariance kernel (Smith and Woolf, 1976). This representation is signal oriented but does permit one to compare retrieval experiments performed on different climatologies if these climatologies share the same vertical covariance kernel.

All three representations will be considered in this Chapter. Three different performance criteria can be used:

(1) The most traditional criterion and most used is the root-mean-square error (rms) for each component of the estimated profile.

(2) This rms error can be normalized to the a-priori rms deviation for the same component to produce a percentage or normalized rms error (nrms). This scaled error allows one to compare operators performing on very small datasets (few degrees of freedoms, small a-priori deviation of the profiles) with operators performing on large datasets. If R denotes the vertical covariance kernel associated with a specific dataset, let $\mathcal{R}$ be the diagonal matrix whose entries are the diagonal entries of R. That is

$$\mathcal{R} = R_{ii}\delta_{ij} \qquad (VI.3.1)$$

where δ_{ij} is the kronecker symbol. The matricial error associated with the nrms is

$$\epsilon' = (I - DW)R\mathcal{R}^{-1} \qquad (VI.3.2)$$

where D and W are defined as in Chapter V. The nrms at a specific pressure level is found by taking the square root of the corresponding element on the main diagonal of ϵ'. The product $R\mathcal{R}^{-1}$ is invariant under point-wise multiplication of the vertical covariance kernel. It's trace equals the number of pressure levels used in the computation of the radiative transfer equation (33 in this dissertation). Note that the product $R\mathcal{R}^{-1}$ is not equal (although related) to the normalized correlation matrix associated with R which is $\mathcal{R}^{-1/2}R\mathcal{R}^{-1/2}$.

(3) Another measure of efficiency for the retrieval operator is the fraction of energy in the estimated signal (normalized energy in estimate or nee) normalized to the a-priori energy (Baumann, 1980, Nathan, 1983). It is directly related to the normalized rms error according to

$$nee = 1 - nrms^2 \qquad (VI.3.3)$$

The energy in the estimate can be used as an optimization criterion instead of the traditional MMSE. Section VIII.2.2 examines this possibility.

VERTICAL COVARIANCE KERNEL: not normalized

	1000	850	700	500	400	300	250	200	150	100	70	50
1000	33.8	15.3	9.7	7.6	6.9	6.8	4.7	-0.8	-5.8	-4.2	-2.7	-0.9
850	15.3	14.1	8.7	6.0	6.2	6.6	4.3	-1.0	-5.8	-5.0	-2.5	-0.1
700	9.7	8.7	10.6	7.2	7.4	7.7	4.5	-2.0	-6.4	-5.6	-1.8	-0.1
500	7.6	6.0	7.2	7.8	7.5	7.4	4.7	-1.7	-6.1	-5.8	-2.2	-1.0
400	6.9	6.2	7.4	7.5	8.5	8.4	5.2	-1.8	-6.9	-6.1	-2.2	-1.0
300	6.8	6.6	7.7	7.4	8.4	10.1	6.9	-0.3	-7.6	-6.5	-2.1	-0.8
250	4.7	4.3	4.5	4.7	5.2	6.9	7.3	3.2	-5.0	-4.8	-1.5	-0.4
200	-0.8	-1.0	-2.0	-1.7	-1.8	-0.3	3.2	10.8	3.2	1.1	1.2	1.5
150	-5.8	-5.8	-6.4	-6.1	-6.9	-7.6	-5.0	3.2	11.5	7.8	4.2	2.5
100	-4.2	-5.0	-5.6	-5.8	-6.1	-6.5	-4.8	1.1	7.8	10.6	4.5	3.0
70	-2.7	-2.5	-1.8	-2.2	-2.2	-2.1	-1.5	1.2	4.2	4.5	5.0	3.0
50	-0.9	-0.1	-0.1	-1.0	-1.0	-0.8	-0.4	1.5	2.5	3.0	3.0	3.7

VERTICAL COVARIANCE KERNEL: normalized

	1000	850	700	500	400	300	250	200	150	100	70	50
1000	1.0	0.7	0.5	0.5	0.4	0.4	0.3	0.0	-0.3	-0.2	-0.2	-0.1
850	0.7	1.0	0.7	0.6	0.6	0.6	0.4	-0.1	-0.5	-0.4	-0.3	0.0
700	0.5	0.7	1.0	0.8	0.8	0.7	0.5	-0.2	-0.6	-0.5	-0.2	0.0
500	0.5	0.6	0.8	1.0	0.9	0.8	0.6	-0.2	-0.6	-0.6	-0.4	-0.2
400	0.4	0.6	0.8	0.9	1.0	0.9	0.7	-0.2	-0.7	-0.6	-0.3	-0.2
300	0.4	0.6	0.7	0.8	0.9	1.0	0.8	0.0	-0.7	-0.6	-0.3	-0.1
250	0.3	0.4	0.5	0.6	0.7	0.8	1.0	0.4	-0.5	-0.5	-0.2	-0.1
200	0.0	-0.1	-0.2	-0.2	-0.2	0.0	0.4	1.0	0.3	0.1	0.2	0.2
150	-0.3	-0.5	-0.6	-0.6	-0.7	-0.7	-0.5	0.3	1.0	0.7	0.6	0.4
100	-0.2	-0.4	-0.5	-0.6	-0.6	-0.6	-0.5	0.1	0.7	1.0	0.6	0.5
70	-0.2	-0.3	-0.2	-0.4	-0.3	-0.3	-0.2	0.2	0.6	0.6	1.0	0.7
50	-0.1	0.0	0.0	-0.2	-0.2	-0.1	-0.1	0.2	0.4	0.5	0.7	1.0

Table VI.3.1: Vertical covariance kernel for Peoria dataset (normalized and unnormalized)

All three measures will be used to characterize the simulations. In most of these simulations, 100 radiosonde measurement taken over a period of three years during the summer over Peoria (Ill.) are used as the statistical database for the vertical covariance kernel. The isobaric poles used for the horizontal correlation are fitted to July 1979 data over the continental United States. This database

VERTICAL COVARIANCE KERNEL: not normalized

	1000	850	700	500	400	300	250	200	150	100	70	50
1000	34.9	23.3	13.4	8.4	8.2	6.0	2.3	-1.4	-3.6	-4.7	-3.7	-2.0
850	23.3	18.9	13.3	9.4	8.6	6.1	2.0	-2.1	-4.2	-5.2	-4.0	-2.3
700	13.4	13.3	12.3	9.8	8.6	5.7	1.4	-2.8	-4.4	-5.3	-4.0	-2.1
500	8.4	9.4	9.8	11.0	10.1	6.8	2.0	-2.5	-4.8	-6.1	-4.5	-2.0
400	8.2	8.6	8.6	10.1	10.6	8.0	3.4	-1.3	-4.3	-6.4	-4.9	-2.2
300	6.0	6.1	5.7	6.8	8.0	8.9	5.6	2.2	-1.9	-5.3	-4.7	-2.8
250	2.3	2.0	1.4	2.0	3.4	6.6	9.0	7.3	2.6	-2.1	-2.7	-1.9
200	-1.4	-2.1	-2.8	-2.5	-1.3	2.2	7.3	10.4	7.6	2.8	0.9	0.2
150	-3.6	-4.2	-4.4	-4.8	-4.3	-1.9	2.6	7.6	9.4	8.0	5.5	3.4
100	-4.7	-5.2	-5.3	-6.1	-6.4	-5.3	-2.1	2.8	8.0	12.9	11.1	8.2
70	-3.7	-4.0	-4.0	-4.5	-4.9	-4.7	-2.7	0.9	5.5	11.1	13.4	12.9
50	-2.0	-2.3	-2.1	-2.0	-2.2	-2.8	-1.9	0.2	3.4	8.2	12.9	18.0

VERTICAL COVARIANCE KERNEL: normalized

	1000	850	700	500	400	300	250	200	150	100	70	50
1000	1.0	0.9	0.6	0.4	0.4	0.3	0.1	-0.1	-0.2	-0.2	-0.2	-0.1
850	0.9	1.0	0.9	0.7	0.6	0.5	0.2	-0.2	-0.3	-0.3	-0.3	-0.1
700	0.6	0.9	1.0	0.8	0.8	0.5	0.1	-0.2	-0.4	-0.4	-0.3	-0.1
500	0.4	0.7	0.8	1.0	0.9	0.7	0.2	-0.2	-0.5	-0.5	-0.4	-0.1
400	0.4	0.6	0.8	0.9	1.0	0.8	0.3	-0.1	-0.4	-0.5	-0.4	-0.2
300	0.3	0.5	0.5	0.7	0.8	1.0	0.7	0.2	-0.2	-0.5	-0.4	-0.2
250	0.1	0.2	0.1	0.2	0.3	0.7	1.0	0.8	0.3	-0.2	-0.2	-0.2
200	-0.1	-0.2	-0.2	-0.2	-0.1	0.2	0.8	1.0	0.8	0.2	0.1	0.0
150	-0.2	-0.3	-0.4	-0.5	-0.4	-0.2	0.3	0.8	1.0	0.7	0.5	0.3
100	-0.2	-0.3	-0.4	-0.5	-0.5	-0.5	-0.2	0.2	0.7	1.0	0.8	0.5
70	-0.2	-0.3	-0.3	-0.4	-0.4	-0.4	-0.2	0.1	0.5	0.8	1.0	0.8
50	-0.1	-0.1	-0.1	-0.1	-0.2	-0.2	-0.2	0.0	0.3	0.5	0.8	1.0

GOODREG is defined by the following latitudes and longitudes

25.00N	60.00N	130.00W	60.00W
60.00N	70.00N	165.00W	90.00W
35.00N	60.00N	15.00W	35.00E
25.00N	45.00N	125.00E	145.00E

Table VI.3.2: Vertical covariance kernel for NMC dataset (normalized and unnormalized)

was chosen because it yields a model characteristics of summer in the northern

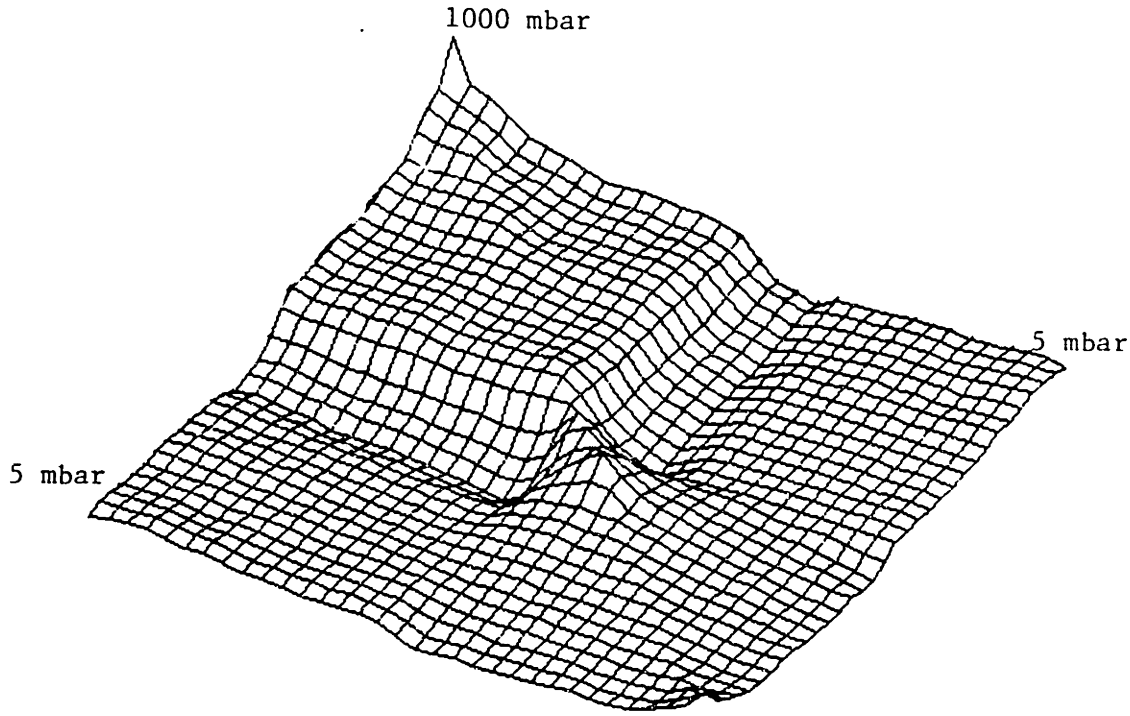


Figure VI.3.1: Vertical covariance kernel for Peoria

hemisphere. Tables VI.3.1 and VI.3.2 compare the vertical covariance kernel of the Peoria dataset with that for the NMC analysis field of the first week of July 1979 over Canada, the US, western Europe and Japan. This comparison is only performed at the 12 pressure levels provided by NMC. Figures VI.3.1 and VI.3.2 present three-dimensional rendering of the vertical covariance kernel for the Peoria set at all 33 pressure levels.

Although the power in the different component is different for both covariances in the stratosphere, the normalized correlations are close. Discrepancies between elements, such as the correlation 250-400 *mbar*, are the results of non-stationarities. Another reason for choosing the Peoria second moment statistics as the basis for the simulations is that retrieval operators derived from the resulting

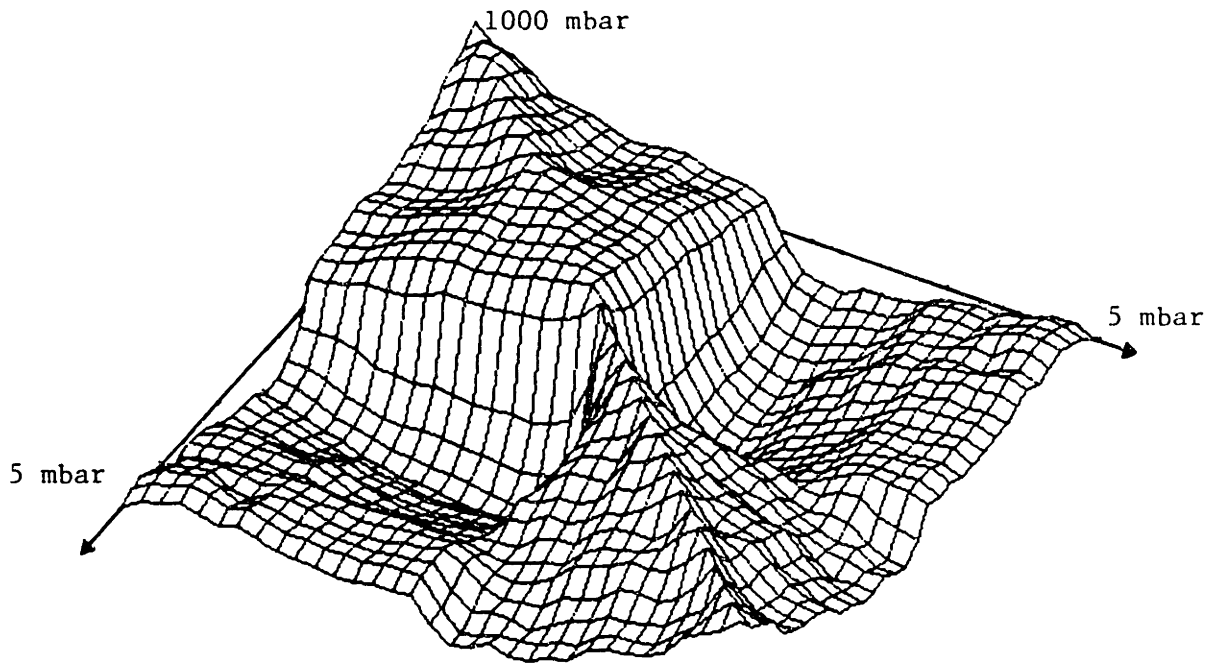


Figure VI.3.2: Vertical covariance kernel for Peoria (normalized)

model will in general perform best in implementation on the MSU instrument. The tropopause inversion appears as a strong anticorrelation in the tables and by a cliff like structure for the three-dimensional plots.

All simulation calculations are performed above land. The conclusions about multidimensional retrievals extend to retrieval above sea, the only difference is in the weighting functions, retrieval operators and retrieval errors at the surface.

Impact of Vertical Statistics

Before examining the impact of multidimensionality in the retrieval process, let us briefly consider the impact of the vertical covariance kernel and of the weighting functions.

As previously mentioned, as the sounding angle departs from nadir, the

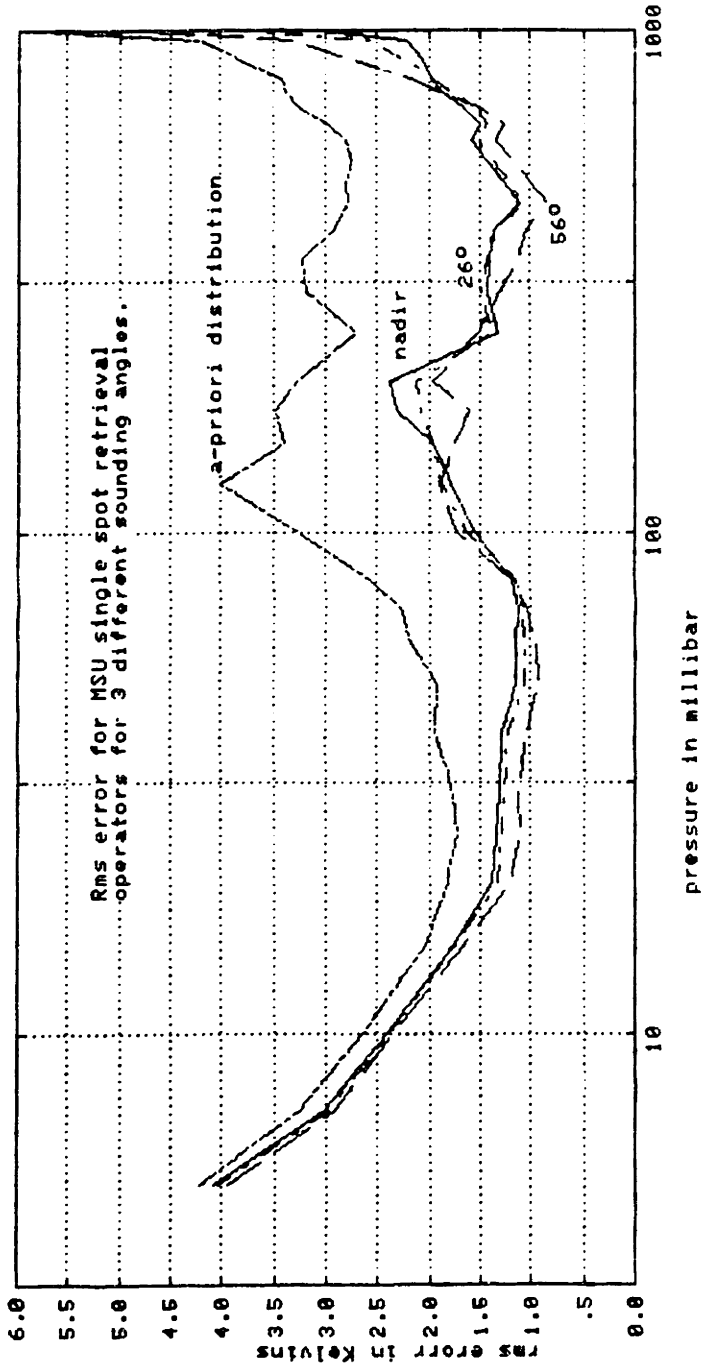


Figure VI.3.3: Predicted rms error for MSU over typical summer climatology for three different sounding angles: nadir, 26°, and 56°

weighting functions rise and the ground contribution diminishes. Figure VI.3.3 presents the effects of this lifting of the weighting functions on the rms retrieval error using MSU at nadir, 26°, and 56°. The atmosphere is typical of summers' atmospheres. The effect of the lifting of the weighting function is not clear around the tropopause inversion because the negative correlation across the inversion overwhelms the retrieval operator. However, in the troposphere the retrievals performed at nadir are slightly better than those performed at 56° and vice versa for the stratosphere. In both regions, one of the weighting functions peaks and the estimation of temperature is directly related to the measurements, hence the dependence of the retrieval error on the sounding angle.

When the signal-to-noise ratio is good, the only source of improvement in single spot estimation over physical retrievals (i.e., minimum information solutions) takes place in the unobservable subspace of the instrument where the inter-pressure correlations translates into correlation between observable and unobservable components. Figure VI.3.4 illustrates the impact of the vertical covariance kernel on the retrieval error. Three different climatologies are considered: a climatology characteristic of summers in the northern hemisphere (Peoria over three summers), a climatology characteristic of a specific summer (July 1979 over the US), and a climatology with no correlation between temperatures at different pressure levels (diagonal covariance kernel). The major difference between the two summer climatologies is a stronger correlation in the vertical direction and in the lower stratosphere and in troposphere for July 1979.

In order to compare fairly the rms errors of these three cases, all three climatologies have been scaled to the same power distribution as the Peoria dataset. Let R_a and R_b denote the covariance kernels associated with the two climatologies and $\mathcal{R}_a$ and $\mathcal{R}_b$ the two diagonal matrices associated with these matrices according to equation (VI.3.1). In order to scale the matrix R_b to the matrix R_a while

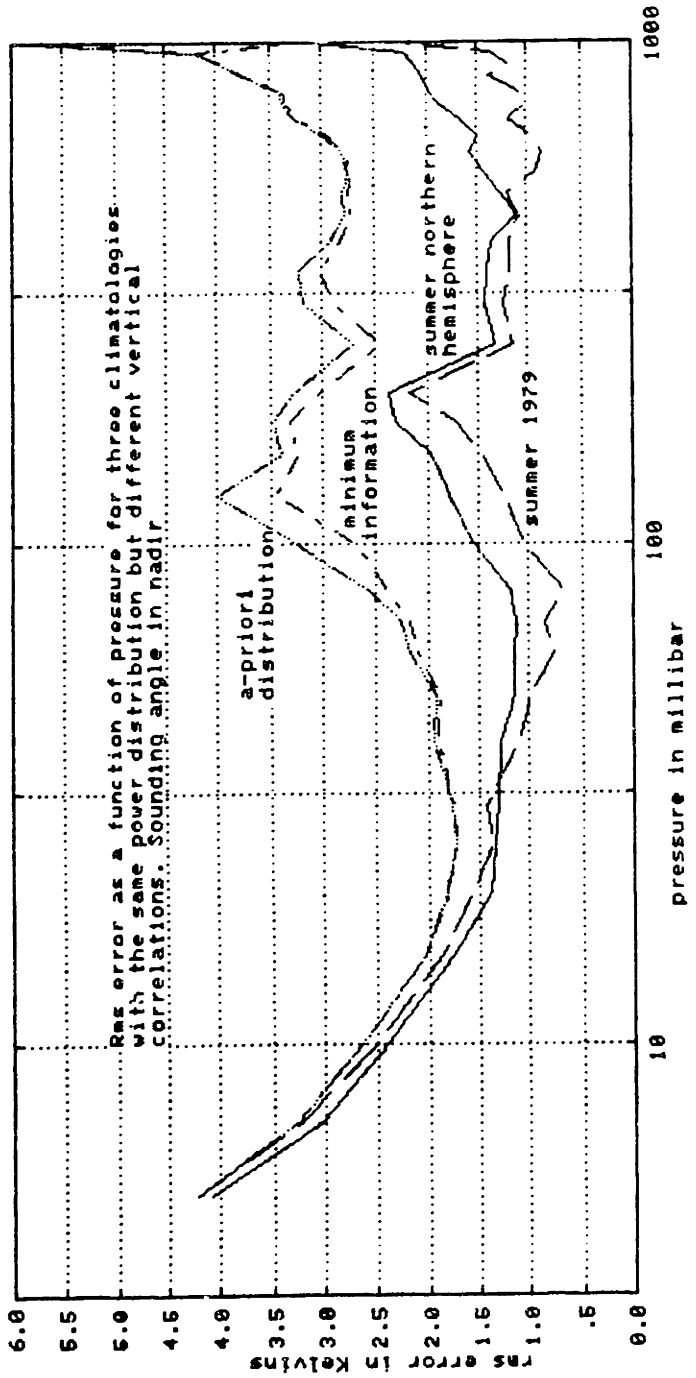


Figure VI.3.4: Predicted rms error for MSU operating at Nadir for three different climatologies: one specific to summer 1979, one characteristic of summer in the northern hemisphere, and one with no correlation between pressure levels

preserving its correlation structure, one performs the similarity transformation

$$\tilde{R}_b = \mathcal{R}_a^{1/2} \mathcal{R}_b^{-1/2} R_b \mathcal{R}_b^{-1/2} \mathcal{R}_a^{1/2} \qquad (VI.3.4)$$

The transformed matrix $\tilde{R}_b$ satisfies

$$trace\{\tilde{R}_b\} = trace\{R_a\} \qquad (VI.3.5)$$

and (with obvious notation)

$$\tilde{\mathcal{R}}_b^{-1/2} \tilde{R}_b \tilde{\mathcal{R}}_b^{-1/2} = \mathcal{R}_b^{-1/2} R_b \mathcal{R}_b^{-1/2} \qquad (VI.3.6)$$

The proposed similarity transformation changes the a-priori rms distribution at each pressure level but does not alter the covariance matrix structure as defined in section V.3.4.

The climatology with no inter-pressure correlation will, once scaled, correspond to a true minimum information solution only if the a-priori power is the same at each pressure level for the Peoria dataset. This is not the case as illustrated by Table VI.3.1. Because the elements on the diagonal matrix are not identical, there will be a correlation between observable and unobservable components, albeit small. However this climatology will still be called minimum information. Of the three cases, the "minimum information" performs the worse (almost as bad as the a-priori distribution) whereas the strong correlation July 1979 case is the best. Note that as the pressure decreases, the statistics becomes unreliable and the rms errors of the three operators tend to the a-priori distribution. The minimum information performance is bad because of the fine discretization of the atmosphere which makes the slabs heights small.

Figure VI.3.5 considers the same climatologies but considers a sounding angle of 56 degrees. The impact of inter-pressure correlation remains the same, especially at altitudes away from the functions's peaks.

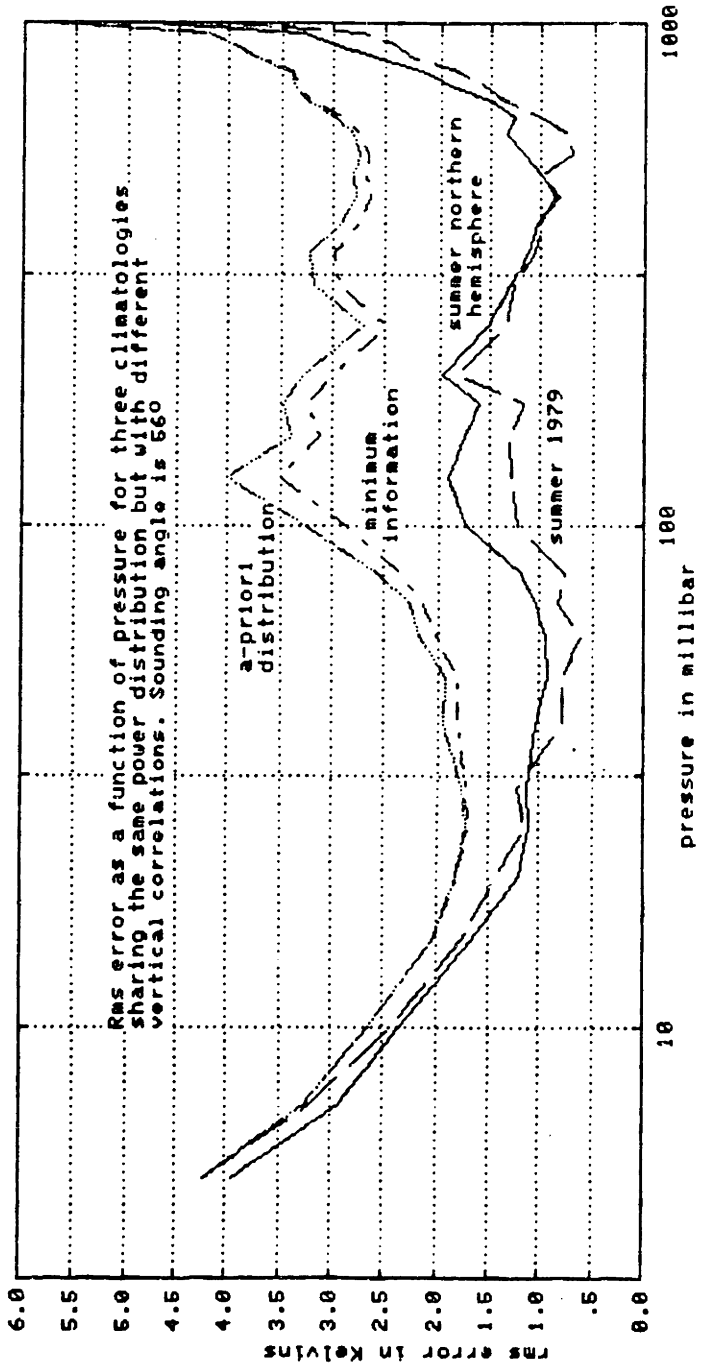


Figure VI.3.5: Predicted rms error for MSU operating at 56° for three different climatologies: one specific to summer 1979, one characteristic of summer in the northern hemisphere, and one with no correlation between pressure levels

Figure VI.3.6 examines the same problem as above in the SVD basis. Because the three climatologies have different vertical covariance kernels, the a-priori power in the observable and unobservable components are different although they are the same in the natural representation. A normalized measure must be considered, in this instance, the nee. The basis vectors for this SVD analysis are described in section V.2.

Most of the energy in the observable components is captured by the retrieval process and is fairly independent of the vertical structure except for the minimum information solution. The reason for the decreased nee in the MIE case is the absence of correlation between observable components. This correlation would reduce the rms error. This decrease in nee also shows the importance of considering all eigenvalues of the observable component covariance matrix and of the sensor noise covariance matrix before considering a retrieval situation as being good or bad from a SNR point of view. Although the power of the temperature profiles is much larger than the noise power for MSU, the impact of this noise is important because the process being examined and the observations are vectors. Because the SNR is not as good as it appear, the coupling matrices characteristic of multidimensional operators will not, for the observable components, be as small as expected, and there will be some improvement in the rms error for the observable components.

On the contrary, the nee for the unobservable components varies greatly with the type of covariance kernel. As a general rule, the more specific the covariance, the lesser the nrms associated with the estimator, and the greater the amount of energy captured by the estimator. For the minimum information, the nee is almost 100 percent for the low index unobservable components which are the smoothly varying vectors. As noted, should the covariance matrix correspond to homoscedascitic profiles (all a-priori power distribution equal) the correlation between observable and unobservable components will be identically zero, and the nee would equal zero

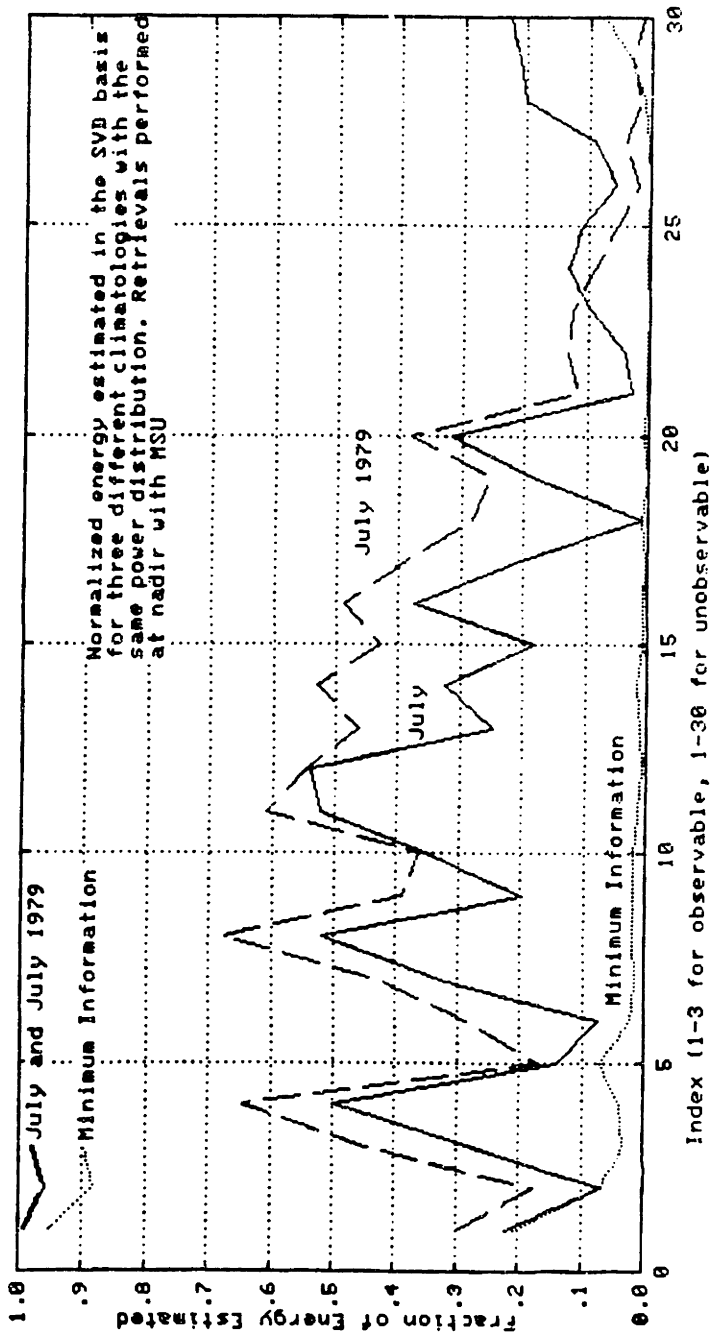


Figure VI.3.6: percentage of energy estimated for three different climatologies using MSU at nadir expressed in the SVD basis (observable/unobservable)

for all unobservable components.

For low indexes, the nee appears to pulse with a period of four indexes. This pulsing is caused by the sharp inversion at the tropopause which aligns or misalign with oscillations of the SVD basis vectors.

As the basis index increases, the proportion of energy captured tends to decrease. This decrease results from the fact that higher order unobservable components are associated with rapidly varying basis vectors which are difficult to estimate. At high indexes, the July 1979 climatology performs worse (from a normalized rms error point of view) than the summer climatology because of the stronger correlation between adjacent pressure levels close to the ground and above the tropopause. Since the high index unobservable basis vectors are rapidly varying quasi-sinusoidally, they will be more difficult to estimate for those covariance kernels which correspond to smooth, ie. correlated, profiles.

Impact of Horizontal Statistics

The impact of multidimensionality on retrievals can be quantified and analyzed different ways. The most obvious one is simply to compare the rms retrieval error associated with optimal one-, two-, and three-dimensional retrieval operators operating on the same climatology.

Figure VI.3.7 compares the rms error as a function of pressure for a single spot estimator, a single-frame estimator (two dimensional retrieval operator perpendicular to the satellite's direction), three-, and five-frame estimators. The estimation is performed with MSU at nadir. The case considered in the plot is typical of summer in the northern hemisphere. Note that most of the improvement in retrieval error is provided by the first frame estimator, and that there is little noticeable improvement between the three- and the five-frame operators. Figure VI.3.8 considers the same situation expect that the sounding angle is 56° and the sounding location is at the edge of the sounding frame. However, the same conclu-

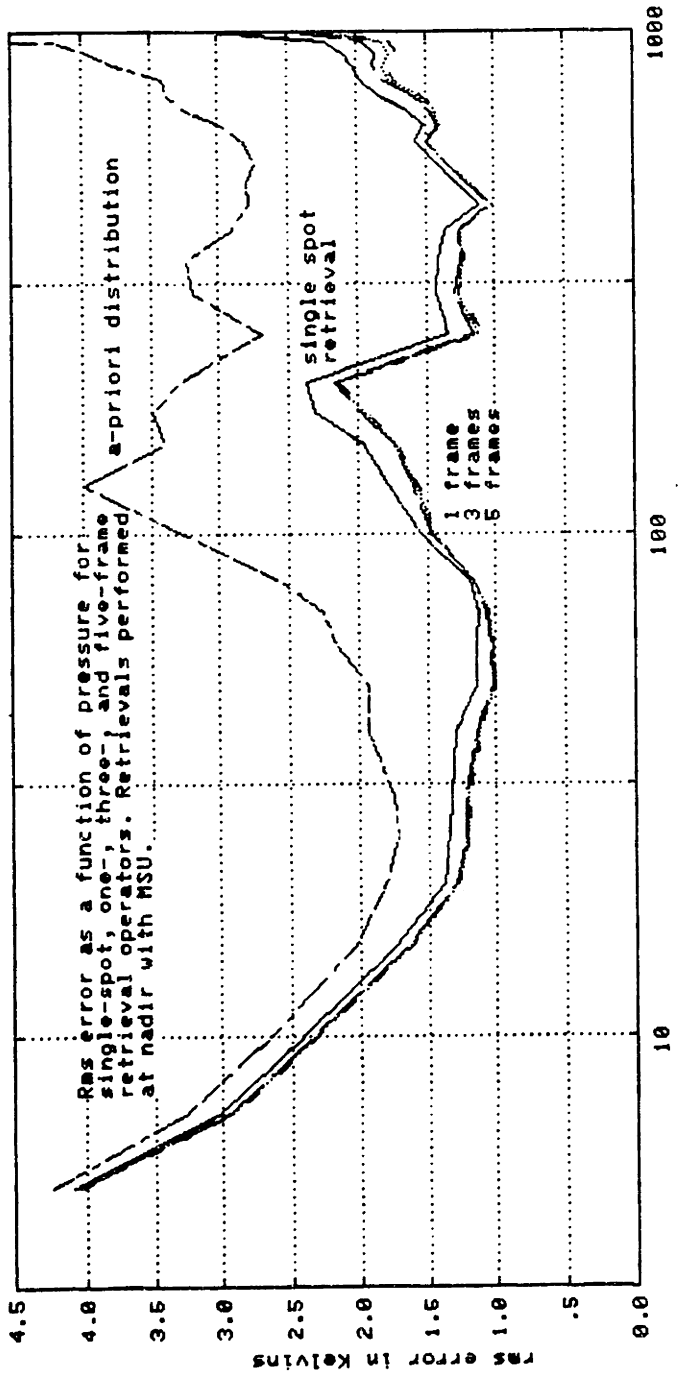


Figure VI.3.7: Predicted rms error at a function of pressure for single-spot, one-, three-, and five-frame retrieval operators for MSU at nadir over typical summer case

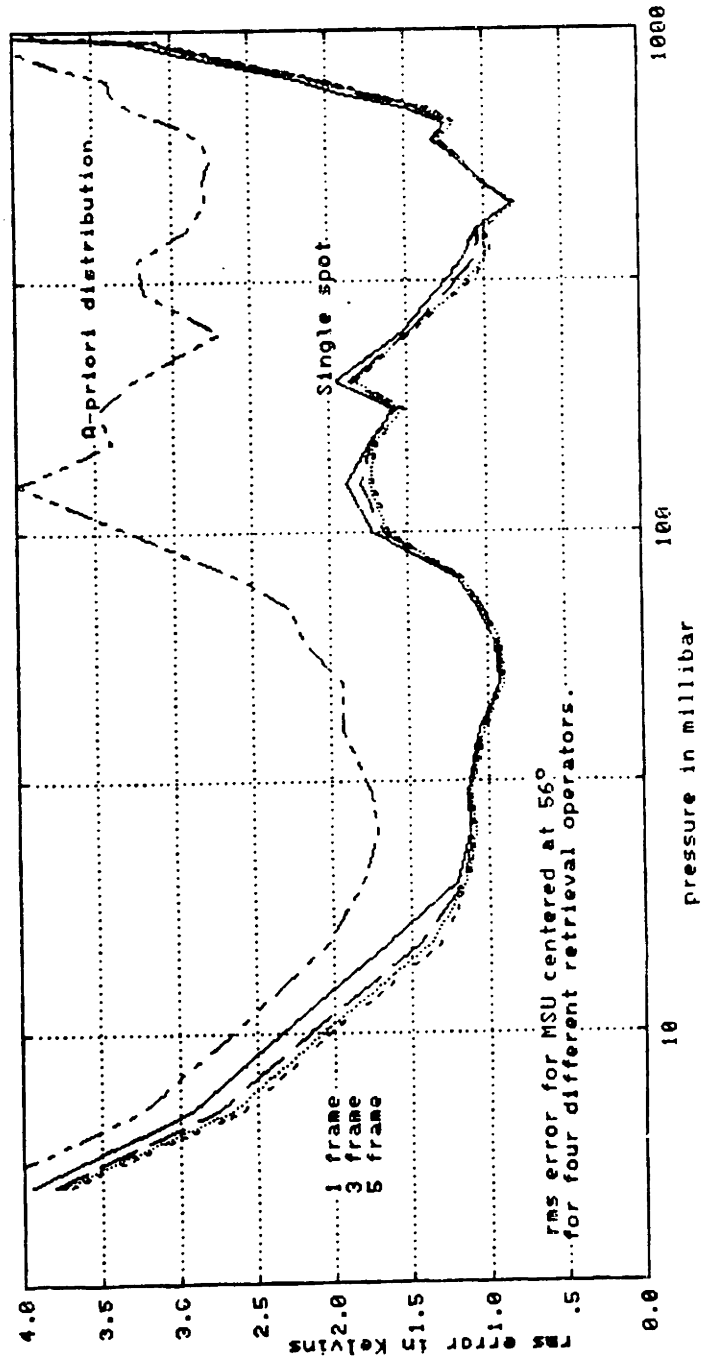


Figure VI.3.8: Predicted rms error at a function of pressure for single-spot, one-, three-, and five-frame retrieval operators for MSU at 56° over typical summer case

sions prevail, that is that most of the improvement is provided by the single-frame estimator. Note that the improvement is larger in the stratosphere for operators centered around the 56° degree measurements than around the one centered at nadir.

Figure VI.3.9 presents the improvement provided by multi-dimensional operators over single spot operator for the same case as above. The improvement I_{mD} is defined (for every component of the profile) as

$$I_{mD} = -\frac{rms_{multiD} - rms_{1D}}{rms_{1D}} \qquad (VI.3.7)$$

The improvement is defined as a fraction in order to get a performance measure independent of the a-priori power distribution, itself a function of the number of degrees of freedom used in the compilation of the vertical covariance kernel. The improvement is largest close to the ground (up to 35 %) and in the upper troposphere. Typical improvement is on the order of 10 to 15 %. Around 200-250 *mbar* a substantial improvement is provided by the inclusion of different frames of data in the retrieval process, this improvement arises from the differences in the horizontal correlation length around and across the tropopause. The lack of improvement for high altitudes ($p \leq 10mbar$) occurs because of the short horizontal correlation length resulting from unreliable statistics at those pressure levels. At 10 *mbar*, the weighting functions associated with 57.95 *Ghz* (the highest probing channel) equals $1/15^{th}$ of its peak value for the nadir case and $1/10^{th}$ of its peak value for 56°. These small values imply that most of the reduction of rms error over the a-priori rms distribution comes from the statistics. As noted in Chapter IV, at high altitudes, measurements are sparse, the isobaric poles' real parts overestimated, thus the correlation length is underestimated. Setting the isobaric poles for $p \leq 10mbar$ to their values in the lower stratosphere is one method to extend the range of validity of the three-dimensional model for the covariance and to further improve the rms error at high altitude. However, because the NMC analysis field does not ex-

tend up to those altitudes, there is no method to test the validity of this extension, which thus remains just a conjecture. Note, that even if the horizontal correlation is strengthened at high altitudes, because these regions lie more in the unobservable subspace than in the observable subspace and because the vertical variations of the horizontal correlation are small, the impact of multi-spot operators over single-spot operators will be small.

As discussed in Chapter V, the improvement going from one- to multi-dimensional retrieval operators is, in cases of good SNR's, mostly occurring the unobservable subspace. This is true because the single-spot estimates of the observable components are already good. Figure VI.3.10 exhibits this fact by plotting the nrms for the observable and unobservable components associated with MSU at nadir. The percentage of rms error is small for the observable components and it barely changes as more brightness temperature measurements are added to the retrieval process. That percentage is much larger for the unobservable components and it decay substantially for multidimensional operators. However, most of the energy in the profiles lies in the observable components. Hence, although multi-dimensional operators reduce a large fraction of the retrieval error for unobservable components, they only reduce a small fraction of the total energy as shown in figure VI.3.10.

In section V.3, the parameters that control the improvement of multidimensional operators over single-spot operators were found to be: The amount of isobaric (horizontal) correlation, its variation with respect to height, and the differences in weighting functions between the points considered. Let us now quantify these different factors.

To analyze the impact of horizontal correlation, consider one- and three-frame retrieval operators during the summer (Peoria dataset) and winter (February 1981 dataset). Figure VI.3.11 presents the rms error as a function of pressure for

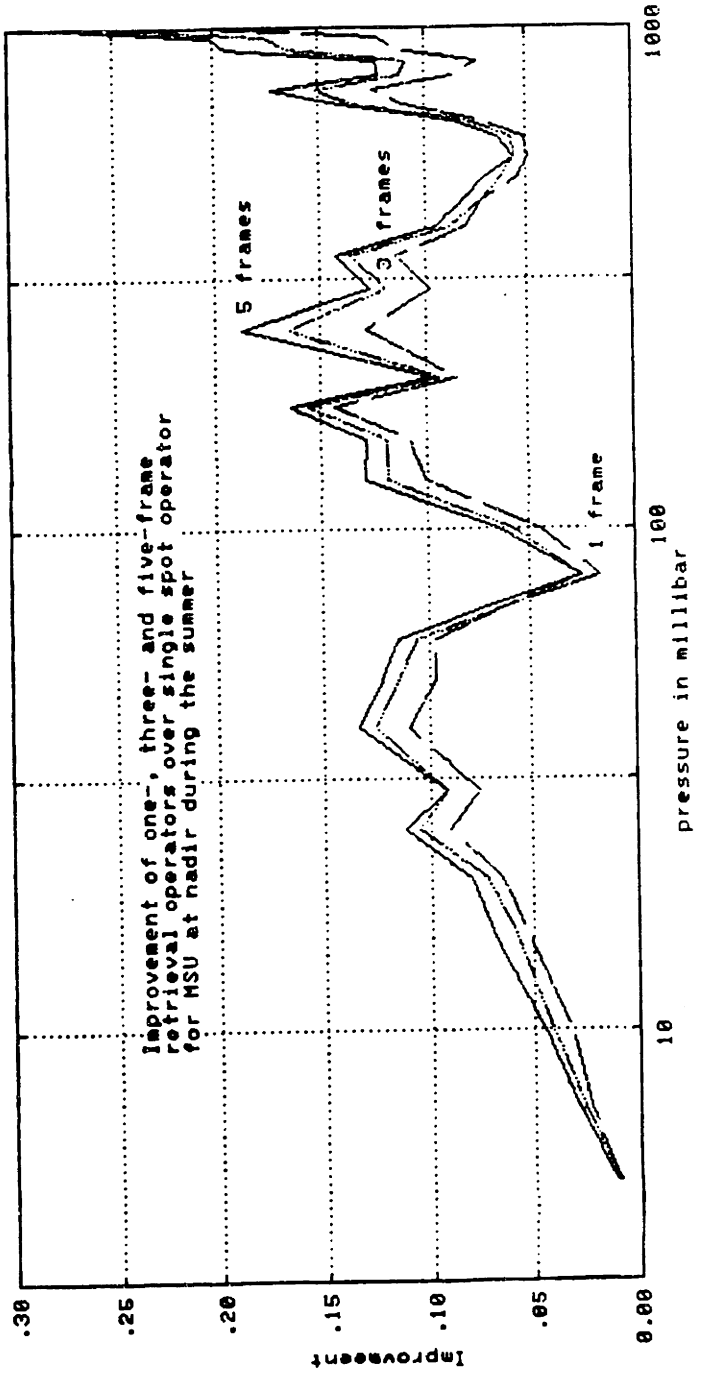


Figure VI.3.9: Improvement in rms error at a function of pressure for single-spot, one-, three-, and five-frame retrieval operators for MSU at nadir over typical summer case

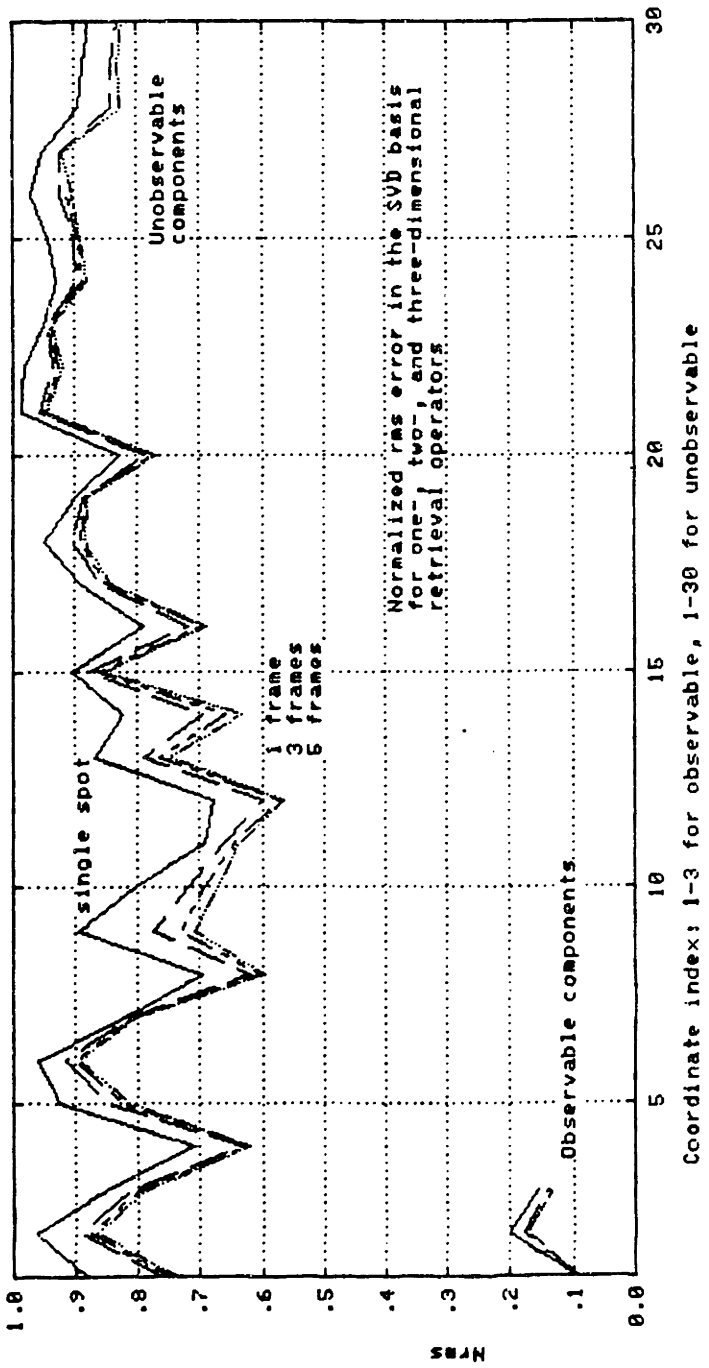


Figure VI.3.10: Nrms error in the SVD representation for single-spot, one-, three-, and five-frame retrieval operators for MSU at nadir over typical summer case

the winter case. As for the summer, most of the improvement in the winter case is produced by the inclusion of the first frame of observations into the retrieval process. However, the amount of rms error removed by the inclusion of the first frame of data is less for winter that it is for summer. This is the result of shorter correlation length. At high altitudes ($p \leq 50 mbar$), the improvement in rms error is solely provided through statistics (the weighting functions are almost zero). At those pressures, the rms error for summer retrievals (Figure VI.3.7) is closer to the a-priori distribution than the rms error for winter retrievals are to the a-priori distribution for winter (Figure VI.3.12). This implies that the vertical correlation extend higher in winter than in summer. This fact agrees with the general behavior of stratospheric atmosphere which are more stable in winter than in summer. At low altitudes, the expected rms error is greatly reduced with respect to the a-priori distribution. This reduction is caused by the strong correlation close to the ground. This correlation is larger than expected and is likely the result of the extrapolation of temperature profiles according to a subadiabatic lapse rate (half of the lapse rate to be exact) below the lowest sensed level. Although this extrapolation (necessary because of limited memory allocation of the computer used to reduce the data) is only performed for profiles where the 850 *mbar* is present, its impact of the estimated temperature profile may be important.

Figure VI.3.13 compares the improvement (as defined by equation VI.3.7) of the one-frame operator in the summer and winter cases over the United States and Europe. The improvement is typically larger in summer than in winter and both curves present local minima at 80 *mbar* and in the lower troposphere. The improvement for winter at 900 *mbar* (35 %) is the result of strong horizontal correlation. As seen in Chapter IV, the evaluation of the isobaric poles which control the horizontal correlation length, is not altered by the vertical extrapolation. This indicates that improvement is most likely to take place at the pressure levels just above the ground when implementing retrieval operators on real data (they will). 35

percent improvement however is unlikely in most implementation because of errors in the reflectivity and changes in the weighting functions caused by the orography. In general, the improvement in summer is larger than in winter. This results from a stronger horizontal correlation and from smaller variations of this correlation with height in the winter case, because of the stronger vertical correlation in winter's stratospheres.

To study the impact of including measurements performed with different sounding angle, consider a two-dimensional retrieval operator 1.1 Mm long. This retrieval operator can be in the direction perpendicular to the satellite, where it occupies the entire frame, or parallel to the track using seven adjacent points. In the first case, the operator makes the best usage of the changes of weighting matrices with angle whereas in the second case, all observations are performed under the same conditions. Figures VI.3.14 to VI.3.15 compare the rms error for the two orientations in the natural and SVD basis representations. The retrieval errors are comparable, with a slight improvement for the perpendicular direction where the lifting of the weighting functions reduces the retrieval error further. The difference between the two retrieval operator is especially noticeable in the 200 to 100 *mbar* range, where through lifting of 54.96 GHz weighting function from 270 to 180 *mbar* (See Figure II.4.2) , the features in that pressure range are observed under very different conditions when the retrieval operator is perpendicular to the satellite's track and the retrieval error is further reduced.

Note the difference between the a-priori rms distribution and rms error in the observable and unobservable subspace. The rms error for the observable varies between $1/9^{th}$ to $1/6^{th}$ of the a-priori rms distribution which corresponds at worse to a mean square error $1/36^{th}$ of the a-priori power. Note also that the three observable components are associated with a large proportion of the a-priori signal power.

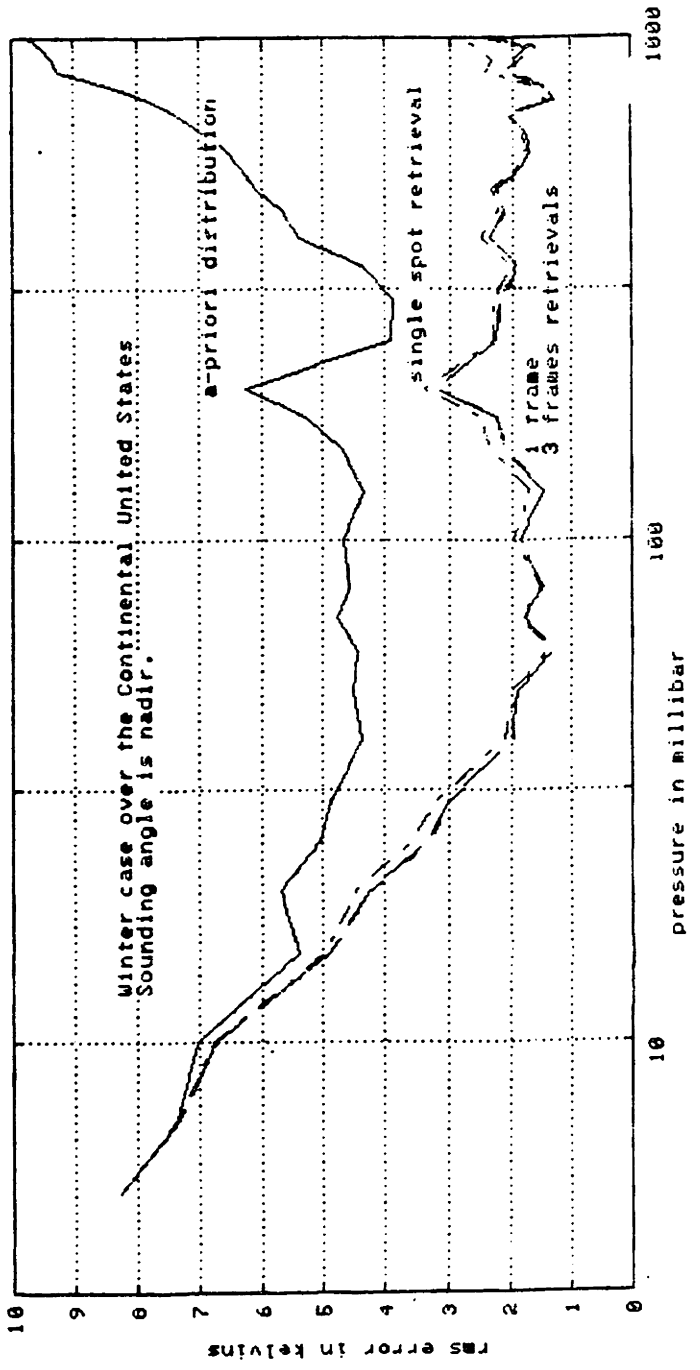


Figure VI.3.11: Rms error at a function of pressure for single-spot, one-, and three-frame retrieval operators for MSU at nadir over winter case

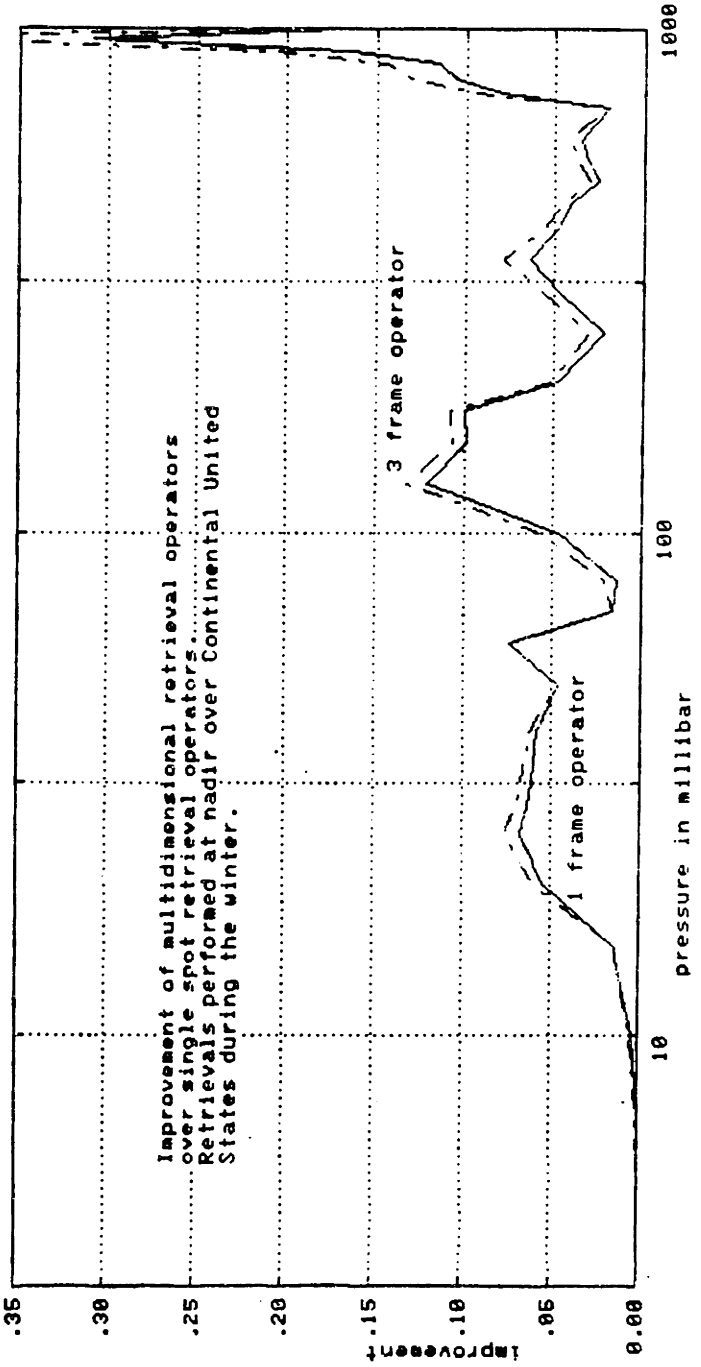


Figure VI.3.12: Improvement in rms error at a function of pressure for single-spot, one-, and three-frame retrieval operators for MSU at nadir over winter case

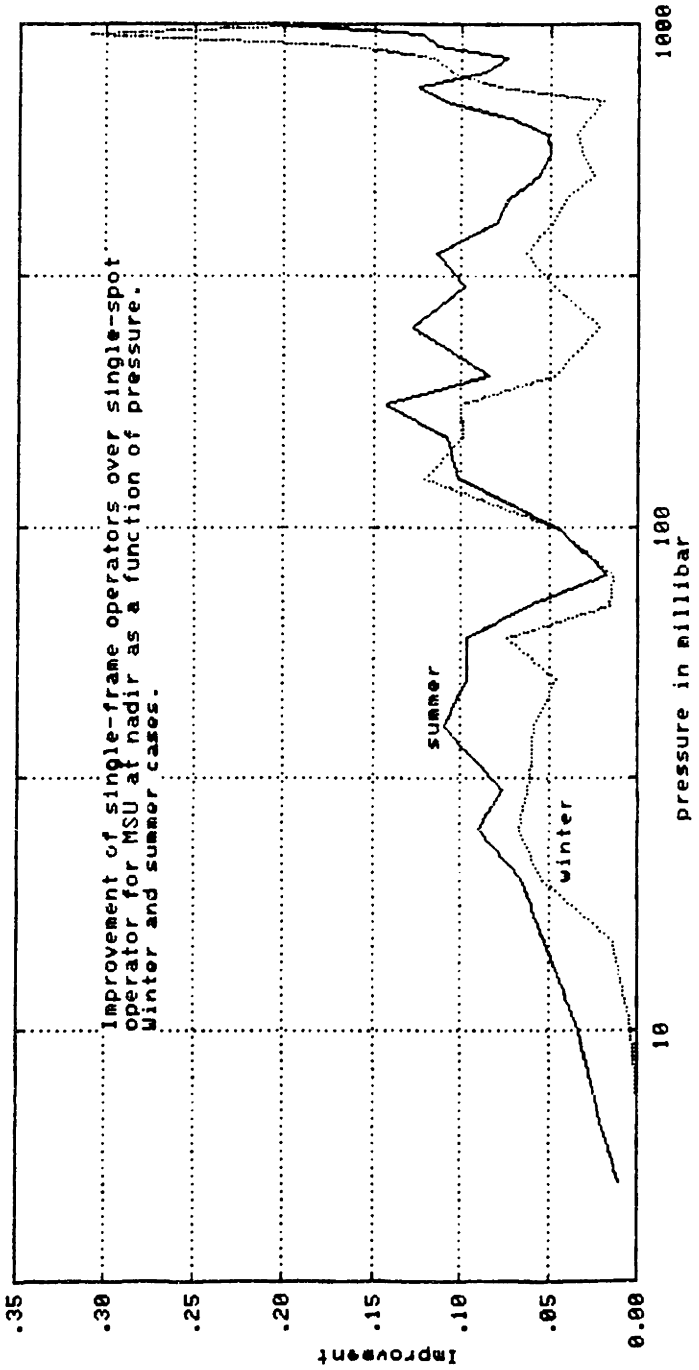


Figure VI.3.13: Improvement in rms error for one-frame retrieval operators over single-spot operators as a function of pressure for MSU at nadir over summer and winter cases

Analyzing the retrieval process in the KL basis permits to reconsider the number of modes required to adequately represent the pressure covariance kernel. Figure VI.3.16 presents the nrms for a single-spot and single-frame estimation. A nrms of 100 percent for a temperature component indicates that component is not estimated at all, that is that the a-priori mean is used as estimate. On the other hand, the lower the nrms for a component, the more efficient the estimation. The figure shows that only the last five modes of the vertical covariances (the modes with the most energy) are being estimated by the one-dimensional LMMSE. The other modes are left unestimated and the corresponding rms error for those eigenvectors equal the a-priori rms distribution. This results indicate that only five modes are required to represent the vertical covariance kernel for retrieval using MSU according to the criterion introduced by Smith (Smith and Woolf, 1976). If one uses a 95 percent nrms as the decision criterion for the number of modes required and look at the retrieval error associated with the single-frame retrieval operator, one must consider eight modes instead of five for the covariance. This number is only adequate for MSU. However, it shows that one must consider multi-dimensional retrieval operators, that is the entire three-dimensional covariance kernel, before expanding the vertical kernel.

To conclude the study of multidimensional operator using expected rms errors, the impact of horizontal correlation is considered. Using the winter data, the covariance kernels corresponding to the United States and Europe are merged together. Because these matrices are similar to one another, this combination does not alter the vertical correlation as defined in section V.3.4. The isobaric correlation structure however is not combined. Figures VI.3.17 and VI.3.18 present the rms error at nadir for single frame optimal retrieval operators for the United States and Europe as well as the corresponding improvements over single spot operators. The two retrieval operators perform similarly. Except close to the ground, Europe's improvement is slightly better than America's improvement. This difference may

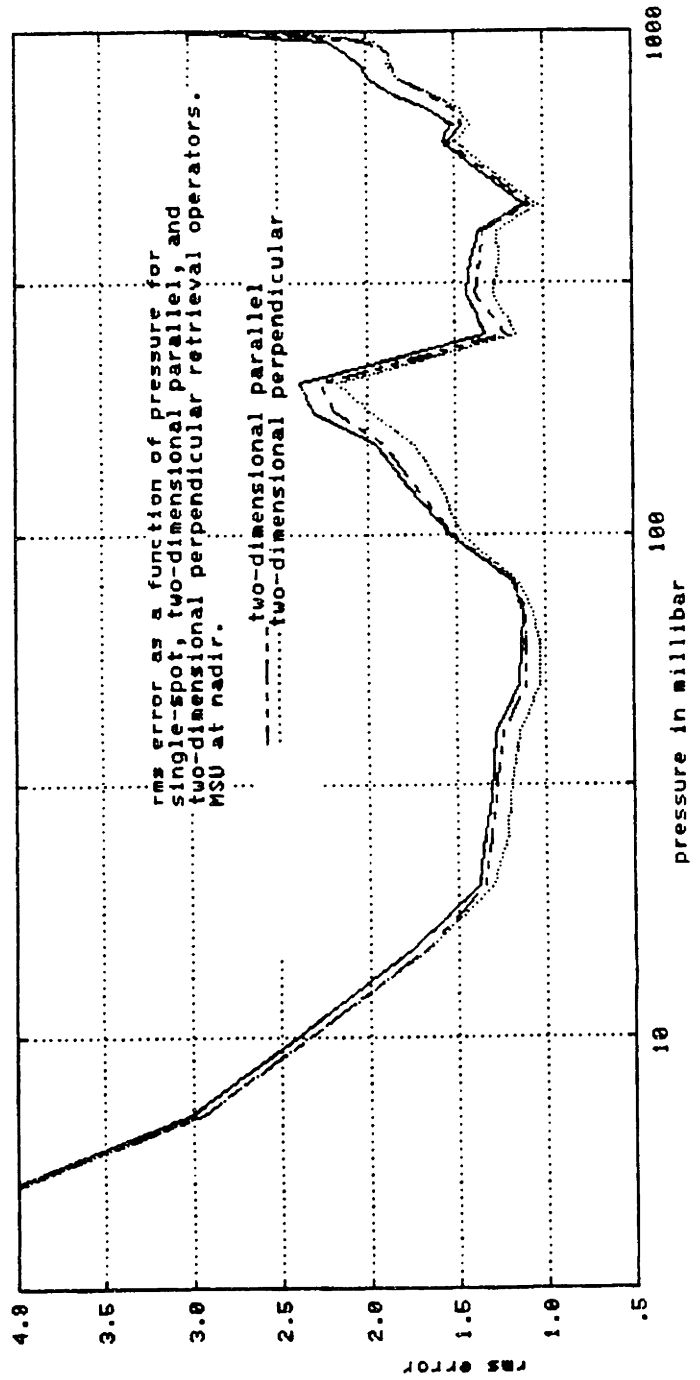


Figure VI.3.14: rms error in the natural representation of temperature profiles for single-spot, two-dimensional perpendicular, and two-dimensional parallel retrieval operators for MSU at nadir in summer

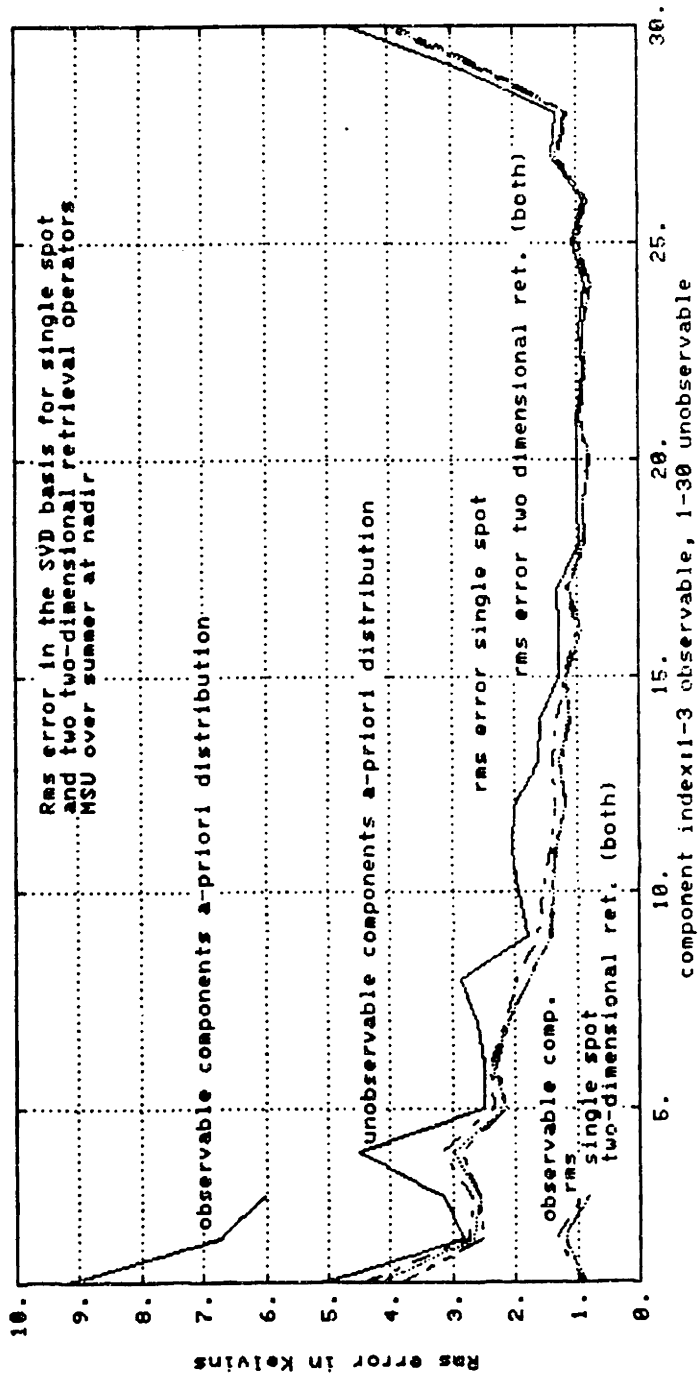


Figure VI.3.15: rms error in the SVD representation of temperature profiles for single-spot, two-dimensional perpendicular, and two-dimensional parallel retrieval operators for MSU at nadir in summer

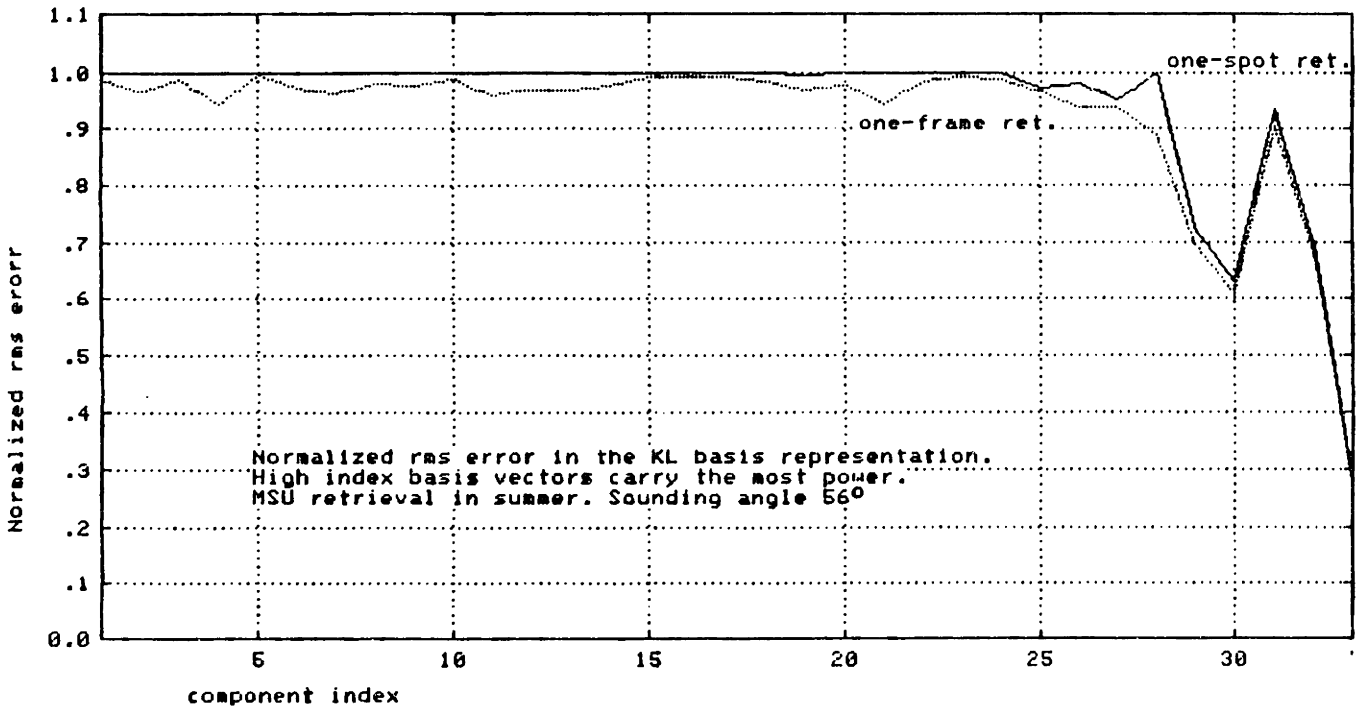


Figure VI.3.16: Normalized rms error in the KL basis for single-spot and two-dimensional retrieval operators.

be attributed to the regularizing impact of the Gulf Stream on western Europe, but more likely to be a reflection of the difference in size between the two regions. The smaller the region, the more valid the model, the longer the correlation length and the larger the improvement. Note also that both improvement have a local minimum around 80 *mbar*.

VI.4 Implementation of Retrieval Operators on MSU

In section, the multidimensional retrieval operators derived in the previous sections are implemented on MSU data for different cases studies characteristic of winter and summer cases. As previously discussed, the design of retrieval operators has been conducted in the spatial domain rather than in the frequency domain for two principal reasons: (1) the finite region of support of the filter in the direction

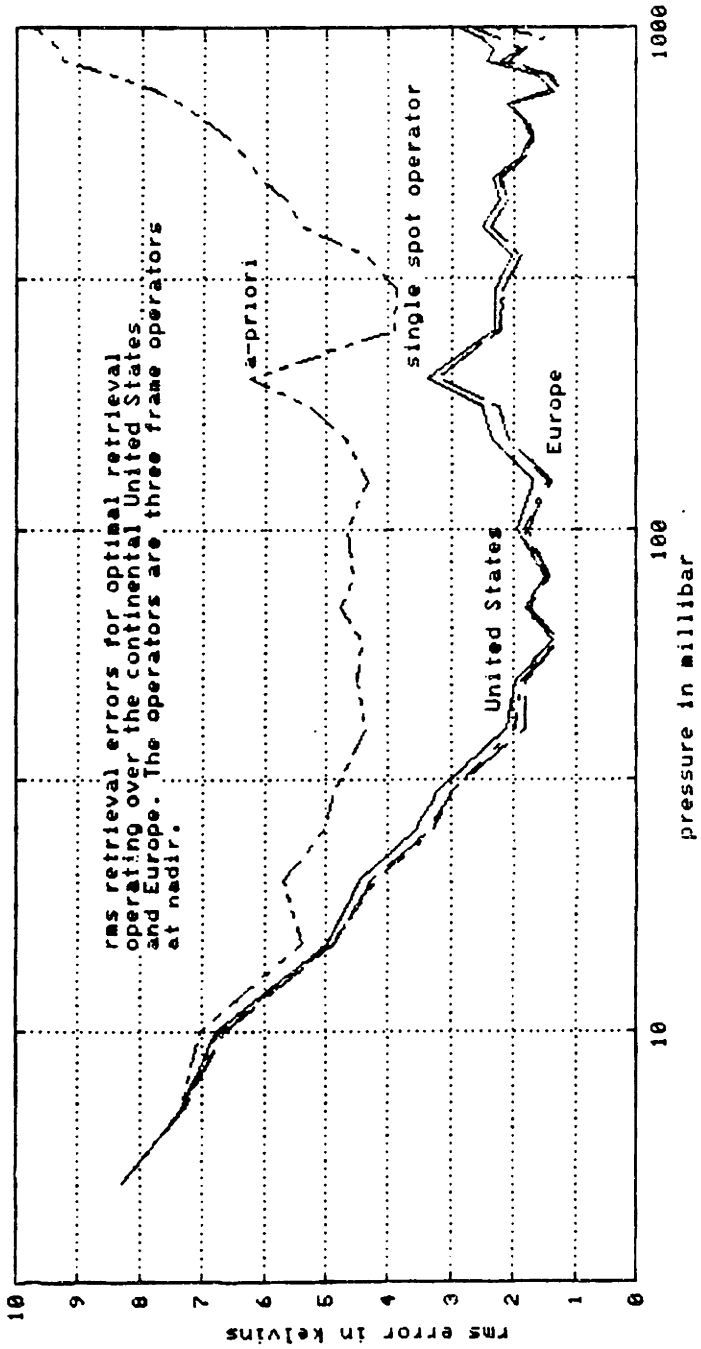


Figure VI.3.17: rms error as a function of pressure for MSU at nadir in winter over Europe and the United States (single frame retrieval operators)

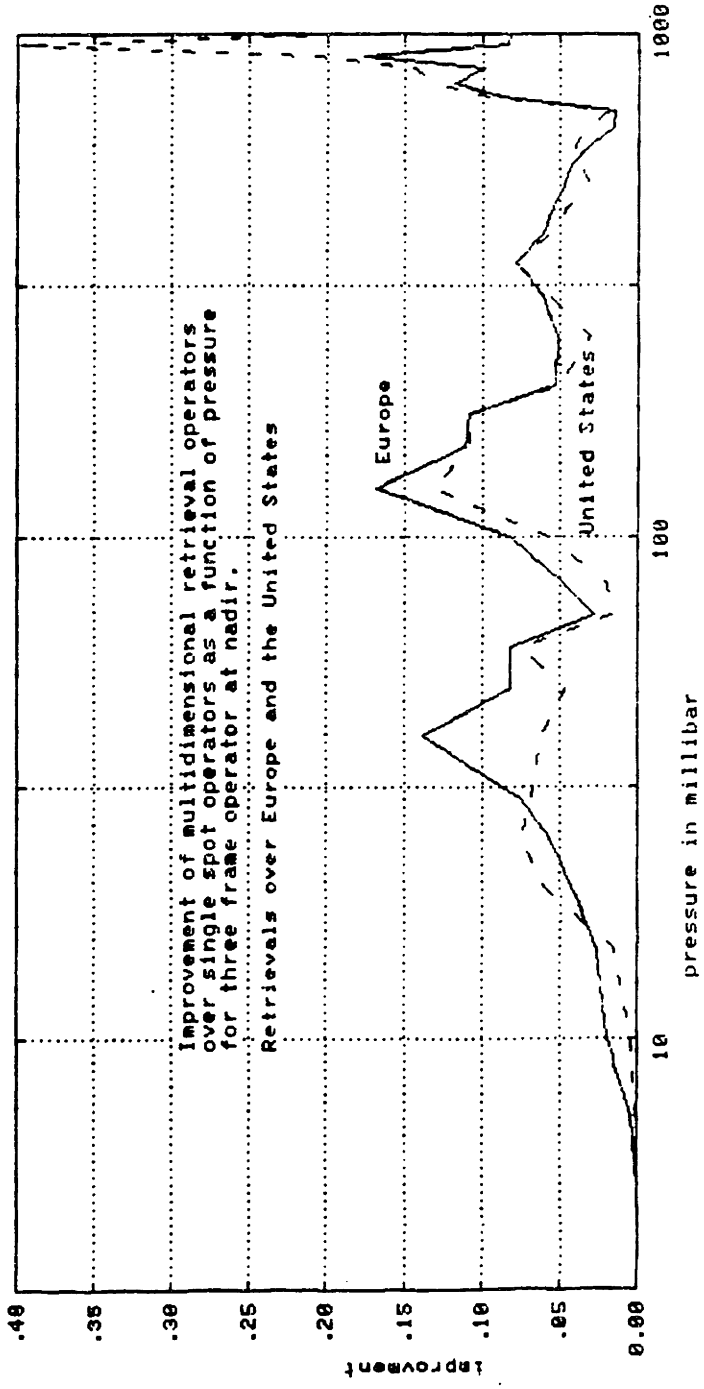


Figure VI.3.18: rms error improvement of single frame over single spot operators as a function of pressure for MSU at nadir in winter over Europe and the United States

parallel to the satellite and (2) the irregular sampling grid in the direction perpendicular to the track. However, when using the retrieval operators, the important factor determining the implementation is the number of multiplications required to estimate the temperature at any location within the grid and pressure level. The operational implementation can by performed in the spatial domain by direct convolution or in the frequency domain by multiplication.

Consider then two-dimensional retrieval operators parallel to the satellite track and compare the number of multiplications of the spatial implementation of this dissertation with the Fourier domain implementation of Nathan (1983). This analysis extends directly to three-dimensional operators. Because the retrieval operators are symmetric with respect to the central frame, for a retrieval operator N frames long (N odd), only $\frac{N+1}{2}C$ multiplications (C is the number of channels) are required per pressure level. This number of multiplications is found by noting that brightness temperatures located at identical locations on either sides of the central frame are multiplied by the same coefficients and thus can be added to each other before performing the multiplication by the retrieval operator. That is

$$\begin{aligned}\hat{T}(m) &= \sum_{n=-M+1}^{M-1} D(n)T_b(m-n) \\ &= \sum_{n=1}^{M-1} D(n)(T_b(m-n)+T_b(m+n)) + D(0)T_b(m)\end{aligned} \qquad (VI.4.1)$$

The implementation of Nathan is based on a filtering in the frequency domain using L=128 points FFTs. The total number of multiplications to estimate the temperature at L points (all with the same pressure level) is (Oppenheim and Shafer, 1975)

$$(\frac{L}{2}\log_2 L + L + \frac{L}{2}\log_2 L)C = CL + CL\log_2 L \qquad (VI.4.2)$$

or per location

$$(1+\log_2 L)C \qquad (VI.4.3)$$

which equals $8C$ for a 128-point implementation. A three-frame retrieval operator which captures most of the improvement going from one-dimensional to two-dimensional improvement requires only $2C$ multiplications. Even a five-frame retrieval operator ($3C$) multiplications presents a significant improvement over Nathan's FFT-based implementation.

Moreover, when the retrieval implementation is performed in the frequency domain, the multiplication of the signal's and operator's DFT's correspond to the circular convolution of the signal with the operator. This circular convolution is equivalent to the linear convolution of the operator with a periodically extended signal. This extension typically introduces discontinuities at the ends of the processed orbit which degrade the performances of the filter by increasing the high frequency error spectrum.

Three methods are available to correct for these discontinuities:

(1) One can pad the signal with zeroes before taking the Fourier transform but this wastes multiplications.

(2) One can reflect the signal with respect to one end before taking the Fourier transform, that is using the Discrete Cosine Transform, but this does not preserve the geographic continuity of the brightness temperature series.

or (3) One can choose the ends of the orbits being Fourier transformed so that the discontinuities are small. Nathan (1983) used that technique and ended the orbits at a latitude of 6° South using the fact that in a 20° band across the equator temperatures are fairly constant. Hence the physical and brightness temperatures are almost equal at the ends of the orbits. Unfortunately, this analysis forgets the fact the retrieval operators do not operate on temperatures but on their deviations from their respective means. Although the temperature are fairly constant across the equator, their deviations from the mean may not be. To be convinced of that fact, it suffices to compare the distance between two adjacent points on the MSU

sampling grid (168 km) with the distance between the ends of the orbits (19 000 km).

The issue of number of multiplications as well as the impact of discontinuities shows that the frequency domain implementation is not adequate for operational purposes.

VI.4.1 Measured Brightness Temperature and Computed Brightness Temperature

Before implementing retrieval operators on measured brightness temperature, two corrections must be considered to improve the quality of the retrievals: (1) the non-ideal beam pattern of the antenna and (2) the discrepancies between the measured brightness temperatures and the one derived from solving the forward problem from coincident NMC field or adjacent radiosonde measurements. These two phenomena must be corrected for two main reasons: (1) to reduce the observation and modeling noise used to derive the filters and (2) to compare the rms retrieval errors with those of regression-based retrieval algorithms which do not account for any physics (Nathan, 1983).

The correction for antenna beam efficiency is simplified by the good performance of sounding system antennas; this correction follows the method reported by Beck (1977). The temperature sensed by the radiometer is given by

$$T_a(\lambda) = \frac{\int_{4\pi} T_b(\theta,\phi,\lambda) P(\theta,\phi) d\Omega}{\int_{4\pi} P(\theta,\phi) d\Omega} \qquad (VI.4.4)$$

where θ and ϕ are the direction angles away from the location the antenna system points to, and $P(\theta,\phi)$ is the antenna pattern. The normalization by the integral of the power pattern sets the antenna temperature equal to the brightness temperature for an ideal beam pencil antenna (infinite resolution). The aim of the correction technique is the deconvolution of the brightness temperature by the antenna pattern.

Assuming circular symmetry (for the convenience of calculation, the results extend to all types of antenna patterns), the antenna temperature becomes

$$T_a(\lambda) = \frac{\int_0^{\pi} T_b(\theta) P(\theta) \sin\theta d\theta}{\int_0^{\pi} P(\theta) \sin\theta d\theta} \qquad (VI.A.5)$$

Recall that even for a constant temperature field over the entire field of view, the brightness temperature is a function of the angle θ because of the lifting of the weighting functions with the observation angle. Assuming the antenna beam efficiency is high, the integral can be separated into the main lobe and the sidelobes according to

$$\begin{aligned} T_a(\lambda) =& \frac{T_b(\lambda) \int_0^{\theta_0} P(\theta) \sin\theta d\theta + \int_{\theta_0}^{\pi} T_b(\theta) P(\theta) \sin\theta d\theta}{\int_0^{\pi} P(\theta) \sin\theta d\theta} \\ =& \eta(\theta) T_b(\lambda) + \frac{\int_{\theta_0}^{\pi} T_b(\theta) P(\theta) \sin\theta d\theta}{\int_0^{\pi} P(\theta) \sin\theta d\theta} \end{aligned} \qquad (VI.A.6)$$

where the brightness temperature is assumed constant over the main lobe (width θ_0) and $\eta(\theta_0)$ denotes the beam efficiency. For MSU, the nominal beam efficiency is set to 90 % but has been measured in the range 97-98%. The second term of the integral can be computed by considering neighboring spots within the same frame or in a different frame, effectively trying to perform an iterative or recursive deconvolution of the brightness with the antenna pattern, but it is time consuming (Njoku *et al*,1980). Equation (VI.4.6) can be approximated by the linear relationship

$$T_a(\lambda) = \eta T_b(\lambda) + (1 - \eta) < T_b(\lambda) > \qquad (VI.4.7)$$

where the mean brightness temperature $< T_b >$ is computed by taking the average over several spots. Any error in the value of the brightness temperatures of the neighboring points is multiplied by $(1 - \eta)$ which typically equals 0.025. A deviation of 5 degrees between the brightness temperature field and its mean in all neighboring points (quite a large deviation) yields an error of about 0.1K which is acceptable. Even if the deconvolution is performed adequately, the finite resolution

of atmospheric sounders will still cause uncorrectable errors between NMC analysis, derived from radio-sonde measurement which are point sounders and brightness inverted temperature profiles which are area sounders (Staelin *et al.*, 1975a, Ledsham, 1978).

The second element that requires correction is the difference between brightness temperature as computed from the NMC field, the so-called "ground truth" and the brightness temperature as measured by the instrument. This error has its origin in different places:

(1) The instrument's electronic is imperfect and an additive white Gaussian noise is added to the measurement. This noise is taken into account in the retrieval process. However systematic errors in the calibration of the instrument (MSU slopes and intercepts) are not taken into account by the observation equation and must be corrected for.

(2) The NMC field is not coincident with the measurement since it has been linearly interpolated in time and space to the satellite's position. This analysis field is computed every 12 hours to be compare with the hour and a half it takes the satellite to cross the equator line. This interpolation is adequate above land where the sampling grid of radiosondes is fine enough when the temperature features are slowly varying. However, it may be adequate to resolve certain rapidly varying structures such as storms.This makes the ground truth unreliable. Moreover, above ocean the NMC field relies heavily on the satellite data, making it extremely unreliable even in case of smooth temporal variations.

(3) The transmittance of the atmosphere, which is the basis for the computation of the weighting function, is not known with a perfect accuracy and may present modeling errors. Two important sources of error for the lowest probing channels are the ground elevation (which changes the bounds

of integration defining the weighting functions (Equation (II.3.14)) and the ground reflectivity, which controls the ground contribution to the overall radiation. Another source of error is the non-linearity in the weighting function's physics as described in appendix II.A. Moreover, in some instances, clouds are in the view of the instrument and perturbs the radiative transfer equation. Standard methods of correction from cloudy to clear profiles are possible but cumbersome to implement in the context of real retrievals (McMillin, 1978).

(4) The NMC field is not sampled at the 33 pressure levels used in the computation of the radiative transfer equation and stops at 50 *mbar*, which is far from appropriate for the highest probing channels of MSU and AMSU, especially when the sounding angle is 56°. Extending the profiles using a standard atmosphere or better, using the lapse rate of standard atmosphere is not adequate. This "shortness" of the ground truth is particularly damaging for the highest probing channels and can compromise the retrieval operator's ability to perform adequately. It also places retrieval's implementations based on the computation of the weighting functions at a net disadvantage with respect to linear regression methods which implicitly assume the temperature field is known exactly above 50 *mbar*. For MSU's channel 4, the weighting matrix has 21 % of its coefficients above 50 *mbar* (as measured by their sum) for nadir and up to 35 % at 56° which is a enormous proportion in either case. This absence of ground truth temperature above 50 *mbar* can translates up to 40 or 50 degrees of brightness temperature which must be compared with the rms of the sensor noise of about 0.2 degrees. This problem is exhibited in Figure VI.4.1 for MSU channel 2 (from Nathan, 1983). In this plot, the measured brightness temperature is compared with the brightness temperature computed from the NMC field and from the estimated temperature. Note that neither of them matches the observed measurements, even above

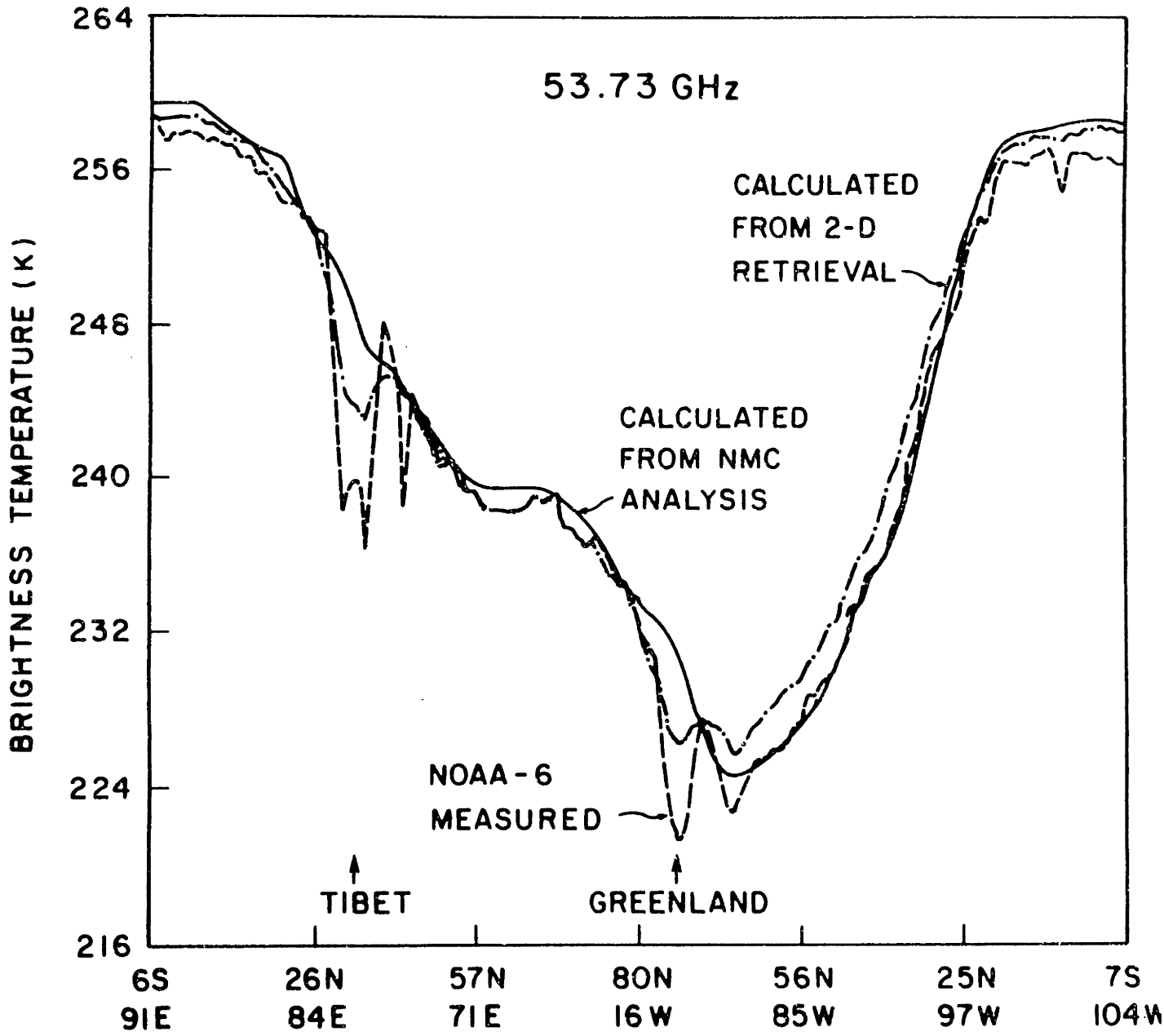


Figure VI.4.1: Observed brightness temperature (channel2) and brightness temperature computed from NMC analysis field (after Nathan, 1983)

ocean. Moreover, the differences between these quantities are much greater than what could expected from either measurement noise or non-linearities. Note, that in the case of MSU, the estimates for the observable components are almost independent of the retrieval operator because of the good signal to noise ratio. Hence, both NMC and estimated profiles observable compo-

nents should yield the same brightness temperature, which they do not. This demonstrate that the ground truth observable components are wrong. The impact of this phenomena will be explored into more details in the following section.

Since the theme of this dissertation is linear processing of data, consider a linear regression scheme to correct for the differences between computed and observed brightness temperature. Two different corrections can be considered: either correcting the would be computed brightness temperatures to the observed ones, as linear regression method do (Grody *et al*, 1985) or the opposite. Because the three-dimensional retrieval operators are derived in terms of the weighting matrices (and not simply in terms of observed cross-correlations), one must correct the observed brightness temperatures to their equivalent computed temperature. Choosing the opposite scheme is equivalent to changing the weighting matrices for the forward problem, this making the retrieval operators incompatible with the observation equation (Susskind and Reuter, 1985).

Let T_b^c and T_b^o denote the brightness temperatures computed from the ground truth and observed respectively. The regression can be written as

$$\hat{T}_b^c = \beta + \alpha T_b^m \qquad (VI.4.8)$$

where the parameters of the regression are (Westwater *et al.*, 1985)

$$\begin{aligned} \alpha &= s(1 + \sqrt{1 - \xi s^{-2}}) \\ \beta &= < T_b^c > - \alpha < T_b^m > \end{aligned} \qquad (VI.4.9)$$

where $< . >$ denotes the expectation and the parameter s equals

$$s = \xi \frac{K_{Tb^m Tb^m} - K_{Tb^o Tb^o}}{K_{Tb^m Tb^o}} \qquad (VI.4.10)$$

The tuning element ξ measures the ratio between the error associated with the measurements and the one associated with the calculation. In this dissertation, the

measurement error is the sum of the error associated with the sensor noise, impact of ground reflectivity, and changes the weighting functions caused by changes in elevation. The errors are independent of each other, and their respective variances can be added. The measurement error was set to five times the MSU nominal noise. The calculation error, assuming the ground truth and transmittances are exact, are caused by the non-linearities of the weighting functions, which was quantified in appendix II.A. Its rms was set to be 0.3 K. For MSU, the regression depends only slightly on the variable ξ.

Note that this correction is performed channel by channel. Theoretically, this correction is not as good as one based on all three channels at once, but it does not use the correlation between channels adjacent in frequency which depends on the climatologies used. The present correction is more robust to errors in the climate' statistics and the weighting functions.

Table VI.4.1 presents the correction coefficients for this correction scheme. The database for this correction is western Europe and the USA during three weeks worth of retrievals (one week in February 1981, July 1979, and October 1981). The ground truth is the 12 level NMC analysis field linearly interpolated in height between available pressure level, and extrapolated above 50 *mbar* using standard atmosphere lapse rate. Each sounding angle (numbered from one to six, one standing for nadir and six for 56° is corrected separately to account for the different weighting functions. Among all sounding angles, the correction coefficients for 56° depart the most from one for α (the slope) and from zero for β (the zero intercept). These departures are caused by the lifting of the weighting functions with angle which put more emphasis on temperature at high altitudes where the ground truth is absent and must be guessed.

VI.4.2 Validity Analysis of the Ground Truth

Because the ground truth for these remote sensing experiments is not pro-

alpha-beta coefficients derived from regression over Europe and USA

sensing angle (1-6, nadir-extreme) 1

	beta	alpha1	alpha2	alpha3
MSU channel 2	-13.547	1.042	0.000	0.000
MSU channel 3	3.130	0.000	0.968	0.000
MSU channel 4	9.286	0.000	0.000	0.942

sensing angle (1-6, nadir-extreme) 2

	beta	alpha1	alpha2	alpha3
MSU channel 2	-9.192	1.042	0.000	0.000
MSU channel 3	-6.937	0.000	1.032	0.000
MSU channel 4	-6.512	0.000	0.000	1.030

sensing angle (1-6, nadir-extreme) 3

	beta	alpha1	alpha2	alpha3
MSU channel 2	-15.465	1.069	0.000	0.000
MSU channel 3	-11.156	0.000	1.050	0.000
MSU channel 4	-9.335	0.000	0.000	1.042

sensing angle (1-6, nadir-extreme) 4

	beta	alpha1	alpha2	alpha3
MSU channel 2	-12.910	1.037	0.000	0.000
MSU channel 3	2.081	0.000	0.973	0.000
MSU channel 4	6.632	0.000	0.000	0.953

sensing angle (1-6, nadir-extreme) 5

	beta	alpha1	alpha2	alpha3
MSU channel 2	-7.437	1.034	0.000	0.000
MSU channel 3	-5.301	0.000	1.025	0.000
MSU channel 4	-4.926	0.000	0.000	1.023

sensing angle (1-6, nadir-extreme) 6

	beta	alpha1	alpha2	alpha3
MSU channel 2	-39.725	1.174	0.000	0.000
MSU channel 3	-20.000	0.000	1.088	0.000
MSU channel 4	-14.047	0.000	0.000	1.062

Table VI.4.1: alpha-beta coefficients for MSU

vided at the exact spatio-temporal locations where the measurements take place, the rms error computed by difference between the estimated profile and the interpolated NMC field is not the exact rms error. This error in the ground truth must thus be taken into account when comparing different retrieval operator, in

particular when comparing the improvement of multi-spot retrieval over single-spot retrieval with the improvement predicted in section VI.3. For instance, it would be incorrect an improvement in retrievals for a reduction of say 0.05 K if the rms error in the NMC field is of the order of 0.2 K. Likewise, quoting rms or bias errors up to the thousandth of Kelvin is outright ridiculous if the precision of the results is only on the order of a few hundredth of degrees. This section attempts to quantify the significance of quoted errors.

As could be expected, there is no method to determine what the error in the NMC field is, for if there was one, one would correct for the error. Moreover, statistics of the NMC error is not available. One can get, however, a measure of the variability of the ground truth by implementing the same retrieval operators on two satellites flying close to each other (TIROS-N and NOAA-6) and seeing basically the same climatology. One might argue that this analysis does not provide a true estimate of the NMC field's variability. However, it produces a good estimate of the variability of the rms retrieval error which is the quantity required to assess the differences between retrieval operators.

Table VI.4.2 presents the a-priori distribution and rms error associated with the minimum information solution for two satellites operating over Europe, the United States, Canada, and Japan in February 1981. The minimum information solution is chosen to avoid the usage of statistics (necessarily erroneous) in the retrieval process. It provides us only with the variability of the ground truth in the observable subspace of the instrument. Remember that the estimate for the unobservable components is a mapped version of the observable components's estimates. Hence any variability for the minimum information solution may be extended to the unobservable components.

The quantities presented in this table are the bias error, root mean square error and the standard deviation of the error (std) defined as the square root of the

February 9,10,11,12, and 13 1981

Latitude between 25 and 60, longitude between -130 and -60

Latitude between 60 and 70, longitude between -165 and -90

Latitude between 35 and 60, longitude between -15 and 35

Latitude between 25 and 45, longitude between 125 and 145

Statistics compiled using 11429 retrievals for TIROS-N

Statistics compiled using 6094 retrievals for NOAA-6

a-priori distribution

pressure	bias	rms	std	bias	rms	std
	NOAA-6			TIROS-N		
1000 mb	20.8	24.0	11.9	22.7	25.7	12.1
850 mb	19.4	21.9	10.1	20.8	23.2	10.4
700 mb	16.2	18.4	8.6	17.2	19.4	8.9
500 mb	14.4	16.4	7.8	15.1	17.1	8.0
400 mb	13.7	15.2	6.7	14.2	15.7	6.7
300 mb	11.5	12.4	4.9	11.9	12.8	4.7
250 mb	9.1	10.2	4.5	9.6	10.5	4.3
200 mb	5.7	7.8	5.4	6.3	8.3	5.3
150 mb	2.3	7.3	7.0	3.0	7.8	7.2
100 mb	2.9	7.8	7.2	3.6	8.3	7.4
70 mb	5.9	8.9	6.6	6.5	9.4	6.7
50 mb	8.7	10.3	5.6	9.3	10.8	5.6

minimum information solution

pressure	bias	rms	std	bias	rms	std
	NOAA-6			TIROS-N		
1000 mb	-6.5	14.9	13.4	-4.0	12.7	12.1
850 mb	11.8	13.6	6.7	11.8	13.5	6.5
700 mb	3.8	5.2	3.6	4.8	5.9	3.5
500 mb	-3.7	4.9	3.2	-1.9	3.3	2.6
400 mb	-4.5	6.1	4.1	-3.1	4.6	3.4
300 mb	-2.4	4.5	3.8	-2.4	4.1	3.4
250 mb	-0.7	3.2	3.1	-2.0	3.5	2.9
200 mb	1.6	3.8	3.4	-0.3	3.2	3.2
150 mb	0.9	5.1	5.0	-1.3	4.8	4.6
100 mb	-1.2	4.8	4.7	-3.2	5.6	4.6
70 mb	2.6	3.8	2.7	1.7	3.1	2.6
50 mb	4.7	6.0	3.6	4.0	5.5	3.7

Table VI.4.2: A-priori distribution and minimum information solution for two MSU satellites over Europe, USA, Canada, and Japan during week of February 1981.

difference between mean square error and bias error squared. The later quantity is sometimes referred to as rms about mean. The standard deviation corresponds the value of the rms error in cases where the a-priori mean for the temperature set is known exactly latitude by latitude.

The difference between the a-priori distribution rms' are in general no greater than 0.5 K for levels higher than 700 *mbar* and can be as large as 1.7 K close to the ground. This variability in the low tropospheric temperature is to expected (see Chapter III), and it is pronounced in this specific case study by the weather phenomena present (see below). It is mostly caused by difference in the a-priori bias as attested by the smaller variations of the a-priori standard deviation from one satellite to the other which does not differ by more than 0.3 K.

Although the a-priori distribution standard deviations are close, the standard deviation for the MIE differs greatly: 1.3 K at 1000 *mbar*, 0.7 K at 400, and less than 0.7 K at higher altitudes. The two instruments used are identical and for the same brightness temperature deviations, they will provide the same estimate. The NMC analysis fields on the other hand differ by the amount of bilinear interpolation (mostly temporal interpolation). This result indicates that the quality of the ground truth varies with the amount of temporal and spatial interpolation required to produce it. It also points out that one cannot compare retrieval operators implemented on close but different datasets, because variations up to 1 K can be expected solely from the impact of interpolation.

The difference between the rms errors of the two minimum information solutions should be less than the difference between the a-priori distributions if both ground truths are correct. This is always the case, expect at 500 and 400 *mbar* where the difference between the rms errors is larger than the difference between a-priori rms distributions by about 1 K, and at 70 and 50 *mbar* where the difference is about 0.2 K. Whereas the latter difference can be attributed to reconstruction er-

rors in the estimated profile, the difference in the troposphere cannot be accounted for by estimation error. It implies that the NMC field may be off by about 0.5 K. Thus, at those pressure levels, unless the improvement in the error's standard deviation is greater that about 0.5 K, it may not be significant.

Note that in the a-priori rms distribution are quite large compared the values used in the simulations. The a-priori rms distributions are up to 6 times as large as the summer cases simulations, up to 4 times the winter cases cases. This correspond to climatologies with up to 36 times the amount of power. This will have two important consequences: (1) All rms errors will be larger than those of simulations, typically four times as large, (2) any improvement provided by multidimensional retrieval operators, if it correspond to the fraction computed in the simulations should be large enough to be significant with respect to the ground truth variability. These large a-priori distributions are caused by the larger number of degrees of freedoms of the climatology, in one case up to 11429 different profiles are considered in the compilation. Finally, one might question the amplitude of the a-priori bias in the troposphere (15 to 20 K). In early February 1981, a intense cold front was over the US and Canada which consists of about 68 % of all soundings. This front is some of the ultimate test for multidimensional operators since its vertical statistics is not too reliable. Note that this case was considered in Nathan (1983) as a case study. This storm presents temperature profile with another inversion close to the ground, where the temperature increases with height. These changes in the lower tropospheric lapse rate explains the 1.7 K difference between the a-priori rms distributions.

Section VI.3 proved that the maximum of percentage one can expect in the error standard deviation (or rms error for unbiased retrievals) is of the order of 10 to 15 % (bigger close to the ground). For this fractional improvement to translate into reduction of rms error greater than 0.3 K requires us to consider experiments

where the single spot retrieval error is of the order of 3 to 4 K. This will achieved by performing most the analyses on the winter cases associated with the a-priori distributions just considered.

VI.4.3 Results

In this section, the same experiments conducted in section VI.3 are performed using measured brightness temperatures. The major difference with the previous section is that because the operators are not implemented on the exact same datasets they are derived from, they are not truly optimal. This non optimality is characteristic of operational implementations. The other difference, somewhat a consequence of the first one, is that the bias error plays an important part of the retrieval rms error. This is true because the a-priori mean profile is chosen independently of the retrieval dataset.

Table VI.4.3 compares the rms and bias error as well as the standard deviation of the error for a single-spot and single-frame retrieval operator derived from summer statistics operating over a week of July 1979.

For the July retrievals, two different a-priori mean vectors were chosen to illustrate the impact of the bias error on the rms error. The impact of the a-priori mean is especially important in the non-observable subspace. Note that the rms error varies substantially between experiments whereas differences between corresponding standard deviations are smaller. As for the simulations, most of the improvement of statistical filters over the minimum information (physical solution) is provided by the incorporation of vertical statistics (vertical covariance kernel) rather than by isobaric statistics. Most of the improvement in using two-dimensional operator takes place in regions where temperature profiles vary greatly with pressure and where one can expect the horizontal correlation's characteristics to change with height: These regions are close to the ground, at the top of the troposphere and bottom of the stratosphere. At the 50 *mbar* pressure, the improvement is larger

latitude between 34 and 54, longitude between -135 and -75
latitude between 35 and 55, longitude between -10 and 30
statistics compiled using 3223 retrievals, all sounding angles
retrievals performed July 2,3,4,5 and 6 1979
a-priori mean: Interpolated standard atmosphere

pressure	bias	rms	std	bias	rms	std	bias	rms	std
1000 mb	-0.8	5.9	5.8	-3.7	5.3	3.8	-1.1	3.8	3.6
850 mb	0.3	3.3	3.3	-1.3	2.5	2.2	-1.3	2.4	2.0
700 mb	0.0	1.9	1.9	-0.6	1.4	1.3	-1.1	1.4	0.9
500 mb	0.3	2.3	2.2	0.8	1.7	1.5	0.8	1.6	1.3
400 mb	-0.5	2.5	2.2	1.6	2.2	1.5	0.9	1.7	1.4
300 mb	0.9	2.1	1.9	1.7	2.4	1.7	0.7	1.7	1.5
250 mb	3.1	3.8	2.3	1.7	2.4	1.7	1.9	2.5	1.6
200 mb	0.2	1.7	1.7	-0.7	2.1	2.0	-0.8	2.2	2.0
150 mb	-1.8	3.2	2.6	-2.1	3.2	2.4	-1.3	2.3	1.9
100 mb	-0.3	2.3	2.3	-1.3	2.4	2.0	-0.4	1.8	1.8
70 mb	-0.6	1.1	1.0	-0.3	0.8	0.7	-0.7	1.1	0.8
50 mb	0.6	2.9	2.8	1.7	2.5	1.9	1.0	1.8	1.6
	minimum information			single spot			single frame		

a-priori mean: sample mean July 7 and 8 1979
statistics compiled using 3223. retrievals, all sounding angles

pressure	bias	rms	std	bias	rms	std	bias	rms	std
1000 mb	-1.6	4.1	3.8	-1.5	4.1	3.8	-1.5	3.8	3.5
850 mb	-0.4	3.3	3.3	-0.5	2.1	2.1	-0.3	2.0	2.0
700 mb	0.2	1.5	1.5	0.0	1.2	1.2	0.1	1.2	1.2
500 mb	0.6	2.3	2.2	0.3	1.5	1.4	0.4	1.4	1.3
400 mb	0.3	1.9	1.9	0.4	1.5	1.5	0.4	1.5	1.4
300 mb	-0.2	1.9	1.9	0.3	1.6	1.5	0.2	1.5	1.5
250 mb	0.3	2.1	2.1	0.0	1.5	1.5	0.2	1.5	1.5
200 mb	-0.3	1.7	1.7	-0.6	1.8	1.7	-1.0	2.1	1.9
150 mb	-0.5	2.8	2.8	0.1	2.4	2.4	-0.1	1.9	1.9
100 mb	-0.1	2.2	2.2	-0.1	2.0	2.0	0.0	1.8	1.8
70 mb	-0.1	0.9	0.9	-0.3	0.7	0.6	-0.3	0.8	0.7
50 mb	0.2	1.8	1.8	0.1	1.8	1.8	0.1	1.5	1.5
	minimum information			single spot			single frame		

Table VI.4.3: Error analysis for week of July 1979 for single-spot and single-frame retrieval operators.

than the amount predicted from fitting the model to summer statistics (22 % measured versus 15 % predicted). This improvement is present for both experiments. It indicates that the isobaric correlation is better at high altitudes ($p \leq 50mbar$) than predicted. Close to the ground, the improvement is of the order of 20 % which is agreement with the predicted results. The improvement is most typically in the 10 to 15 percent range. However, at 200 and 150 *mbar*, the two-dimensional retrieval operators performs alike or slightly worse than the corresponding single spot estimation, at 150 the error is even 0.3 to 0.5 K worse than the minimum information solution. The statistics used to derive the retrieval operator are characteristics of northern hemisphere's summers. The covariance kernel used to derive the filter presents a sharper discontinuity at the tropopause than the temperature field under analysis (compare tables VI.3.1 and VI.3.2) The increase in error is the penalty paid for using an inadequate negative correlation structure across the tropopause. This experiment shows that the use of multidimensional retrieval algorithms will not always improve the remote sensing of temperatures if the covariance kernel used to derive the filters is not close to the sample covariance kernel. At 70 *mbar* the error also slightly increases but this error is dominated by an increase in bias error, this implies that the error standard deviation barely degrades. This result must be put in perspective with Figure VI.3.12 where the improvement was found to present a local minimum. Note also that this degradation is small in magnitude and may be blamed on inaccurate ground truth.

The impact of using measurements performed under different observation conditions was studied in the simulations by comparing the rms error for two two-dimensional retrieval operators sharing the same length of support, on parallel to the satellite's track where all measurements are performed at nadir and one perpendicular to the satellite's track that takes advantage of all possible sounding angles. Table VI.4.4 present the implementation results for these two operators over the July dataset. As for the simulations, the perpendicular operator outperforms the

latitude between 34 and 54, longitude between -135 and -75
latitude between 35 and 55, longitude between -10 and 30
statistics compiled using 3223 retrievals, all sounding angles

pressure	bias	rms	std	bias	rms	std
1000 mb	-1.1	3.8	3.6	-2.0	5.0	4.6
850 mb	-1.3	2.4	2.0	-1.5	2.5	2.1
700 mb	-1.1	1.4	0.9	-1.0	1.8	1.6
500 mb	0.8	1.6	1.3	1.0	1.7	1.5
400 mb	0.9	1.7	1.4	1.3	2.1	1.7
300 mb	0.7	1.7	1.5	1.3	2.2	1.8
250 mb	1.9	2.5	1.6	2.2	2.8	1.8
200 mb	-0.8	2.2	2.0	-1.0	2.2	2.0
150 mb	-1.3	2.3	1.9	-2.0	3.2	2.5
100 mb	-0.4	1.8	1.8	-1.0	2.3	2.1
70 mb	-0.7	1.1	0.8	-0.4	0.9	0.8
50 mb	1.0	1.8	1.6	1.4	2.4	1.9
	perpendicular			parallel		

Table VI.4.4: retrieval errors for two two-dimensional retrieval operators: one parallel and one perpendicular

parallel operator, by an amount even larger than the one predicted in section VI.3. Note that the results presented are compiled using all sounding angles, that is the results for the parallel operator is compiled from six different retrieval operators, one for each sounding angle. Because the two retrieval operator have the same length (1.1 Mm), the rms error for the parallel filter is typical of what would be obtained using a perpendicular operator after using a limb correction algorithm. This result reinforces the importance of using measurements performed at different angles for MSU and of not using limb correction schemes in order to take advantage of the different SVD decompositions along the frame.

Table VI.4.5 presents the rms, bias, and standard deviation errors for the single-spot and single-frame operators applied on the February dataset. The same type of results found for the July case applies to the February retrieval, albeit for a factor. Note that the rms error are substantially larger than for the summer

February 9,10,11,12, and 13
A-priori mean interpolated standard atmosphere
Latitude between 25 and 60, longitude between -130 and -60
Latitude between 60 and 70, longitude between -165 and -90
Latitude between 35 and 60, longitude between -15 and 35
Latitude between 25 and 45, longitude between 125 and 145
Statistics compiled using all sounding angles and 6094. retrievals

pressure	bias	rms	std	bias	rms	std
1000 mb	0.6	5.7	5.7	4.4	5.5	3.3
850 mb	-2.7	4.4	3.5	-2.7	4.4	3.5
700 mb	-1.5	2.9	2.4	-1.6	2.8	2.4
500 mb	-0.6	2.5	2.5	-0.4	2.4	2.3
400 mb	-0.2	2.2	2.2	-0.1	2.3	2.3
300 mb	-0.6	2.3	2.2	-1.0	2.2	2.0
250 mb	-0.2	2.7	2.7	0.5	2.7	2.7
200 mb	-0.7	3.0	2.9	-1.3	3.4	3.2
150 mb	-2.9	4.3	3.2	-3.9	4.9	2.9
100 mb	2.5	4.7	4.0	1.3	3.8	3.5
70 mb	3.2	4.2	2.6	1.7	3.5	3.2
50 mb	4.2	5.3	3.3	3.6	5.1	3.6
	single spot			single frame		

Table VI.4.5: Error analysis for week of February 1981 for single-spot and single-frame retrieval operators.

case; this is caused by the larger spread of the distribution of profiles. The a-priori distribution as well as the minimum information solution associated with the same dataset is presented in Table VI.4.2 The improvement in the troposphere is less substantial than for the July case, where as the improvement in the stratosphere is larger (in proportion). The tropospheric improvement is mostly the result of the reduction of the bias error. At for the summer case, the two-dimensional operator has difficulty performing around the tropopause inversion (around the 150 *mbar* level). Contrary to the July case, the rms error does not degrade at 70 *mbar* going from one to two-dimensional retrievals although the standard deviation worsens because of a reduction of the bias error.

February 9,10,11,12, and 13 1979.

a-priori mean: interpolated standard atmosphere.

latitude between 60 and 90

statistics compiled using 8008 retrievals, all sounding angles

pressure	bias	rms	std	bias	rms	std	bias	rms	std
1000 mb	-3.0	7.5	6.9	-3.1	7.5	6.9	-3.5	7.7	6.9
850 mb	-10.0	10.5	3.4	-8.6	9.3	3.3	-7.8	8.5	3.3
700 mb	-1.3	3.4	3.1	-1.3	3.3	3.1	-1.2	3.3	3.1
500 mb	2.9	4.7	3.7	2.2	4.2	3.8	1.9	4.0	3.5
400 mb	1.9	3.4	2.9	1.5	3.2	2.9	1.3	3.2	2.9
300 mb	-3.0	3.5	1.7	-3.0	3.5	1.7	-3.0	3.5	1.8
250 mb	0.9	2.0	1.8	0.5	1.9	1.9	0.1	1.9	1.9
200 mb	-2.3	4.0	3.3	-1.8	3.7	3.2	-1.6	3.5	3.2
150 mb	-0.6	2.3	2.2	-0.1	2.2	2.2	0.3	2.2	2.1
100 mb	7.9	8.3	2.5	7.6	8.0	2.5	7.4	7.9	2.6
70 mb	7.1	7.7	2.9	6.8	7.3	2.9	6.5	7.1	2.9
50 mb	2.8	4.5	3.5	3.0	4.6	3.5	3.1	4.6	3.4

latitude between 0 and 30

statistics compiled using 9966 retrievals, all sounding angles

pressure	bias	rms	std	bias	rms	std	bias	rms	std
1000 mb	-11.3	12.1	4.2	-11.5	12.2	4.2	-11.6	12.3	4.2
850 mb	5.8	6.9	3.8	5.4	6.6	3.7	5.3	6.4	3.6
700 mb	3.7	4.2	2.0	3.8	4.4	2.0	3.9	4.4	2.0
500 mb	0.1	1.8	1.8	0.5	1.8	1.7	0.6	1.8	1.7
400 mb	-0.4	1.8	1.7	-0.2	1.7	1.7	-0.2	1.7	1.7
300 mb	-5.4	5.6	1.6	-5.5	5.7	1.6	-5.7	5.9	1.6
250 mb	-8.2	8.5	2.3	-8.5	8.8	2.3	-8.7	9.0	2.3
200 mb	-11.5	11.7	2.3	-11.8	12.0	2.3	-11.9	12.1	2.3
150 mb	-0.5	1.6	1.5	-0.6	1.5	1.4	-0.5	1.5	1.4
100 mb	1.7	2.4	1.7	2.1	2.6	1.6	2.4	2.9	1.6
70 mb	3.9	5.1	3.3	4.2	5.3	3.3	4.4	5.5	3.3
50 mb	6.3	6.8	2.6	6.2	6.7	2.6	6.0	6.6	2.6
	1 frame operator			3 frame operator			5 frame operator		

Table VI.4.6: Error analysis for week of February 1981 for single-, three- and five-frame retrieval operators: severe climates (polar and tropical)

February 9,10,11,12, and 13 1979.

a-priori mean: interpolated standard atmosphere.

latitude between 45 and 60

statistics compiled using 4345 retrievals all sounding angles

pressure	bias	rms	std	bias	rms	std	bias	rms	std
1000 mb	-4.5	7.4	5.9	-4.7	7.5	5.8	-5.0	7.6	5.8
850 mb	-6.5	7.5	3.8	-5.7	6.8	3.8	-5.2	6.4	3.8
700 mb	-0.6	2.9	2.8	-0.6	2.9	2.8	-0.5	2.8	2.8
500 mb	1.6	3.1	2.6	1.2	2.8	2.5	1.0	2.7	2.5
400 mb	0.7	2.3	2.2	0.5	2.2	2.1	0.4	2.1	2.1
300 mb	-3.2	3.6	1.8	-3.1	3.6	1.8	-3.1	3.6	1.8
250 mb	0.5	2.0	2.0	0.2	2.0	2.0	-0.1	2.0	2.0
200 mb	-1.1	3.4	3.2	-0.8	3.3	3.2	-0.7	3.3	3.2
150 mb	0.3	2.3	2.2	0.7	2.3	2.2	0.9	2.4	2.3
100 mb	6.0	6.7	2.9	5.8	6.5	2.9	5.7	6.4	2.9
70 mb	5.7	6.3	2.8	5.5	6.1	2.7	5.3	6.0	2.7
50 mb	3.8	4.5	2.4	3.9	4.6	2.4	4.0	4.6	2.4

latitude between 30 and 45

statistics compiled using 4708 retrievals, all sounding angles

pressure	bias	rms	std	bias	rms	std	bias	rms	std
1000 mb	-10.1	12.6	7.4	-10.3	12.7	7.4	-10.6	12.9	7.4
850 mb	-1.6	4.6	4.3	-1.3	4.5	4.3	-1.1	4.4	4.3
700 mb	0.1	3.1	3.1	0.1	3.1	3.1	0.1	3.1	3.1
500 mb	0.0	3.1	3.1	-0.1	3.0	3.0	-0.2	3.1	3.1
400 mb	-0.2	2.5	2.4	-0.3	2.5	2.5	-0.3	2.5	2.5
300 mb	-3.6	4.1	2.0	-3.6	4.1	2.1	-3.6	4.1	2.1
250 mb	-1.6	3.4	3.0	-1.9	3.6	3.0	-2.2	3.7	3.0
200 mb	-4.2	5.7	3.8	-4.2	5.7	3.8	-4.1	5.7	3.9
150 mb	-0.7	2.2	2.0	-0.6	2.1	2.1	-0.4	2.1	2.1
100 mb	3.3	3.6	1.6	3.3	3.6	1.6	3.3	3.7	1.6
70 mb	4.3	4.7	1.7	4.3	4.6	1.8	4.3	4.6	1.8
50 mb	4.2	4.5	1.8	4.2	4.6	1.8	4.2	4.6	1.8
	1 frame operator			3 frame operator			5 frame operator		

Table VI.4.7: Error analysis for week of February 1981 for single-, three- and five-frame retrieval operators: temperate climates

Contrary to the summer case, the bias error plays an important role in the

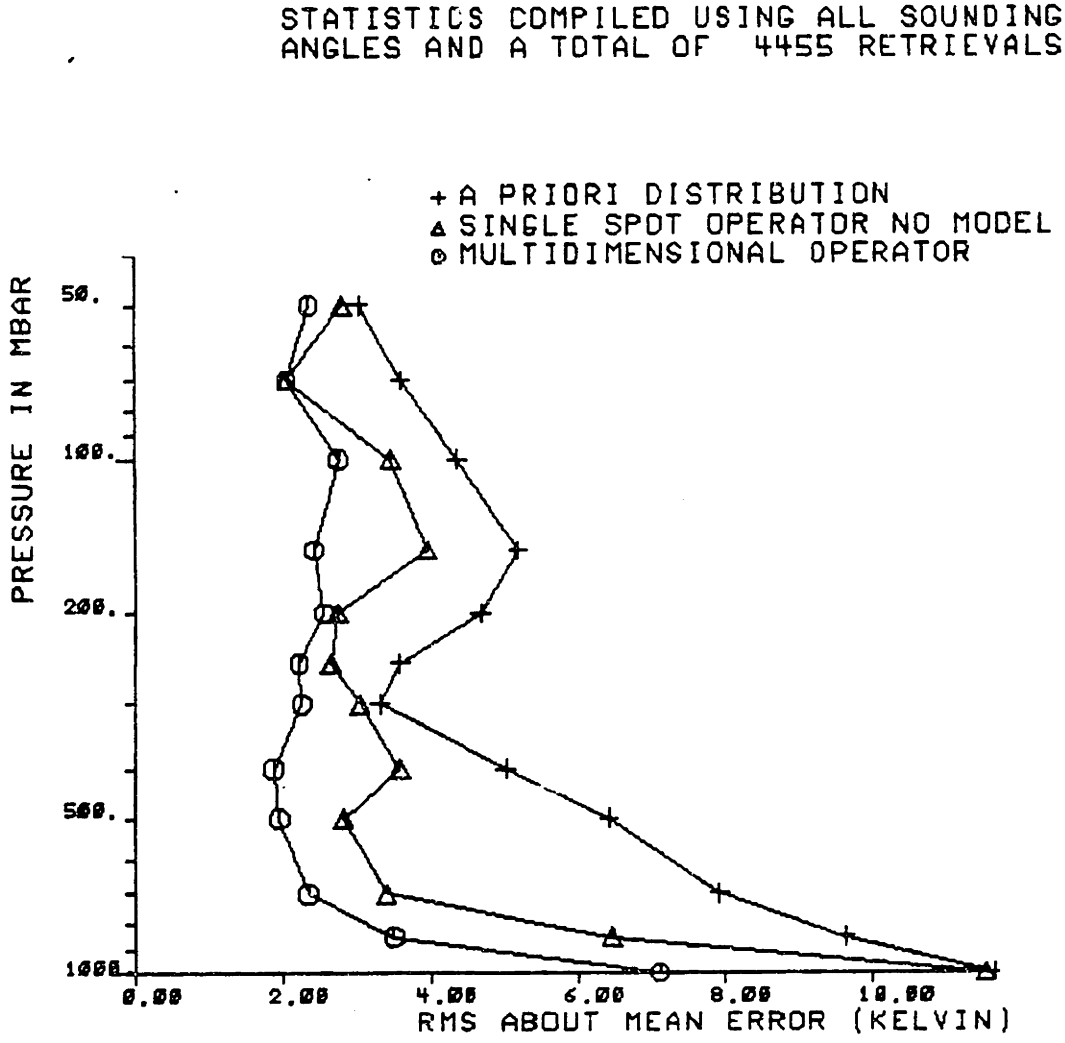


Figure VI.4.2: a-priori distribution, minimum information solution and three dimensional operator rms about mean error over Europe

overall performances of the operator. This is the case because the a-priori mean error, the result of the presence of the a storm above the United States and Canada for most of the retrieval period, is larger than in the July case.

Figures VI.4.2 and VI.4.3 present the rms about mean errors (standard deviation) over Europe and the United States for single spot estimation (minimum information and D matrix) as well as the error associated with the three-frames

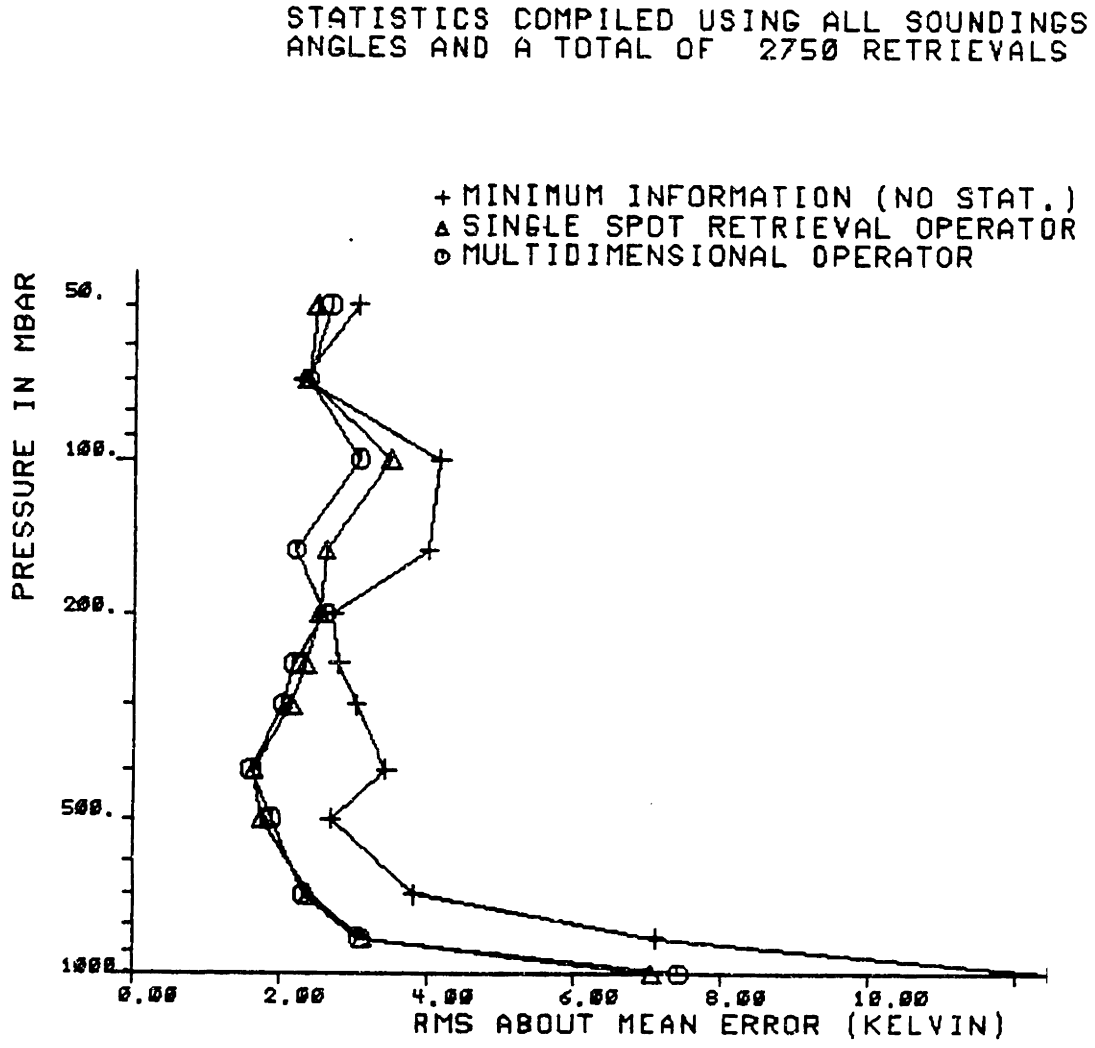


Figure VI.4.3: a-priori distribution, single spot and three dimensional operator rms about mean error over USA

three-dimensional retrieval operator. At low altitudes most of the error reduction is provided by the measurements and there is little room for improvement using multi-spot retrieval operators. At high altitudes, the inclusions of horizontal correlation reduces error except in the vicinity of 70 *mbar* where the horizontal and vertical statistics becomes unreliable.

The difficulty in retrieving the temperature around the tropopause inversion

and the possible worsening of the estimation with the region of support of the retrieval operator comes from the variations of the tropopause inversion height with climatologies and the fact that the model for the three-dimensional covariance kernel reinforces the impact of the negative correlation across the inversion. Hence, if a single-spot operator places the inversion at the wrong height (thus degrading the retrieval error), the associated multi-spot operators based on the same statistics will also place the inversion at the wrong height and the error will increase.

Tables VI.4.6 and VI.4.7 present the rms error for three retrieval operators during February 1981 for 4 bands in the northern hemisphere: $0^o - 30^o$, $30^o - 45^o$, $45^o - 60^o$, and $60^o - 90^o$. These bands include radiosonde rich areas such as the US and Europe as well as radiosonde poor areas such the Atlantic and Pacific Ocean. Bear in mind that when no other sources of data is available, the NMC analysis field relies heavily on the satellite data. Thus the rms errors will in general be lower than the one on land only. The operators considered have one, three, and five frames. These filters are derived using winter statistics over the US. The improvement in rms gained by using an extended three-dimensional operator over a two-dimensional operator varies typically from 0 to 0.3 K. The largest improvement is usually at 850 *mbar* were it varies from 0.3 K for the $30 - 45^o$ band to 2 K for the polar band. For the tropical band, the inclusion of multiple spots in the retrieval process has a worsening effect for the stratosphere. This is true because the covariance kernel used to derive the retrieval operators is not representative of the sample statistics over that band. Note that, with very few exceptions, most of the improvement in the rms error comes from a reduction of the bias error (the standard deviation remains almost constant from operator to operator)

VI.5 Predicted Retrieval Errors for AMSU

The major difference between MSU and AMSU is the number of channels

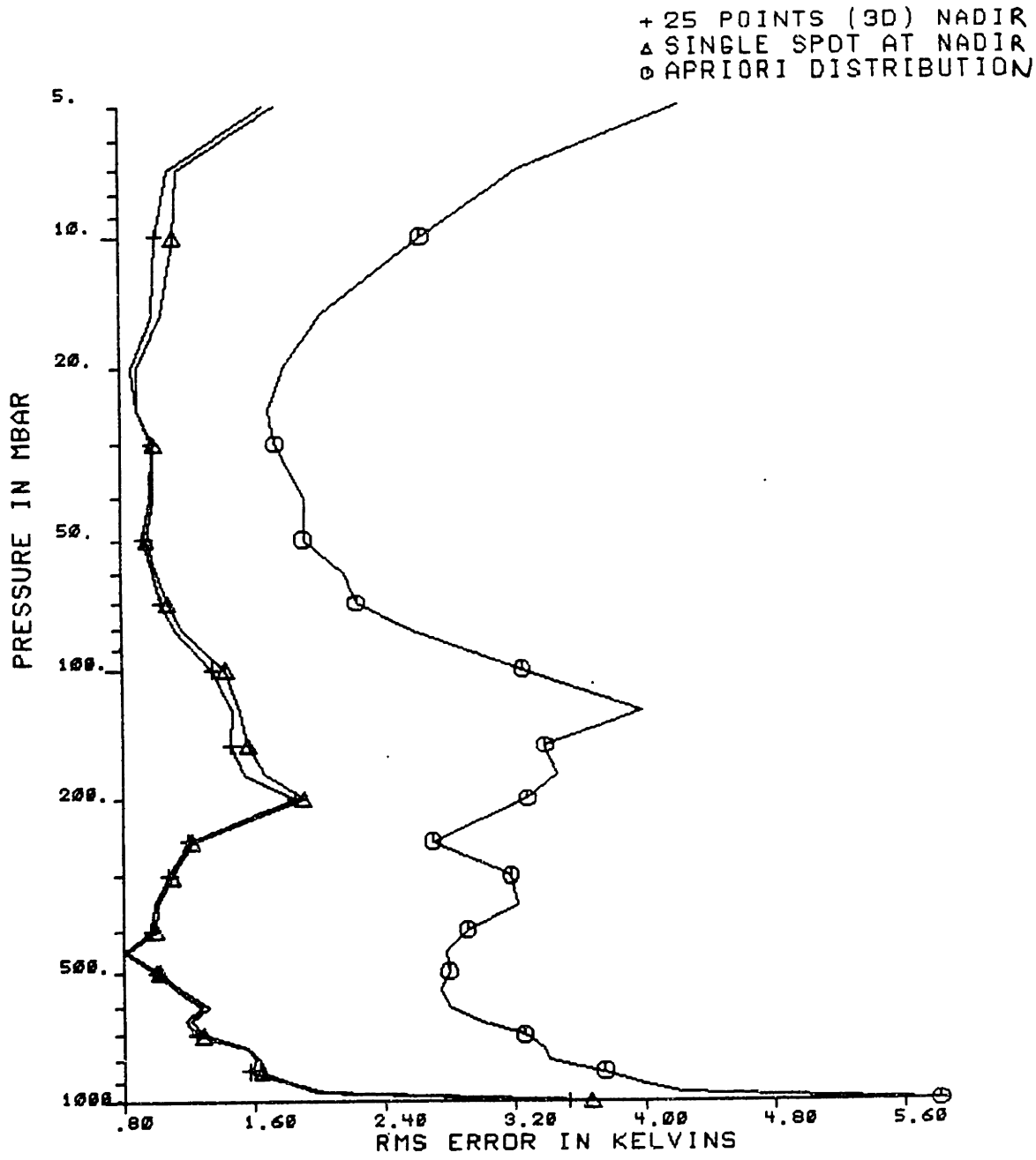


Figure VI.5.1: Expected rms error for single-spot and three-dimensional retrieval operators for AMSU at nadir in summer and the corresponding a-priori distribution.

used to sense the atmosphere. This increased sounding ability puts a larger proportion of the signal in the observable space of the instrument, up to 92 % (Baumann,

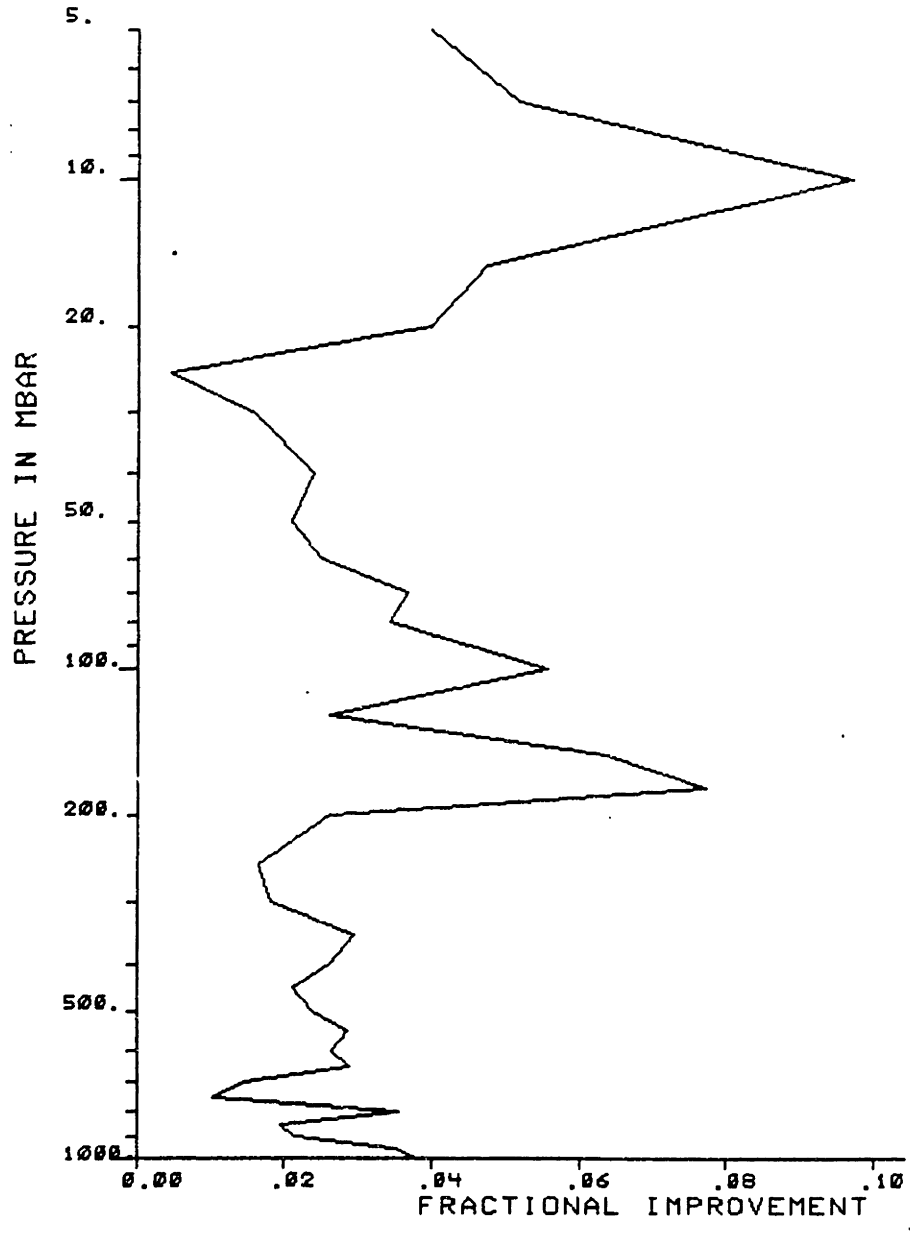


Figure VI.5.2: Improvement of three-dimensional operator over single-spot operator for AMSU over summer case

1980). Hence, in view of the Chapter V results, one can expect the retrieval operators to be narrower (for a fixed region of support) for the AMSU than for the MSU. Moreover, the impact of multidimensional retrieval operators is dictated in

the case of AMSU by the elimination of the receiver noise through low-pass filtering of the temperature fields. Moreover, because the sampling raster is more compact for AMSU, the differences in sounding angle between adjacent points in a frame are much smaller than the differences for MSU. This results implies that the lifting of the weighting functions, instrumental in the reduction of rms error for MSU, will not play as important a role for AMSU. For these reasons, the impact of multidimensional retrieval operators on the retrieval error is greatly diminished for AMSU.

While sweeping in the direction perpendicular to the satellite's track, the AMSU antenna samples the 12 temperature sounding channels at 30 different locations. This large number of measurements makes the computation of two-dimensional perpendicular and three-dimensional retrieval operator time consuming. Moreover, the covariance matrix corresponding to these 360 brightness temperatures (there were only 33 measurements per frame for MSU) may become ill-conditioned.

Because the expected improvement through the inclusion of multiple spots in the retrieval process was small, the analysis of the impact of statistics for AMSU was not conducted as thoroughly as for MSU. Figure VI.5.1 presents a typical expected rms error for a single spot retrieval operator and a centered 5×5 three-dimensional retrieval operator in summer. The summer dataset is given by the three-dimensional for the covariance fitted to July and August data over the United States (see section IV.4) Both operators are centered at nadir. The three-dimensional operator has a square support. Note that the improvement is much smaller than for MSU. In the troposphere, the improvement is almost non-existent as illustrated by Figure VI.5.2 which plots the variation of the improvement with height. The improvement is of the order of a couple of percent in the troposphere; it is slightly larger at the tropopause and in the stratosphere. The improvement at higher altitudes ($p < 25mbar$) comes

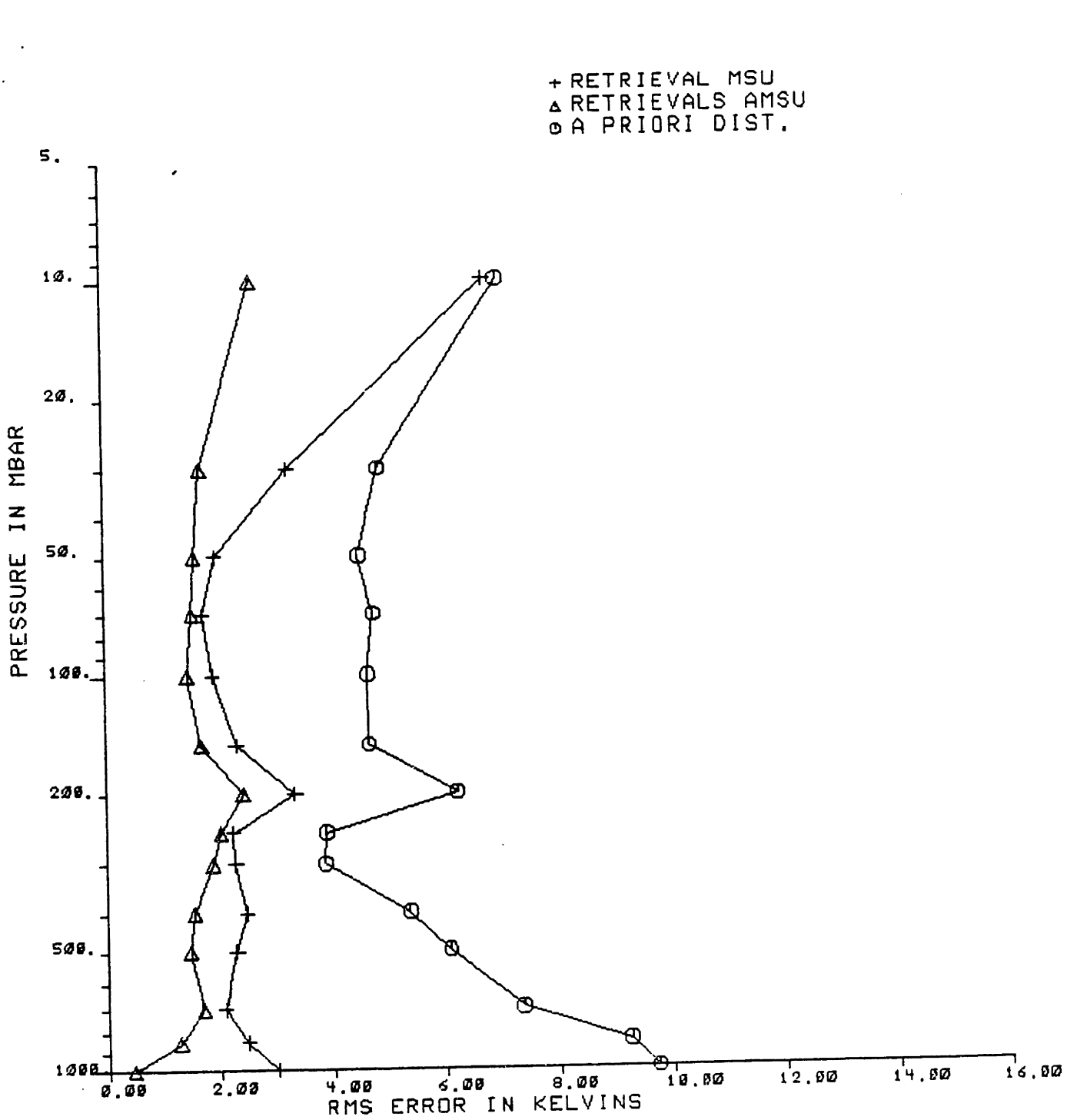


Figure VI.5.3: Rms errors for MSU, AMSU and a-priori distribution for winter case. Retrievals are performed at nadir above land

from the fact that the a-priori variance of temperature at high altitudes is smaller than close to the ground. Moreover, the high altitude sounding channels of AMSU have narrow frequency bandwidth and are characterized by large receiver noise

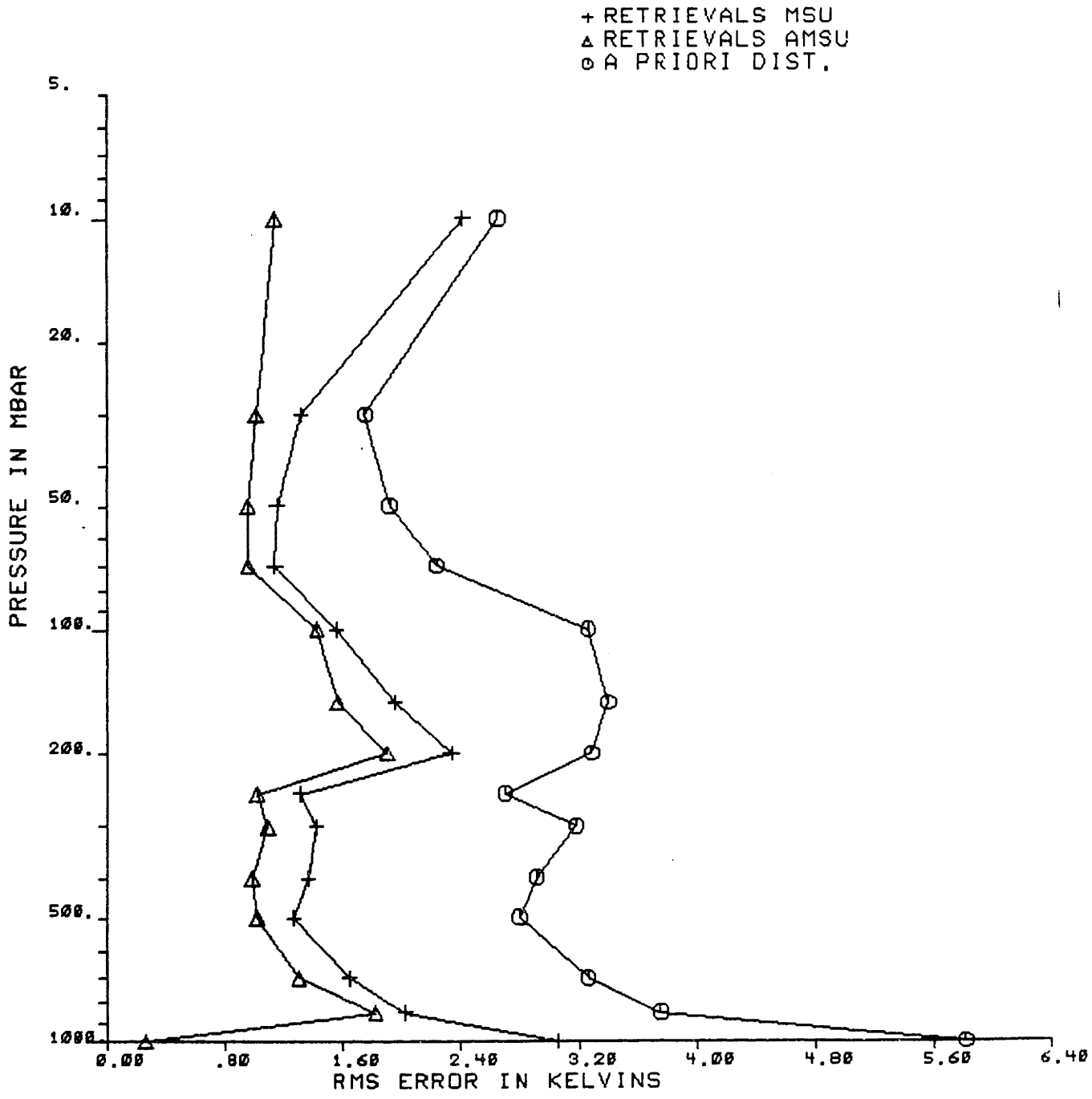


Figure VI.5.4: Rms errors for MSU, AMSU and a-priori distribution for summer case. Retrievals are performed at nadir above land

(rms levels greater than 1K). Thus at those altitudes, the improvement provided by multidimensional filtering comes from the low-pass filtering of the sensor noise. The 36 points (6×6) retrieval operator found by extending the region of support

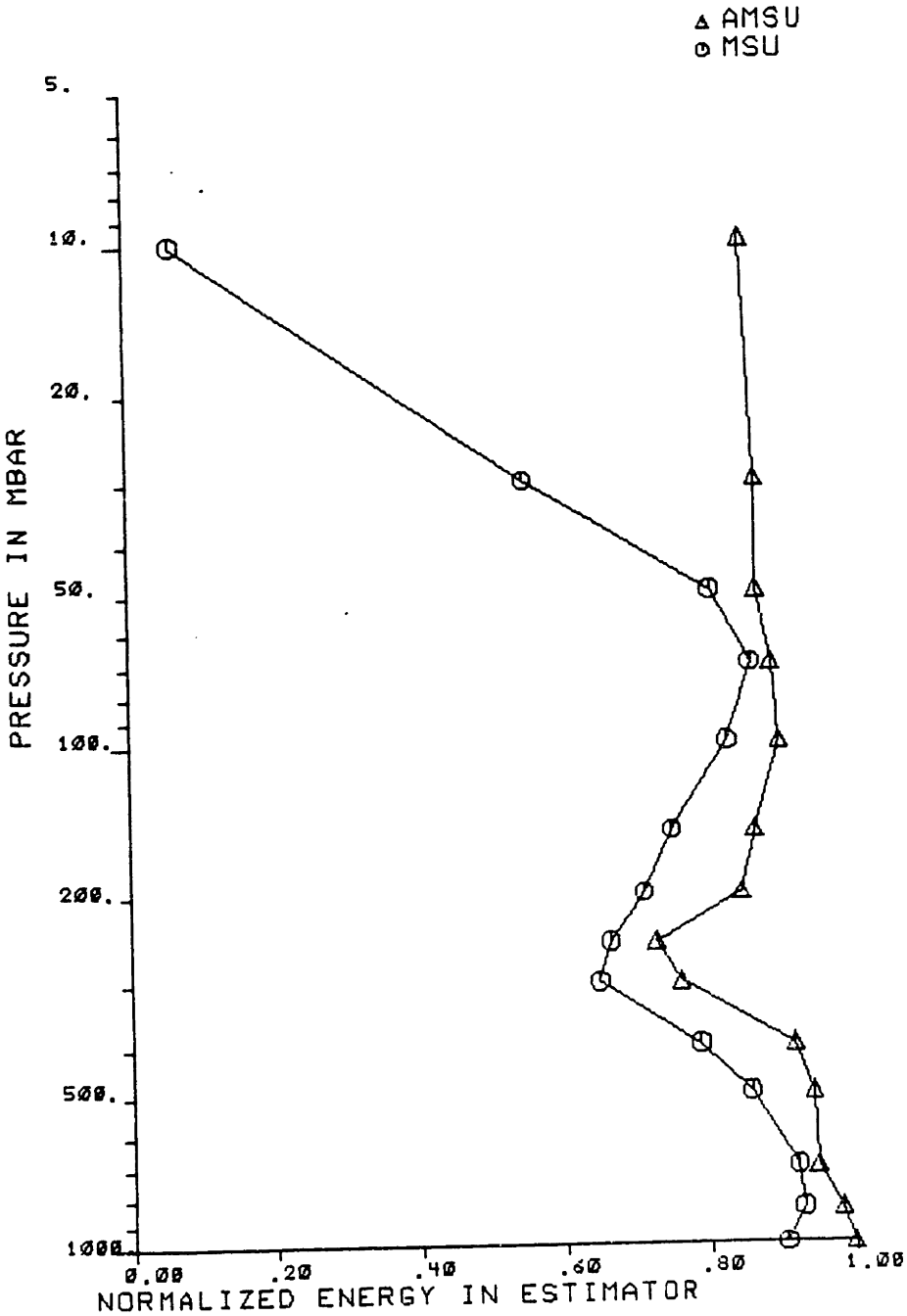


Figure VI.5.5: Nee for MSU and AMSU for winter case. Retrievals are performed at nadir above land

one measurement in each direction provided the same improvement curve (to an accuracy of $0.05K$).

Finally, let us compare rms errors for the MSU and AMSU systems probing

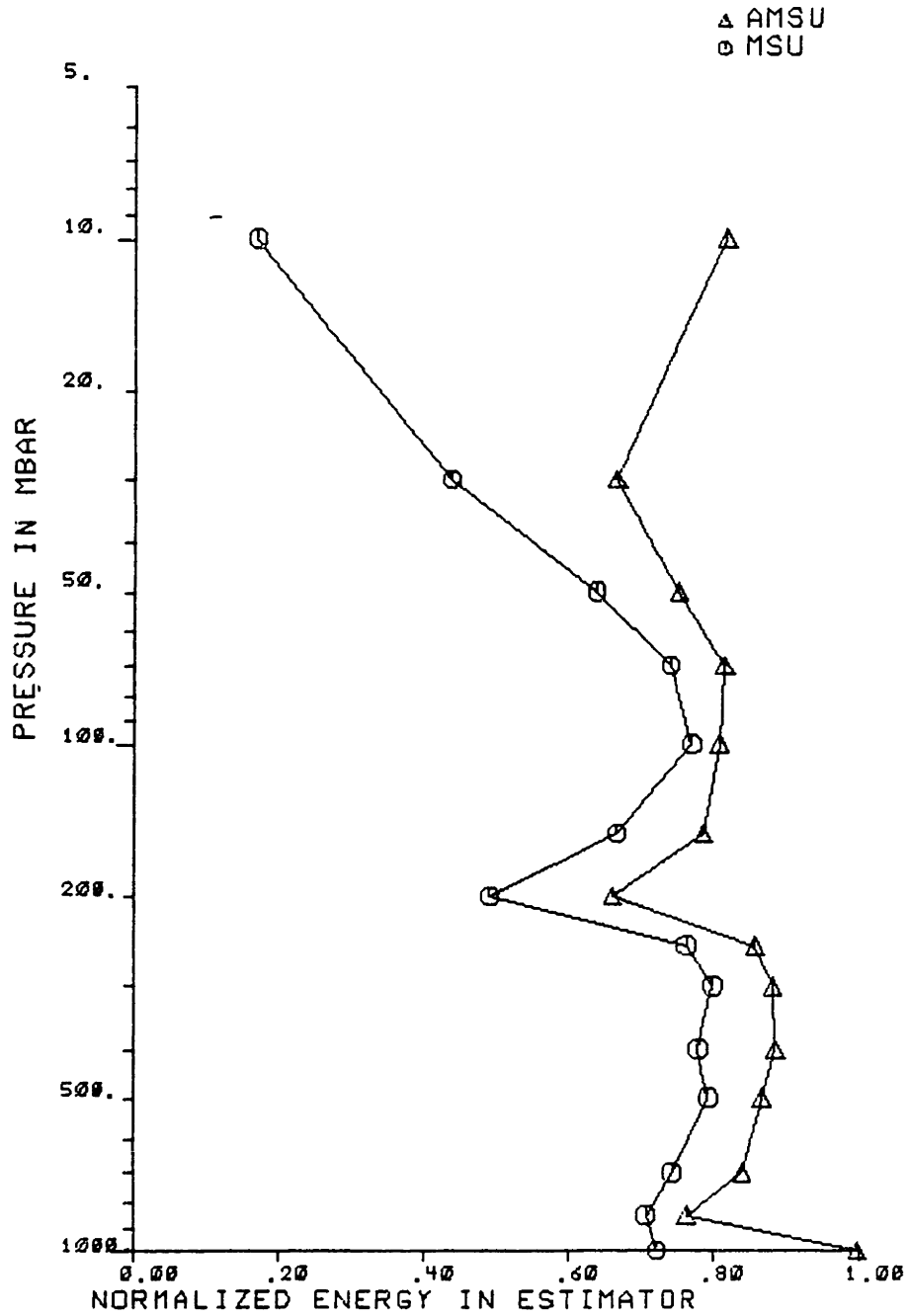


Figure VI.5.6: Nee for MSU and AMSU for summer case. Retrievals are performed at nadir above land

the same datasets. Figures VI.5.3 and VI.5.4 consider the rms errors for single spot operators associated with these two systems at nadir for a winter and a summer case. The improvement provided by AMSU is most important at the ground and at

high altitudes. Note that the ground emissivity is assumed known exactly. In the middle stratosphere, as well in the lower stratosphere, the improvement is less than expected. Figures VI.5.5. and VI.5.6 compare the normalized energy estimated (the nee) for both cases. AMSU estimates typically 75 % of the signal energy for the summer case and 80 % for the winter case. Note that for the winter case, except around the tropopause area (300 to 200 *mbar*), the nee varies between 90 and 100 %, which indicates the AMSU performs very well from a mean-square point of view even if the rms error plots do not convey such a impression. For both winter and summer, the nee is lowest at the tropopause. Note also the drastic change in the nee between MSU and AMSU for pressure less than 50 *mbar*.

The drastic improvement close to the ground is the result of the lowest probing channel of AMSU. Note even in the summer case where the vertical correlation is not that strong close to the ground, the rms error is greatly reduced because of the good probing of lower levels of the atmosphere.

The lack of significant improvement around the tropopause (around 200 *mbar*) demonstrates that the added channels are still not able to resolve that sharp discontinuity. It appears that the error at the tropopause will be the last major error to be eliminated in future retrieval systems.

VI.6 Conclusions

In this chapter we computed and implemented the retrieval operators on MSU data and compared the resulting errors with theoretically computed errors. Both analyses showed that the improvement in the retrievals concentrated in the lower troposphere (up to 20 %) and in the lower stratosphere (up to 15 %). This improvement was found to be larger when the spots used were arranged in the direction perpendicular to the satellite rather than in the parallel one. This difference came from the lifting of the weighting function with view angle which provides different

observable and unobservable subspaces at each location. Most of the improvement was obtained going from one- to two-dimensional retrievals. For implementation on real data, a good fraction of this improvement is provided by a reduction in the bias error. Implementation on real data proved that the use of multidimensional operators is not always profitable when the statistics used to derive the filters is different from the sample statistics, thus the need to correct or adapt for it. This adaptation as well as the quantification of the impact of inadequate statistics, is the subject of the following Chapter. The need for reliable ground truth was also demonstrated.

For AMSU, the improvement from three-dimensional retrieval was found to be much smaller (up to 8 %) and was mostly the result of low-pass filtering in the observable subspace of the instrument. For AMSU, the inclusions of few nearby points in the retrieval process was found to provide most of the improvement.

Appendix VI.A: Isobaric Correlation, Measurements and Number of Degrees of Freedoms.

The purpose of this appendix is to quantify the number of degrees of freedom in a specific retrieval experiment (independently of the form of the retrieval operator and of its dimensions) as a function of the isobaric structure of temperature fields and the total number of frames where retrievals are performed. This analysis is related to the definition of a scale of fluctuation (VanMarcke, 1984) for discrete stochastic processes.

Let $u[n]$ denote a first-order Markov process with variance $\tilde{\sigma}_o^2$ and one lag correlation coefficient ρ. Although the sign of ρ is irrelevant to this analysis, the interpretation of the result is relevant only to positive correlation. Let $\overline{u}_N$ denote the N points sample mean of this process

$$\overline{u}_N = \frac{1}{N}\sum_{n=1}^{N} u[n] \qquad (VI.A.1)$$

Let $\tilde{\sigma}_{u_N}^2$ denote the variance of the sample mean $\overline{u}_N$.

The purpose of this analysis is to compute $\tilde{\sigma}_{u_n}^2$ as a function of N and ρ. The correlation coefficient will be related to the model for the three-dimensional structure in a latter part of the analysis.

$$\begin{aligned}\tilde{\sigma}_{u_N} &= \overline{\frac{1}{N}\sum_{n=1}^{N} u[n]\frac{1}{N}\sum_{n=1}^{N} u[n]} \\ &= \frac{1}{N^2}\sum_{n=1}^{N}\sum_{m=1}^{N}\overline{u[n]u[m]}\end{aligned} \qquad (VI.A.2)$$

where the overline denotes the expectation. Plugging the Markovian assumption

into (VI.A.2) yields

$$N^2\tilde{\sigma}_{u_N} = \sum_{n=1}^{N}\sum_{m=1}^{N}\tilde{\sigma}_o^2\rho^{|n-m|}$$
$$= \tilde{\sigma}_o^2\sum_{n=1}^{N}\Big(\sum_{m=1}^{n}\rho^{n-m} + \sum_{m=n+1}^{N}\rho^{m-n}\Big) \qquad (VI.A.3)$$
$$= \tilde{\sigma}_o^2\sum_{n=1}^{N}\Big(\frac{1-\rho^n}{1-\rho} + \rho\frac{1-\rho^{N-n}}{1-\rho}\Big)$$

for $\rho \neq \pm 1$ After regrouping the terms, this yields

$$N^2\tilde{\sigma}_{u_N} = \tilde{\sigma}_o^2\sum_{n=1}^{N}\frac{1+\rho}{1-\rho} - \frac{\tilde{\sigma}_c^2}{1-\rho}\sum_{n=1}^{N}(\rho^n + \rho^{1+N-n})$$
$$= \tilde{\sigma}_o^2 N\frac{1+\rho}{1-\rho} - \frac{\tilde{\sigma}_o^2}{1-\rho}\Big(\rho\frac{1-\rho^N}{1-\rho} + \rho^N\frac{1-\rho^{-N}}{1-\rho^{-1}}\Big) \qquad (VI.A.4)$$

and

$$\tilde{\sigma}_{u_N}^2 = \frac{\tilde{\sigma}_o^2}{N}\frac{1+\rho}{1-\rho} - \frac{2\tilde{\sigma}_o^2}{N^2}\frac{\rho(1-\rho^N)}{1-\rho} \qquad (VI.A.5)$$

For N large, the standard deviation of the sample mean is given by

$$\tilde{\sigma}_{u_N} = \frac{\tilde{\sigma}_o}{\sqrt{N}}\sqrt{\frac{1+\rho}{1-\rho}} \qquad (VI.A.6)$$

The equivalent number of independent measurements N_e is defined as the number of measurements it would take to have the same standard deviation of the standard mean, namely

$$\tilde{\sigma}_{u_N} = \frac{\tilde{\sigma}_o}{\sqrt{N_e}} \qquad (VI.A.7)$$

which yields

$$N_e = N\frac{1-\rho}{1+\rho} \qquad (VI.A.8)$$

Let us also define a discrete scale of fluctuation θ as

$$\theta = \lim_{N\to\infty} N\frac{\tilde{\sigma}_{u_N}^2}{\tilde{\sigma}_o^2} \qquad (VI.A.9)$$

The scale of fluctuation measures the "width" of an independent measurement. For the first order Markovian process, it equates

$$\theta = \sqrt{\frac{1+\rho}{1-\rho}} \qquad (VI.A.10)$$

The first order Markovian approximation is not, as demonstrated in Chapter III, a adequate model for the isobaric description for the covariance of atmospheric temperature fields. It provides a good approximation only if the oscillatory term is smaller than the decay constant (equation (III.5.)). Furthermore, the temperature fields being estimated cannot been modeled as scalars, but as vectors (for two-dimensional retrievals along the track of the satellite) and more generally as matrices (using the concept of frame).

However, this Markovian analysis still permits us to obtain a rough estimate of the number of degrees of freedom associated with a typical retrieval experiment. Let us then approximate the one step correlation along the track of the satellite by the one derived from the model. That is

$$\rho \approx \exp(-\alpha d) \qquad (VI.A.11)$$

where d denotes the distance between two sampling points along the track of the satellite (in the case of the Microwave Sounding Unit , d equates 0.168 Mm). This approximation yields

$$N_e = N\frac{1-\exp(-\alpha d)}{1+\exp(-\alpha d)} = N sinh(\frac{\alpha}{2}d) \qquad (VI.A.12)$$

In order to get an order of magnitude, let us consider the decay constant fitted to the 500 *mbar* level over the United States ($\alpha = 1.61 Mm^{-1}$). For MSU, 250 independent measurements require about 1400 consecutive retrievals.

CHAPTER VII

ADAPTATION OF RETRIEVAL OPERATORS

"Bring me to the test,
And I the matter will re-word; which madness
would gambol from."
William Shakespeare, Hamlet.

VII.1 Introduction

As described in chapter III, the atmospheric flow and the temperature fields vary with location and time with different dynamics at different scales (Lorenz, 1969). Because of these trends, the temperature field cannot be modeled as a stationary and/or ergodic stochastic process. At best, it can be characterized as locally (piece wise) homogeneous (Charney, 1971). This non-stationarity implies that the first and the second moments of the pdf describing the temperature field will vary between sample sets. These variations have a major impact on the retrieval process. The statistics used to compute a retrieval operator, in general, differ from the sample statistics of the retrieval dataset. Thus, the retrieval error is not be minimum. This effect has been already encountered and quantified in previous works (and this dissertation) where retrieval algorithms did not perform as well as expected because the training set used to derived the filter was different from the sample set (Nathan, 1983, Toldalagi, 1980).

The second impact of these trends in the temperature fields is related to the system identification aspect of the modeling of non-stationarity or non ergodicity of random process, where the separation of the first moment ("deterministic part of the signal") from its second moment ("stochastic part") is non trivial.

The third and most important impact of non-stationarity is, from a engineering point of view, the need to adapt retrieval operators to changes in statistics.

This adaptation, common in several aspects of signal processing (such as speech, image, and video processing), is complicated in the context of remote sensing from satellite in that the observed signal and the signal of interest are vectorial and the observation is incomplete. There are, as discussed extensively in chapter V, observable and unobservable subspaces associated with each observation matrix. Given our previous experience, one might suspect that the adaptation in either of these subspaces must be different. However, in contrast to these other fields mentioned, the satellite is only part of the total observation network for atmospheric temperature. Other sources of information about profiles (numerically predicted temperatures, radiosondes, rocketsondes) and ways of predicting their evolutions (NWP, fluid dynamics and thermodynamics) are available and usable. As noted in chapter IV, the measurements used to fit the model for the three-dimensional covariance kernel are taken using all but satellite data. By restricting the temporal and spatial extent of the measurements used, one can adapt the covariance kernel and the multidimensional operators. However, this adaptation is "off-line" and not practical from an operational point of view. This chapter will concentrate on "on-line" adaptation schemes relying on statistics derived from traditional ways to measure temperatures as well as satellite measurements.

The chapter's organization is as follows: (1) the impact of non stationarities and wrong statistics from two empirical points of view is examined, (2) the traditional models for modeling and characterization of non-stationary scalar processes are quickly presented and analyzed in the context of remote sensing, and (3) a measure more appropriate to the modeling of vectorial non-stationary processes is introduced. In a latter part of this chapter, (4) the optimum criterion for adaptation is developed and examined, a analytical suboptimal criterion is then introduced to allow an analytical treatment of the adaptation problem, (5)the optimum adaptation operator for a Fisher and Bayesian frameworks are introduced, then implemented on MSU and simulated for AMSU.

VII.2 Non Stationarities in the Temperature Fields: Impact and Modeling.

In this section, the characterization of non-stationary processes is considered from a modeling point of view. The importance of the stationarity is briefly discussed using two simple examples which yield simple analytical analyses.

The remainder of this section deals with possible modifications (parameterization) to the model for three-dimensional covariance kernel are considered. Alterations in the model parameters (isobaric complex poles) are based on an analytical description of the statistics and are thus best suited for variations of the horizontal correlation. As discussed in Chapter III, there are no analytical expressions for the vertical covariance kernel which is, in views of the discussion in chapter VI, the most important element in the determination of the retrieval operator rms error. Vertical covariance matrices are typically evaluated by taking a finite sum of outer products of temperature profiles. Depending on the number of profiles used, on the fact that they do belong or not to the same climatology, the vertical covariance kernel will vary dramatically. Although it is possible to change each coefficient of the vertical covariance kernel once the compilation has been performed (preserving the positive definition however), and thus change the effective climatology of the dataset, it is difficult to fully characterize from an intuitive point of view the impact of such a tuning. The later part of this section will deal with the characterization of vertical covariance kernel, more precisely by coming with a quantitative measure of how specific a given kernel is.

VII.2.1 Two Empirical Demonstrations

The first demonstration of the impact of wrong statistics has been illustrated extensively in the previous two chapters. Let $K_{TT}(\sigma, p, p')$ denote the true covariance kernel of a specific temperature field and let $\tilde{K}_{TT}(\sigma, p, p')$ be a specific

approximation of the form

$$\tilde{K}_{TT}(\sigma, p, p') = K_{TT}(\sigma, p, p')\delta(\sigma) \neq K_{TT}(\sigma, p, p') \qquad (VII.2.1)$$

The approximation to the covariance field is uncorrelated in the isobaric direction. All multidimensional retrieval operators derived using the approximate model are single-spot operators. In this case, the impact of badly specifying the isobaric covariance results in the loss of rms error reduction otherwise possible using multi-spot operators. This amount was quantified by chapter VI and found to equal up to 15 to 20 percent of the single spot rms error for MSU and up to 5 to 10 percent for AMSU.

The next example deals with our ability to distinguish trends or modes in the retrieval dataset. For instance, it is possible to cut a specific dataset into several smaller subregions and use different retrieval operators for each one of them. Consider then a set of realizations of temperature profiles which has two modes H_0 and H_1. The definition of mode is left vague intentionally. The mode H_0 is characterized by a covariance matrix R_0 whereas H_1 is characterized by R_1. Both modes share the same mean profile $\overline{m}$ assumed known. Half of the set is in mode H_0 and half is in mode H_1. The overall sample covariance matrix is naturally

$$R = \frac{1}{2}R_0 + \frac{1}{2}R_1 \qquad (VII.2.2)$$

To simplify the analysis, assume that the sounding system which observes this set of temperature profiles has enough channels, and that the covariance matrices R_0 and R_1 are such that the profiles lay in the observable subspace of the instrument. The retrieval error is hence the standard observable retrieval error when the sensor noise covariance is N.

When detecting the presence of two regimes, the retrieval error over the entire set is

$$E^{(2)} = \frac{1}{2}trace\{R_0(R_0+N)^{-1}N\} + \frac{1}{2}trace\{R_1(R_1+N)^{-1}N\} \qquad (VII.2.3)$$

If one fails to use two different retrieval operators, the retrieval mean square error becomes

$$E^{(1)} = trace\{(\frac{R_0 + R_1}{2})(\frac{R_0 + R_1}{2} + N)^{-1}N\} \qquad (VII.2.4)$$

The penalty p for such a misjudgment is defined by

$$\begin{aligned} p =& E^{(1)} - E^{(2)} \\ =& trace\{R_0\big((R_0 + R_1 + 2N)^{-1} - (2R_0 + 2N)^{-1}\big)N\} \\ &+ trace\{R_1\big((R_0 + R_1 + 2N)^{-1} - (2R_1 + 2N)^{-1}\big)N\} \end{aligned} \qquad (VII.2.5)$$

This computation is considerably simplified if the covariance matrices R_0 and R_1 commute (ie., share the same set of eigenvectors). Under such circumstances,

$$p = trace\{\frac{1}{2}(R_0 - R_1)^2 N(R_0 + N)^{-1}(R_1 + N)^{-1}(R_0 + R_1 + 2N)^{-1}N\} \quad (VII.2.6)$$

This penalty vanishes whenever the signal to noise ratio deteriorates and the determination matrices tend to zero. The penalty varies proportionally to the Euclidian distance between the covariance matrices of the two modes.

Note that both retrieval methods (with one or two retrieval operators) are optimum in a mean square sense at their respective resolutions. If the user is capable of distinguishing between the two modes and adapt the retrieval operators, the overall rms error decreases. Such detection may not be easy to obtain in cases where the two regimes are characterized by almost equal statistics. However in such cases, the resulting penalty for not resolving the modes is small.

The above characterization can readily be extended to two modes characterized by different a-priori mean profiles and sharing the same covariance kernel. The same type of results obtained for the second moment extent to the first moment, i.e. the penalty for not recognizing the two modes varies proportionally to the Euclidian distance between the two mean vectors. A notable difference with the previous case is that it does not require all profiles to lie in the observable subspace

of the instrument. This result will be put in perspective in the "mean replacement" schemes of section VII.5.1.

Section VII.5 will consider a more traditional quantification of the impact of inaccurate statistics by implementing single-spot retrieval operators derived from different climatologies on the same dataset. The impact of inaccurate statistics is most important across the tropopause and at pressure levels where no weighting functions peak.

Although the implementation of retrieval operators is performed in a multi-spot framework, one may ask if the adaptation should be performed in such a framework. Multidimensional adaptation (vertical and horizontal adaptation) may be better than single spot (vertical) adaptation but it requires a more extensive processing of data which may not be worthwhile in terms of possible improvements.

In Chapters V and VI, the maximum improvement going from single-spot to multi-spot operators was found to range between 10 to 20 % depending upon the pressure level and climatology. The maximum errors for both single- and multi-spot retrieval operators occur at the ground and at the tropopause where sharp discontinuities in the lapse rate are difficult to locate and represent. Because of the structure of the three-dimensional covariance kernel (itself a reflection of atmospheric flow), whenever the tropopause height is mispecifyed by the vertical covariance kernel, it is mispecifyed by the three-dimensional covariance kernel. This indicates that the penalty for using a covariance derived from a different climatology on a specific dataset will not be lessened by using a multidimensional retrieval operator. As a matter of fact, as illustrated by Table VI.4.4, the multidimensional operator may perform worse than the associated one-dimensional operator. These results show that the vertical covariance kernel is essential to the good performances of multidimensional operator, a results confirmed by the theoretical analysis of section V.3.

Furthermore, as characterized in Chapter IV, the isobaric correlation length is shortest around the tropopause and the surface which results in a reduction of the ability of multidimensional filters to improve the retrieval error around the tropopause, even when the tropopause height is correctly inferred.

Moreover, as the number of channels increases, so does the proportion of signal's energy in the observable subspace of the instrument. For such cases, multidimensional retrieval operator are essentially single-spot operators, or if they are not, their performances is essentially characterized by the single-spot retrieval error. Hence, their performances solely rely on the vertical (pressure) covariance matrix.

Based on these items, one may now establish that, in terms of adaptation, the one dimensional covariance kernel has more importance than the full three-dimensional covariance kernel. Most of the attention of this Chapter will be devoted to the vertical covariance kernel. However, all results can be directly applied to the three-dimensional covariance by grouping measurements taken at different frame locations at one large brightness temperature vector and extending the definitions developed below to that vector.

VII.2.2 Modeling and Characterization of non Stationarity and Informational Content

• *Modeling and Characterization of non-stationarity in the horizontal domain*

The easiest method to modify the statistics characterizing a scalar process to model non-stationarities is to change the parameters describing the spectrum, covariance or recursion formula describing the signal (the later description is associated with ARMA modeling). These coefficients' variations can be piecewise constant (Toldalagi, 1980) or vary smoothly along the path of the satellite (time series index) (Robinson and Treitel, 1978). This later technique is efficient in processing time-series that presents trends or which are pseudo-periodic, such as seismic

traces (Weggi and Dimri, 1984) or speech (Jayant and Noll, 1984).

The only mathematical constraint on the variations of the isobaric poles is to guarantee a valid model for the three-dimensional covariance. The positive definiteness of the three-dimensional model will be guaranteed if the adapted poles obeys the role that real part be larger in magnitude than the imaginary part.

An interesting approach to modeling non-stationary time series consists of the introduction of so-called "hyper-parameters" which allows to write an ARMA type regression for the ARMA parameters of the time-series (Kitagawa and Gersh, 1985a, 1985b). The hyper-parameters are supposed constant for each time series, and the signal ARMA model are estimated using localized windows and traditional system identification techniques, before any processing of the series is performed. This approach is appealing in the context of remote-sensing because the isobaric poles of the model (equivalent to the ARMA coefficients) are representing the characteristics of the two-dimensional flow. These two-dimensional flow characteristics could be obtained from numerical weather prediction (NWP) sources, and this information could then be incorporated in the design of multi-spot retrieval operators. Given the difficulties encountered by Toldalagi (1980) in using NWP-based parameters to modify retrieval operators in a "on-line" fashion, such an approach was not considered in this thesis.

The identification of isobaric poles does not require, however, a parametric description of those parameters. Assuming that the variations of this parameters are smooth, one can window the brightness temperatures and identify the poles by extending the method developed for scalar processing (Kumaresan and Tufts, 1982). Such methods would extend directly to the observable subspace of the instrument, but there is no clear method to extend them to the unobservable subspace where most of the improvement of multidimensional operators over single spot operators comes from. The only clear way to extend these identification schemes is to assume

the isobaric poles to be the same in the observable and the unobservable subspace. The impact of such an assumption would be to reduce the impact of horizontal correlation for retrieval operators based on measurement performed under similar or near similar conditions (see chapter V). Finally, the horizontal correlation of brightness temperatures depends on the vertical covariance kernel for the physical temperature. Hence, any error in the vertical covariance kernel (height of the tropopause, lapse rate close to the ground) will impair one's ability to identify these isobaric poles using satellite data. For these reasons and because of time and flood constraints, the horizontal adaptation was not considered.

Note, however, that adapting the vertical covariance kernel will alter the form of multidimensional retrieval operators in both vertical and horizontal direction by modifying the impact of certain pressure levels and their corresponding isobaric poles.

• *Modeling and Characterization of non-stationarity in the horizontal domain*

As described in Chapters III and IV, because of non-stationarities in the vertical direction, the vertical covariance matrix is not parameterized and thus cannot be altered except by adding or subtracting outer products of temperature profiles. A parametric model for the covariance such as Bergmann (1979) (check section III.3.2) can be changed by tuning the correlation height K_p in equation III.3.9. However this model is not appropriate because it does not consider the sharp inversion of the tropopause.

Even if the modeling of that matrix is difficult (or impossible), the characterization of the vertical covariance kernel is still possible. The main idea behind this characterization is to quantify if a given covariance matrix is obtained by taking few temperature profiles and thus specific of a small air mass, or if it is compiled using numerous and unrelated profiles.

In Chapter VIII, a purely information theoretic analysis of the remote sensing process will be advanced which analyzes the radiative transfer equation and the moment characterization of the retrieval process using the concepts of entropy and mutual information. However, one can try to use information theory concepts to analyze the informational content of vertical covariance kernels. The overall goal of this analysis is to quantify how specific a given vertical covariance kernel is.

The main thrust of this analysis is the distribution of eigenvalues of the vertical covariance matrix. The proposed technique is an extension of the factor analysis technique introduced by Watanabe (1965) for speech processing. Denote by $\lambda_i, i = 1, \ldots, N$ the N eigenvalues of a vertical covariance kernel R computed by taking an average of the outer products. The normalized eigenvalues γ_i are defined as (Unser, 1984)

$$\gamma_i = \frac{\lambda_i}{\sum_j \lambda_j}, i = 1, \ldots, N \qquad (VII.2.7)$$

Each normalized eigenvalue is a number between zero and one and their sum equates one, thus similar to a discrete probability density whose entropy (Shannon,1949) can be computed according to

$$H = -\sum_{i=1}^{N} \gamma_i \log \gamma_i \qquad (VII.2.8)$$

This measure of uncertainty does not have any connection with the entropy of a random process with covariance matrix R. This is true, because any reversible operation on the covariance matrix does not alter the entropy of the random variable but greatly modifies the eigenvalues of R. This definition is mostly an extension of the entropy measure used in spectral estimation (Jaynes, 1982) which measures the "non-whiteness" of the spectrum. The proposed entropy measure becomes the entropy defined in terms of the spectrum , should the temperature statistics be stationary in the vertical direction and the number of pressure levels large.

Two limiting cases of vertical covariance kernels can be thought of as providing bounds on the entropy measure H. The first bound is the case of minimum information (from an estimation point of view) where all eigenvalues are equal and infinite. This distribution corresponds to temperature profiles uncorrelated from pressure level to pressure level with infinite variances. A diagonal matrix is the limit of the sum of outer products when the temperature profiles are assumed independent of each other and characteristic of numerous unrelated climatologies. As the number of uncorrelated outer products increases, the eigenvalues of the covariance matrix tend towards those diagonal entries. Such as convergence is a $0(N^{-1})$ convergence (Brillinger, 1975). For a constant entry diagonal matrix, all normalized eigenvalues equal $\frac{1}{N}$ and the entropy measure

$$H = -\log \frac{1}{N} = \log N > 0 \qquad (VII.2.9)$$

which represents a white spectrum. This covariance matrix is as unspecific as possible since it corresponds to the combination of an infinite number of climatologies. It yields a minimum information solution where observable and unobservable components are all uncorrelated.

The other limiting bound to the entropy measure is when the covariance kernel is characteristic of a very specific climatology. Such is the case if the compilation is performed using only two temperature profiles, say $\overline{T}_1$ and $\overline{T}_2$. The trivial case of only one profile yield a sample covariance matrix identically null and is pathological. For the two profiles case, the sample covariance matrix is given by

$$R_{\mathbb{B}} = \frac{1}{4}(\overline{T}_1 - \overline{T}_2)(\overline{T}_1^t - \overline{T}_2^t) \qquad (VII.2.10)$$

Note when plugging such a covariance matrix in the design equations for retrieval operators developed in Chapter III, one will obtain a unique optimum retrieval operator if there is sensor noise in order to guarantee the existence of the inverse of $WR_{\mathbb{B}}W^t + N$. Using a retrieval operator derived using such a matrix is similar to

the linear variety constrained estimation to be introduced in Chapter VIII. When the covariance matrix R_{Ξ} has only one eigenvector with a non-zero eigenvalue, the estimated temperature profile, that is the deviation from the a-priori mean, is simply a scaled version of that eigenvector. Hence, the sample covariance matrix corresponds to a very specific distribution (collection) of temperature profiles.

The vertical covariance kernel R_{Ξ} has a unique non-zero eigenvalue (equal to $(\overline{T}_1 - \overline{T}_2)^2/4$). Thus all normalized eigenvalues γ_i are equal to zero, except for the first one equal to one. The entropy measure of the matrix is (taking $0 \log 0$ as the limit of $x \log x$ as x tends towards zero)

$$H = \log 1 = 0 \qquad (VII.2.11)$$

The uncertainty associated with such a matrix is null. The proposed climatology is very specific. This is equivalent to the spectrum of a single sinusoidal wave. For practical reasons, let us normalize the entropy measure by dividing it by its upper bound, namely

$$H_{spec} = -\frac{1}{\log N} \sum_{i=1}^{N} \gamma_i \log \gamma_i \qquad (VII.2.12)$$

Any covariance kernel will have a normalized entropy measure between zero and one (Gallager,1963).

Now that this proposed normalized entropy measure has been introduced, let us prove that it is an appropriate measure of how specific a covariance matrix is. As aforementioned, sample covariance matrices are computed as finite sums of outer products of temperature profiles whose sample mean has been removed. The computation of the sample covariance matrix as the number of profiles used increases can be written as a recursion equation. Let $\overline{m}_N$ denote the sample mean using N samples and R_N the corresponding covariance. The recursion for the mean profile is

$$\overline{m}_{N+1} = \frac{1}{N+1} \sum_{i=1}^{N+1} \overline{T}_i = \frac{N}{N+1} \overline{m}_N + \frac{1}{N+1} \overline{T}_{N+1} \qquad (VII.2.13)$$

Likewise for the covariance matrix

$$R_{N+1} = \frac{1}{N+1}\sum_{i=1}^{N+1} \overline{T}_i \overline{T}_i^t - \overline{m}_{N+1}\overline{m}_{N+1}^t$$
$$= \frac{N}{N+1} R_N + \frac{N}{(N+1)^2}(\overline{T}_{N+1} - \overline{m}_{N+1})(\overline{T}_{N+1} - \overline{m}_{N+1})^t \qquad (VII.2.14)$$

The above equation yields a recursion on the eigenvalues of R_N which, in turn, yields a recursion on the entropy measure. Unfortunately, it is impossible to indicate exactly, even in the limiting case when N is a large number and the perturbation theory of eigenelements can be applied (Wilkinson, 1965), if and when the entropy measure at step $N+1$ is larger (a less specific covariance) or smaller (a more specific covariance) at step $N+1$ than at step N. The increase or decrease in the entropy measure depends on the inner product of the new profile $\Delta\overline{T}_{N+1} = \overline{T}_{N+1} - \overline{m}_{N+1}$ with the different eigenvectors of R_N. If the new profile is more aligned with the eigenvector having the largest eigenvalue at step N, the spread of eigenvalues at step $N+1$ will be larger, the entropy measure will decrease and the new covariance matrix will be more specific. On the other hand, if the new profile is most aligned with any of the other eigenvectors, the entropy measure will increase and the new covariance matrix will be less specific.

Although an exact analysis of the evolution of the entropy measure is impossible, one can examine the likelihood of evolution of this quantity. To simplify the analysis, assume that the temperature profiles follow a Gaussian pdf.

If the new temperature profile belongs the ensemble of profiles (climatology) characterized by R_N, its likelihood $\mathcal{L}$ is

$$\mathcal{L} \sim \exp\left(-\frac{1}{2}\sum_i \frac{< \Delta\overline{T}_{N+1}^t \overline{\theta}_i >}{\lambda_i}\right) \qquad (VII.2.15)$$

where $\overline{\theta}_i$ and λ_i denote the eigenvectors and eigenvalues of R_N respectively. This equation shows that, all things considered, the temperature profile is most likely to

July USA	0.3957
August USA	0.4035
Summer USA	0.4175
July Europe	0.3456
July USA	0.3957
July both	0.3981
August Europe	0.4317
August USA	0.4035
August both	0.4364
July both	0.3981
August both	0.4364
Summer both	0.4168

Table VII.2.1: Normalized entropy measures for several vertical covariance kernels

lay along the eigenvector associated with the largest eigenvalue. Hence, if the new temperature profile belongs to the same climatology, the entropy measure is likely to decrease.

A temperature profile unrelated to the ensemble of N profiles can be characterized by a Gaussian pdf whose eigenvalues are all equal (no preferred orientation). When this is the case, the new temperature profile is most likely to lay along one of the eigenvectors not associated with the largest eigenvalue. This is true because simply there are more than one of them. In this case, the entropy measure is likely to increase, and the overall set of temperatures becomes less specific.

Let us compute this measure for several matrices compiled over different locations and time of the year. Table VII.2.1 presents the value of H_{spec}. The measure behavior follows what could be expected from an intuitive point of view, that is, for a region, the longer the considered time period, the less specific the climatology. Likewise, for a given time period, the larger the region is, the less specific the climatology is. The notable exception is the last case where summer over the USA and Europe is quantified as more specific than August over the same

region.

There is, unfortunately, no direct relationship between specificity and sensitivity of retrieval errors to statistics (this analysis is pursued with more details in section VII.3.4). This is true because the basis vectors for the SVD analysis do not coincide with the eigenvectors of the vertical covariance kernel. However, the specificity measure is a valuable tool to compare statistical description. Finally, note that the measure is an extension to a finite length random process of the predictability index developed in speech and video processing (Jayant and Noll, 1984).

VII.3 Criteria for Adaptation

In this section, the penalty for not choosing the optimal retrieval operator is derived. The minimization of this penalty forms the basis for the design of adaptation operator. A suboptimal optimization criterion more suited for adaptation is then derived and characterized with respect to the best optimization criterion. Finally the adaptation are derived and implemented on MSU and AMSU.

VII.3.1 General Case

For any retrieval experiment, the retrieval error can be written as the difference between the a priori distribution and the error reduction provided by a specific retrieval operator. The error reduction is bounded above by e_{max} the maximum possible reduction obtained for the optimal operator. Expressed with the notation of chapter V (signal covariance R, weighting matrix W, and noise covariance N), this maximum reduction is given by

$$e_{max} = trace\{RW^t(WRW^t + N)^{-1}WR\} \qquad (VII.3.1)$$

Note that e denotes the square error. For operator with all zero-entries (feeding back the a-priori mean as estimate for the temperature profile), the error reduction is null. The retrieval operator as well as the error reduction depend only on

the covariance of the observable components and on the cross-covariance between observable and unobservable components. To analyze the impact of misspecifying these two quantities, two possible approaches are possible: The former fixes the retrieval operator and changes the statistics of the temperature field under analysis (this approach is close to the implementation of retrieval operators), the latter fixes the statistics and varies the retrieval operator.

To reference the misspecification of the statistics to the determination matrix (or retrieval operator), add and subtract the covariance kernel R_a from the sample covariance kernel R_s in the expression for the retrieval error. Namely,

$$e = trace\{(D_aW - I)(R_a + R_s - R_a)(W^tD_a^t - I) + D_aND_a^t\} \qquad (VII.2.2)$$

After factorization, this equation yields

$$e = trace\{D_aW - I)R_a(W^tD_a^t - I) + D_aND_a^t\} + trace\{(\mathcal{T}_a - I)(R_s - R_a)(\mathcal{T}_a^t - I)\} \qquad (VII.2.3)$$

or equivalently

$$e = e_a + trace\{(\mathcal{T}_a - I)(R_s - R_a)(\mathcal{T}_a^t - I)\} \qquad (VII.2.4)$$

where e_a is the optimum retrieval error associated with the statistical description R_a and $\mathcal{T}_a$ the corresponding transfer matrix. The rightmost term in the RHS is the penalty for not using the correct statistics. The main problem with such a referencing of the penalty is the inability to determine if it is positive or negative. Formulated in this framework, deriving an adaptation operator based on the penalty is impossible. For that reason, all of the following work is referenced to the statistical description of temperature fields.

The penalty for wrong statistics is defined as the difference between the retrieval error associated with a dataset and a retrieval operator and the minimum retrieval error that can be obtained on that dataset. This penalty is instrument

dependent. The minimization of this penalty will form the basis for adaptation of retrieval operators. Using the same notation as above,

$$p = trace\{(I - D_a W)R(I - W^t D_a^t) + D_a N D_a^t + (DW - I)R\} \qquad (VII.3.5)$$

To derive a simpler expression for the penalty, express the diverse retrieval operators in the SVD basis using the expression normalized to the temperature profile as developed in Chapter V. Namely,

$$D_a = \begin{pmatrix} R_{aoo}(R_{aoo} + \Sigma^{-1}N\Sigma^{-1})^{-1}\Sigma^{-1} \\ R_{auo}(R_{aoo} + \Sigma^{-1}N\Sigma^{-1})^{-1}\Sigma^{-1} \end{pmatrix} \qquad (VII.3.6)$$

To simplify the computation, denote the normalized noise covariance matrix $\Sigma^{-1}N\Sigma^{-1}$ by N'. Moreover, let R_{ao} denote the vertical concatenation of the covariance matrix of the observable components R_{aoo} and the cross-covariance between unobservable and observable components R_{auo}.

Projecting the matricial error using the retrieval operator derived from the a priori statistics, after expansion, yields

$$\begin{aligned} e =& trace\{R_{ao}(R_{aoo} + N')^{-1}N'(R_{aoo} + N')^{-1}R_{ao}^t\} \\ &+ trace\{R_{ao}(R_{aoo} + N')^{-1}R_{oo}(R_{aoo} + N')^{-1}R_{ao}^t\} \\ &+ trace\{-R_{ao}(R_{aoo} + N')^{-1}R_c^t - R_o(R_{aoo} + N')^{-1}R_{ao}^t\} \end{aligned} \qquad (VII.3.7)$$

where R_o is the concatenation of the true covariance matrix of observable components with the true cross-covariance of unobservable components with the observable ones. With this notation, R_{ao} equals R_o for the optimal retrieval operator. After manipulations,

$$\begin{aligned} e =& trace\{(R_{aoo} + N)^{-1}(R_{oo} + N)(R_{aoo} + N)^{-1}R_{ao}^t R_{ao}\} \\ &- trace\{(R_{aoo} + N)^{-1}(R_{ao}^t R_o + R_o^t R_{ao})\} \end{aligned} \qquad (VII.3.8)$$

Thus, the penalty for wrong statistics is

$$\begin{aligned} p =& trace\{(R_{oo} + N')^{-1}\big((R_{oo} + N')(R_{aoo} + N')^{-1}(R_{oo} + N')(R_{aoo} + N')^{-1}R_{ao}^t R_{ao}\} \\ &+ trace\{-(R_{oo} + N')(R_{aoo} + N')^{-1}(R_{ao}^t R_o + R_o^t R_{ao}) + R_o^t R_o)\} \end{aligned} \qquad (VII.3.9)$$

When the penalty p equals the maximum reduction e_{max}, performing the estimation is equivalent to not performing any estimation. The expression for the penalty can further be reduced to

$$p = trace\{(R_{oo} + N')^{-1}\big((R_{oo} + N')(R_{aoo} + N')^{-1}R_{ao}^t - R_o^t\big) \big((R_{oo} + N')(R_{aoo} + N')^{-1}R_{ao}^t - R_o^t\big)^t\} \quad (VII.3.10)$$

Factoring the true covariance kernel of the brightness temperature $(R_{oo} + N')$ out yields

$$p = trace\{(R_{oo} + N')\big(R_{aoo} + N')^{-1}R_{ao}^t - (R_{oo} + N')^{-1}R_o^t\big) \big(R_{aoo} + N')^{-1}R_{ao}^t - (R_{oo} + N')^{-1}R_o^t\big)^t\} \quad (VII.3.11)$$

In the above equation, one can recognize the expressions of the determination matrices for the temperature profile expressed in the SVD basis using the a priori and the true covariance matrices. Thus, the adaptation criterion, ie. the penalty, to be minimized becomes

$$p = trace\{(R_{oo} + N')(D_a^t - D^t)(D_a - D)\} \quad (VII.3.12)$$

The penalty is proportional to the square difference between the retrieval operator and the optimum one.

The above derivation shows that the mean square design of the determination matrix (or in multidimensional cases of the retrieval operator) is optimal when the determination matrix is later used to optimally estimate (in a mean square sense) the temperature profiles. In other words, to optimally estimate temperature profiles in a mean square sense, one must derive retrieval operator also optimal in a mean square sense.

The major difficulty with such a penalty function is the characterization of the optimal estimator as the sample covariance varies. This is true since the optimal retrieval operator is the product of two matrices, one of them being the inverse of

the sample covariance matrix. Defined as a function of the realized temperature profiles, the optimal determination matrix (set dependent) is

$$D_S = \sum_{i \in S} \overline{T}_i \overline{T}_{oi}^t (\sum_{i \in S} \overline{T}_{oi} \overline{T}_{oi}^t + N)^{-1} \qquad (VII.3.13)$$

where $\overline{T}_i$ denotes the temperature profile whose sample mean has been subtracted and N is the covariance matrix corresponding to the sensor noise. This determination matrix is defined only if enough temperature profiles are compiled in the sample set S. A lower bound of such a number is the number of channels used as inputs in the retrieval process (3 for single spot MSU, 3×11 for the two-dimensional operator perpendicular to the track of MSU). Even for simple cases, such as modeling of the temperature profiles as Gaussian stochastic processes, the computation of D_S and of $D_S D_S^t$ (also required in the optimization) is analytically burdensome (if possible). However, the criterion is a convex function of the determination matrix, thus allowing for optimization by conventional numerical algorithms (Ortega and Rheinboldt,1970).

Because the average behavior of the criterion is not computable, one must sacrifice optimality to derive a suboptimal criterion under which estimation can be performed with greater facility. One method that achieves this task restricts the retrieval operator to a specific form.

VII.3.2 Consistency with Observations

Consistency with observations pertains to the special case where the retrieval operator is derived using a covariance matrix that matches the observed covariance matrix. Consistency with observation is unrelated to the asymptotic behavior of statistical estimators such as discussed in Ibragimov and Has'minskii (1981).

By denoting R the estimated covariance matrix, R_o the sample covariance matrix, and R_{TbTbo} the covariance matrix of the observed brightness temperature,

the statistics is sample consistent (or consistent with the observations) if

$$R_{TbTbo} = WR_oW^t + N = WRW^t + N \qquad (VII.3.14)$$

In these instances, the determination matrix takes the form

$$D = RWR^{-1}_{TbTbo} \qquad (VII.3.15)$$

By forcing consistency of the retrieval operator with the observation, the future adaptation is made optimal only for a specific form of retrieval operators. The benefit of such a restriction is that one can obtain a simple optimization criterion for which optimal estimators can be derived. Unfortunately, the proposed criterion is suboptimal. In other words, because an optimal or suboptimal adaptation operator cannot be derived for the optimal adaptation criterion, one changes this criterion. For the new sub-optimal criterion, the optimal adaptation criterion can be derived, as will be demonstrated momentarily.

Plugging this determination matrix into the error term yields (assuming that the mean temperature profile is perfectly known)

$$e = trace\{(RW^tR^{-1}_{TbTbo}W - I)Ro(W^tR^{-1}_{TbTbo}WR - I) + RWR^{-1}_{TbTbo}NR^{-1}_{TbTbo}W^tR\} \qquad (VII.3.16)$$

After manipulations,

$$e = trace\{RW^tR^{-1}_{TbTbo}WR + R_o - R_oW^tR^{-1}_{TbTbo}WR - RW^tR^{-1}_{TbTbo}WR\} \quad (VII.317)$$

Now, introduce the symmetric "Impact" matrix $H = W^t(WR_oW^t + N)^{-1}W$. This impact matrix has the same rank as the weighting matrix W and thus has a non empty null space. The latter coincides with the unobservable space of the weighting matrix W. The retrieval error can be expressed as

$$e = trace\{RHR + R_o - R_oHR - RHR_o\} \qquad (VII.3.18)$$

Introducing the difference ΔR between the sample covariance matrix R_o and the estimated covariance matrix R,

$$e = trace\{\Delta R H \Delta R + R_o - R_o H R_o\} \qquad (VII.3.19)$$

Note that all components of R and R_o do not play the same role, hence the name "Impact matrix" to denote those components that matter in the computation of the penalty. The product of the sample covariance matrix and the impact matrix is the transmission matrix of the optimal retrieval problem. Thus, the last two terms in the trace operator corresponds to the retrieval error when the statistical description is optimal.

The penalty for using the wrong covariance matrix in the computation of the determination matrix becomes

$$e = trace\{\Delta R H \Delta R\} = trace\{\Delta R W^t (W R_o^t W + N)^{-1} W \Delta R\} \qquad (VII.3.20)$$

where $\Delta R = R_o - R$. This is a convex relationship of the difference in covariance matrices, which is well suited for iterative search algorithms.

In the case of retrieval operators consistent with observations, the mean square difference between the estimated covariance matrix and the sample covariance matrix determines the penalty for not using the sample statistics. Hence, to optimally estimate, in a mean square sense, the temperature profiles, one must optimally estimate, in a mean square sense, the covariance matrix associated with those profiles.

For more insight into the impact of wrong statistics and the implications of observations consistent on the retrieval process, one can analyze graphically a simple case. Consider a two-element vector, where only the first component is observed according to

$$z = \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \eta \qquad (VII.3.21)$$

The observable subspace of this "sounding" is the x component whereas the unobservable component is the y component.

One can compute the penalty for using the wrong noise covariance matrix, the wrong variance of the x component, and the wrong cross-covariance between x and y. Three different cases are considered to analyze the impact of the signal-to-noise ratio. In all three cases, the variance of the x-component K_{xx} as well as the cross-covariance K_{xy} between x and y are equal to one. In the first case, the true covariance of the noise is equal to one $Kelvin^2$ (average SNR), in the second case it equals ten $Kelvin^2$ (bad SNR), and in the third and final case one tenth of $Kelvin^2$ (good SNR).

To scale the penalty, it is normalized by dividing it by the maximum amount of improvement of the retrieval error over the a priori distribution. That is, a normalized penalty of one corresponds to a case where the retrieval error is, due to inadequate statistics, equal to the a priori distribution. When this is the case, performing the retrieval with the wrong statistics and giving for estimate the a priori mean are equivalent in terms of rms error and for all purposes worthless. A normalized penalty of zero corresponds to the optimum retrieval operator.

Figure VII.3.1 presents such a normalized penalty for the case of good signal-to-noise ratio for the observable component. The form of the isopenalty contours does not vary when the SNR varies, they are all straight lines function only of the SNR. The straight line superimposed on the graph corresponds to the observed power, that is the power of the variable z. This quantity which is the sum of the signal variance and the noise variance is directly computable from the observations. The quantity z is the equivalent of the brightness temperature's brightness temperature for multidimensional remote sensing. For any dataset, the covariance of the brightness temperature, the sum of the covariance of the observable components (after scaling) and of the sensor noise can readily be computed/ When

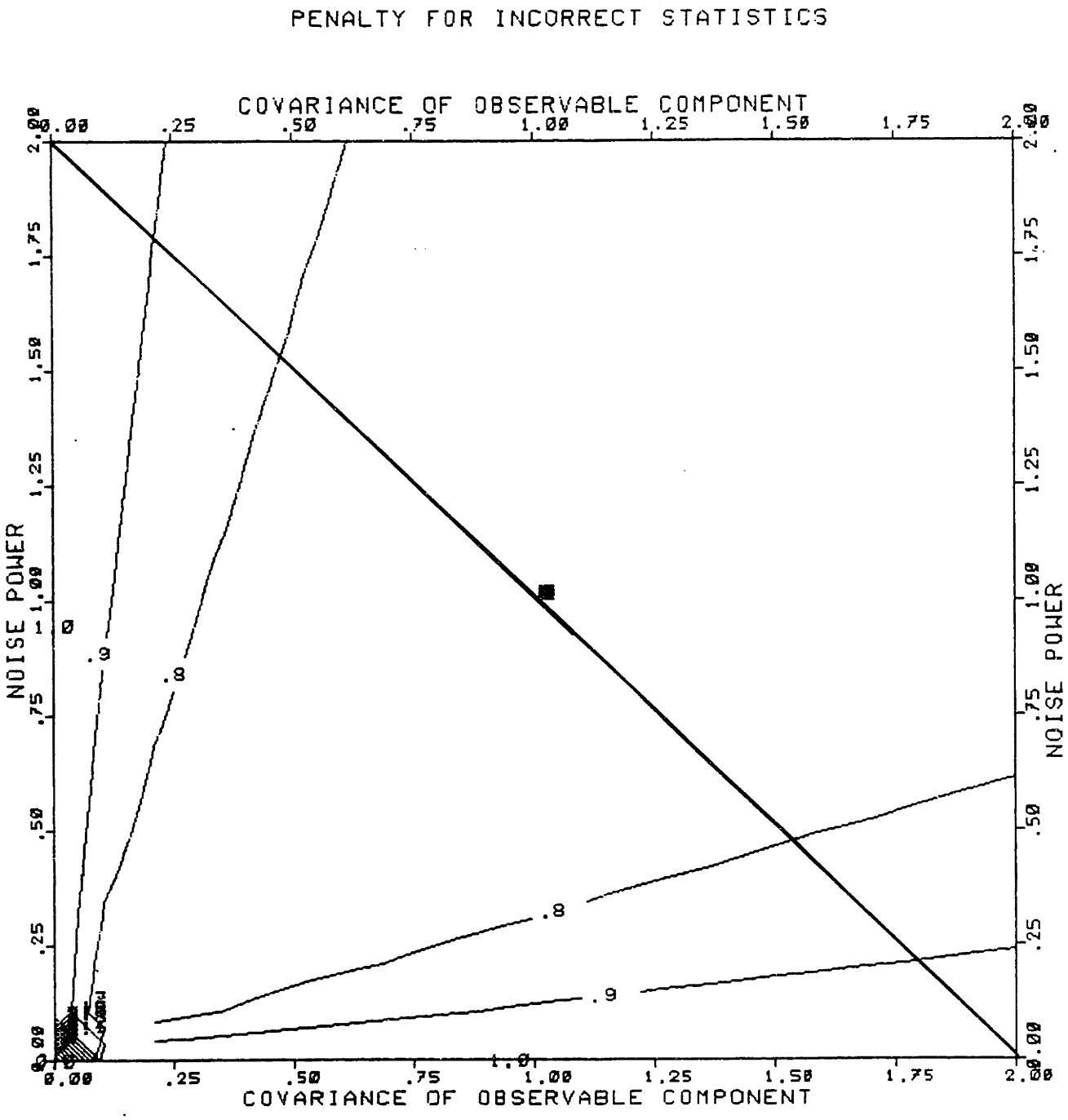


Figure VII.3.1: Normalized penalty for the observable component as a function of assumed receiver noise and signal covariance matrices. $K_{xx} = K_{\eta\eta} = 1$.

performing any type of adaptation for the observable components, one lays on that straight line: The position along the line depends on the value of the receiver noise covariance. Because such a line is not perpendicular to the isopenalty lines, a simple relationship does not exist for "sample consistent" retrieval operators. That is to say, there is simple analytical relationship between the penalty for misspecifying the experiment's SNR and the abcisse along the line.

Figures VII.3.2, VII.3.3 and VII.3.4 present the same penalty function for the unobservable components. The impact of the signal-to-noise ratio is apparent by looking at the slopes of the isopenalty curves. For cases where the SNR is small, the isopenalty curves are almost parallel to the signal covariance matrix whereas for cases with large SNR's, the angle of the isopenalty curves vary with the estimated noise covariance to become almost parallel to the cross-covariance axis for small a priori distribution of the observable components. This case corresponds to a priori statistics where the power in the unobservable components is much larger than the power in the observable components and where the correlation between those components is larger, which translates to a cross-covariance much larger than the variance of the power in the observable components.

For both cases, it should be noted that on the vertical axis $R_{a-priori} = R_{true}$, the penalty follows a parabolic behavior. This shows the quadratic relationship between error between the true and the a priori covariance matrices and the normalized penalty unobservable component when the retrieval operator is assumed to be consistent with the observations.

In cases where the signal-to-noise ratio is large, the square difference between the covariance matrices is almost equivalent to the optimum criterion for the adaptation which is the square difference between the retrieval operators.

The signal-to-noise ratio measures, in this context, more than just the performances of the sounding unit (antenna gain, front end amplifier, mixer and other

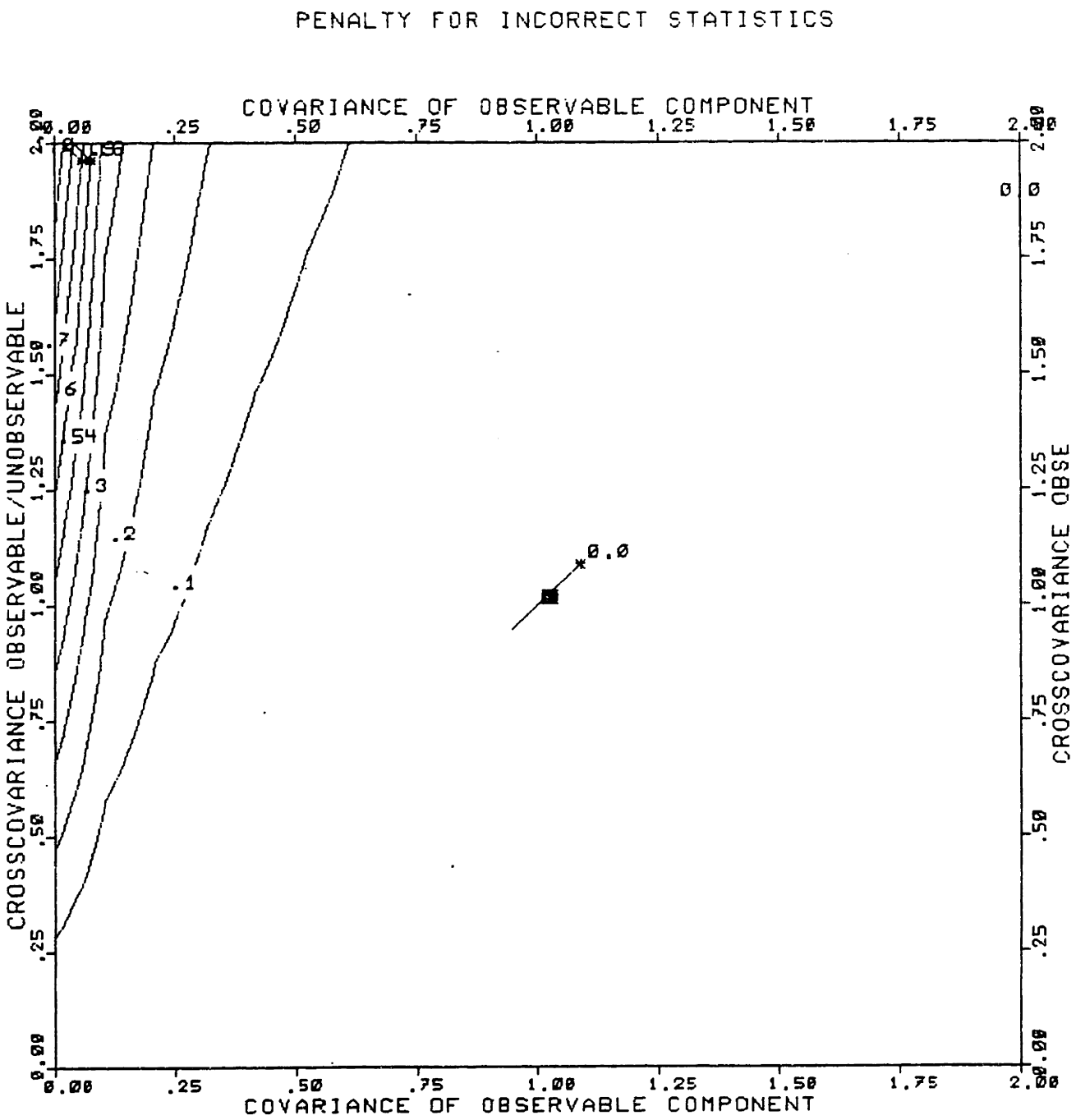


Figure VII.3.2: Normalized penalty for the y components in the case of good signal-to-noise ratio.

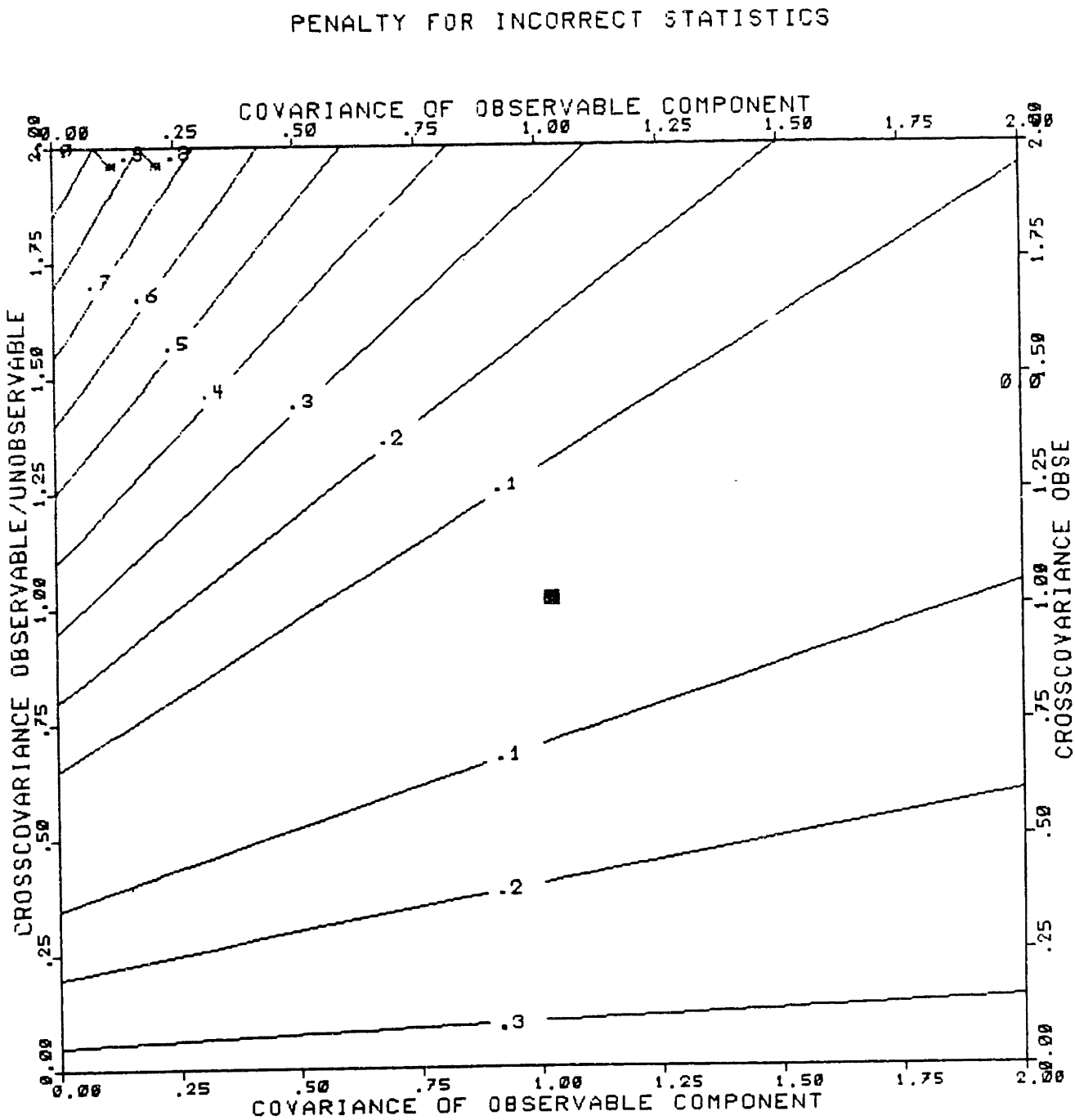


Figure VII.3.3: Normalized penalty for the y components in the case of average signal-to-noise ratio.

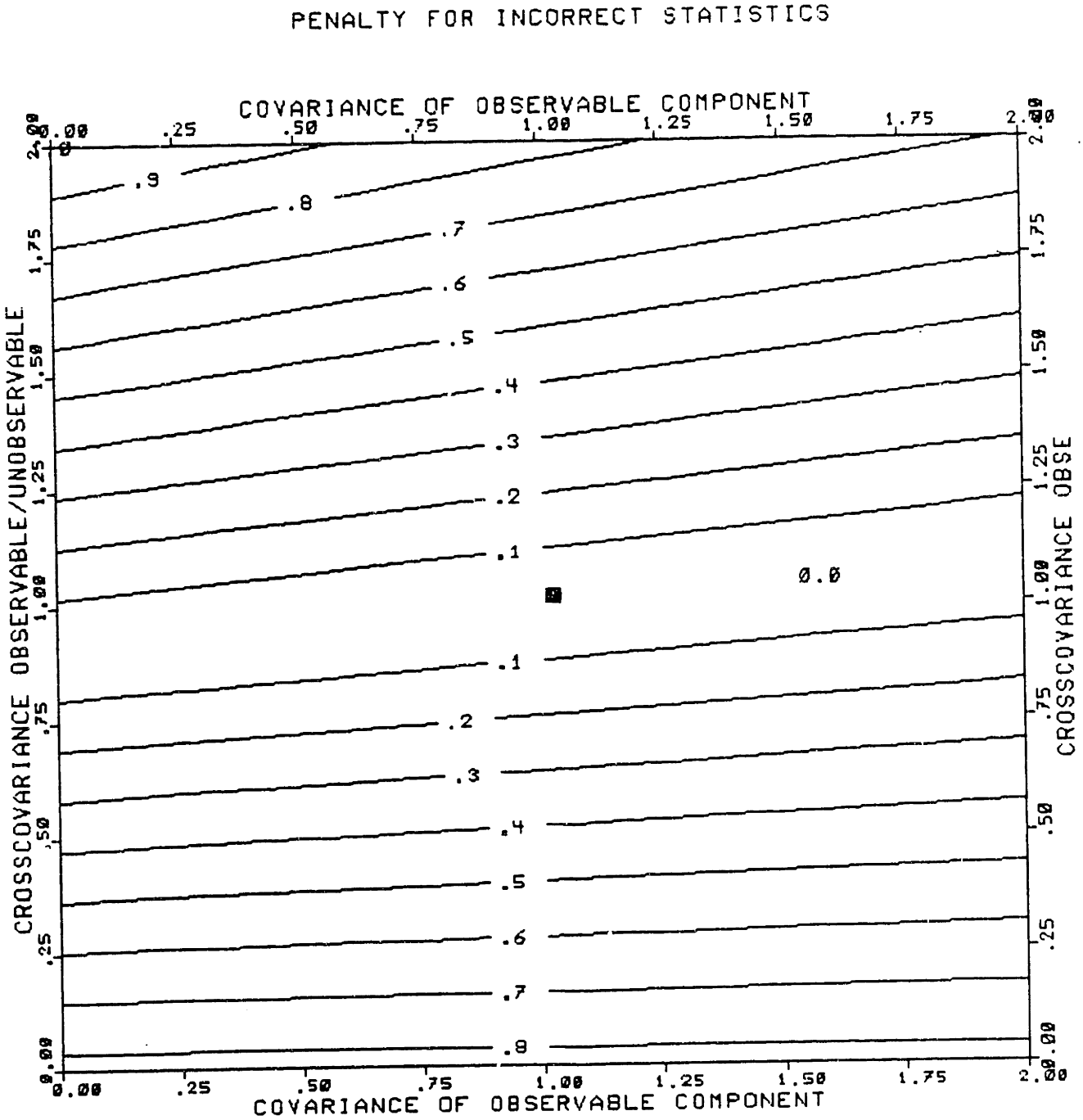


Figure VII.3.4: Normalized penalty for the y components in the case of bad signal-to-noise ratio.

on board electronics). Rather, it measures the relative strength of the deviation of the temperature profile from its a-priori mean compared to the noise power of the sounder. Because the deviations of a set of temperature from their mean value varies from set to set, this SNR varies even when keeping the instrument's characteristics constant.

From a adaptation point of view, the signal to noise should be kept to a low level through the proper choice of signal power. This somewhat counter-intuitive results can be explained as follows: The purpose of adapting the statistics and retrieval operators to changes in the air masses being estimated is to avoid large excursions of temperature profiles and resolve as many modes as possible (modes are defined as in section VII.2.1). This goal is accomplished by partitioning the temperature field into stationary subfields, each one with its own mean and covariance. The means of these subfields are, in general, different and demonstrate the variations of atmospheric temperature field with time and space. These variations model the deterministic variation of the mean with time and space. For the adaptation to succeed, the deviations of the temperature must be small, thus the signal power and noise must be comparable, within a single order of magnitude or so, and the SNR low. The SNR is low because of small variance in the signal and not because of large noise.

A second argument support that thesis: Figures VII.3.3 and VII.3.4 show that when the signal power is comparable to the noise power, the sample consistent optimization criterion is almost equivalent to the general optimization criterion. Hence, the derivation of adaptation operators based on that criterion will almost be optimal.

Another advantage at keeping the signal power low is that, as discussed in Chapter V, multidimensional retrieval operators will extend more within their region of support for the estimation of the observable components of the temperature

field more in the cases of low SNR than in cases high SNR. In doing so, any error of an impulsive nature, such as mountain or localized weather phenomena, will be smoothed out by the somewhat low pass filter effect of the retrieval operator for the observable components.

Finally, as for the discussion for multidimensional retrieval operators, one should note that the SNR is of matricial nature and that the spread of eigenvalues of the vertical covariance kernel plays an important role in the discussion of low versus high SNR's.

The impact of the SNR and shape of the statistics will be further studied in section VII.3.4 as the sensitivities of the retrieval error and of the optimal retrieval error to changes in the true and the a priori statistics are examined.

VII.3.3 A-Priori Mean Profile

The impact of a wrong a-priori mean was established in Chapter V and will be concisely summarized for completeness. Denoting by T_a the a-priori mean profile (usually a function of latitude, pressure and season) and by T_s the sample mean temperature profile, the bias error is

$$b = (I - D_a W)(T_a - T_s) \qquad (VII.3.22)$$

where D_a is the determination matrix. The impact of the bias on the mean square error is b^2.

VII.3.4 Sensitivity Analysis of the Retrieval Error to Error in the Statistics

In this section, we depart slightly from the previous discussion to examine when the retrieval error will be more sensitive to the statistics used in the retrieval process.

Using the previously derived results one can express the retrieval error obtained using the retrieval operator derived from the covariance kernel R_a over

the sample set described by the covariance R. Namely,

$$e = trace\{R - DWR\} + trace\{(R_{oo} + N')(D_a^t - D)(D_a - D)\} \qquad (VII.3.23)$$

The first term is the minimum possible error which is a function of solely the instrument (through its weighting matrix and the receiver noise power) and the dataset sample covariance R. The second term is the penalty for inaccurate statistics The impact of inaccurate statistics on the retrieval error is independent of the true value of the covariance kernel (except for multiplication by the $R_{oo} + N'$ term). On the contrary, the influence of the statistics on the optimum retrieval error varies with the sample covariance kernel R.

Expressing the minimum error in the SVD basis yields

$$e = trace\{N'(R_{oo} + N)^{-1}R_{oo}\} + trace\{R_{uu} - R_{uo}(R_{oo} + N')^{-1}R_{ou}\} \quad (VII.3.24)$$

Take the partial derivative of the error with respect to the three principal submatrices of R expressed in the SVD basis. After trivial manipulations,

$$\frac{\partial e}{\partial R_{oo}} = (R_{oo} + N')^{-1}(N^2 + R_{ou}R_{uo})(R_{oo} + N')^{-1} \qquad (VII.3.25)$$

$$\frac{\partial e}{\partial R_{uo}} = R_{uo}(R_{oo} + N')^{-1} \qquad (VII.3.26)$$

$$\frac{\partial e}{\partial R_{uu}} = I \qquad (VII.3.27)$$

The sensitivity does not depend on the absolute values of the diverse covariances, but rather on their relative values. In other words, sensitivity does not change under point wise multiplication of R and N by a scalar. This is not true of the sensitivity of the error to inaccurate statistics, which is proportional to R_{oo}.

The covariance of the unobservable components R_{uu} has no impact on the sensitivity. This is to be expected since it does not enter into account in the computation of the determination matrix.

The sensitivity increases as the signal-to-noise ratio decreases. A low SNR is typical of cases where the deviation of the profiles from their mean value is minimal. Such cases are encountered with vertical covariance kernels compiled using few degrees of freedoms or of very stable air masses which are characterized by a small uncertainty measure H_{spec} (Equation (VII.2.11)).

Sensitivity also increases as the cross-covariance between observable and unobservable components increases. This case corresponds to instances where the number of degrees of freedoms in the temperature profile is small compared to the number of pressure levels used in the computation of the radiative transfer equation or, again, to stable air masses. High sensitivity is typically associated with cases where the sounding process does not lose a significant amount of information about the variable being observed.

As the cross-covariance increases ($R_{uo} >> N$), the derivatives are rewritten using D_u, the determination matrix for the unobserved components. Hence,

$$\frac{\partial e}{\partial R} = \begin{pmatrix} \frac{\partial e}{\partial R_{oo}} & \frac{\partial e}{\partial R_{ou}} \\ \frac{\partial e}{\partial R_{uo}} & \frac{\partial e}{\partial R_{uu}} \end{pmatrix} = \begin{pmatrix} D_u^t D_u & D_u^t \\ D_u & I \end{pmatrix} \qquad (VII.3.28)$$

or

$$\frac{\partial e}{\partial R} = \begin{pmatrix} D_u^t \\ I \end{pmatrix} \begin{pmatrix} D_u & I \end{pmatrix} \qquad (VII.3.29)$$

The norm of $D_u^t D_u$ thus provides a global measure of sensitivity. Expressed in the natural representation of temperature profiles, the sensitivity norm is

$$\| \frac{\partial e}{\partial R} \| = \| (I - W^t (W W^t)^{-1} W) D^t D (I - W^t (W W^t)^{-1} W) \| \qquad (VII.3.30)$$

Sensitivity is small for the MIE (minimum information estimator) and for the NAPIE (non a-priori information estimator) which are both characterized by a null D_u.

A possible interpretation of the conditions of high sensitivity that combines the results for both observable and unobservable components is that more a-priori

information is given about the profile when the cross-covariance between different components of the temperature profile are large. Because more information is available, more information is used by the retrieval operator which, in turns, becomes more sensitive to it. On the other hand, in situations where the sensor noise power overwhelms the signal covariance matrix, the determination matrix does not use any information about the statistics, and the sensitivity as well the retrieval operator become null. This lack of robustness of the estimator in instances where the knowledge about the signal overrides the measurement noise (perfect observation) is a classical phenomenon in detection and estimation theory (Van Trees, 1963) but it was not analyzed in the context of remote sensing.

The above discussion demonstrates that the error is most sensitive to statistics and error in the radiative transfer equation when the number of profiles used in the compilation is small, that is when the retrieval operators are derived from small and specific datasets.

VII.4 Adaptation Schemes

In this section, expressions for diverse optimal estimators associated with the optimization criteria developed in section VII.3 are computed.

Adaptation is performed first in terms of the second moment (vertical covariance kernel and three dimensional covariance) and second in terms of the first moment (a-priori mean).

VII.4.1 Framework for the Adaptation of the Second Moment

It is natural to consider only second moments to adapt second moments. Hence, only the covariance matrix or the three dimensional statistics will be the input elements to any adaptation scheme. To guarantee symmetry and positive definiteness of the estimated vertical covariance matrix, one must consider a sym-

metric estimator of the form

$$\hat{R} = CR_{T_b T_b} C^t \qquad (VII.4.1)$$

where the adaptation operator C will be optimized momentarily.

This adaptation is performed as a three step process: First, compile the observed brightness temperature, correct them to clear column radiances, then compute mean and covariances and generate the adapted retrieval operator, and, finally, apply the retrieval operator on the same set.

As was previously discussed, the optimization of the optimum cost function for the minimization of the overall rms error (Equation VII.3.12) is an impossible task as far as the characterization of the vertical covariance kernel is concerned. Thus, the derivation of the adapted statistics and retrieval operators will be performed using the criterion optimum for observation consistent operators (Equation VII.3.20). Recall that, in cases where the SNR is low, that criterion is almost optimal.

VII.4.2 Observation-Consistent Retrieval Operators: Fisher Case

For this specific analysis, the diverse sample covariance matrices R are assumed to be completely unknown in advance, in the sense that their probability density function is not specified. One can think of the different sample covariance matrices as independent from dataset to dataset. The objective criteria to minimize is as before

$$e = trace\{\Delta R W^t (WRW^t + N)^{-1} W \Delta R\} \qquad (VII.4.2)$$

where $\Delta R = \hat{R} - R$. Plugging (VII.4.1) into (VII.4.2) yields

$$e = trace\{H(CWRW^t C^t - R)^2\} \qquad (VII.4.3)$$

where H is the impact matrix introduced in section introduced in section VII.3.2. Because each dataset covariance matrix is independent from other datasets matrices,

the minimization of the penalty must be performed set by set. Hence, a minimal norm solution to the design equation is advisable. As for the estimation of the temperature profiles, this minimum norm solution coincides with the Moore Penrose pseudo-inverse of the weighting matrix (see section V.2 and Moore (Rao, 1973)). That is

$$C = W^t(WW^t)^{-1} \qquad (VII.4.4)$$

This minimum type solution does not alter the components of the covariance matrix which involves the unobservable components of the temperature profile. Denote by the subscript h the historic covariance pressure kernel and by s the sample one. The adaptation equation is

$$\begin{aligned} R_a =& R_h + W^t(WW^t)^{-1}(R_{TbTbs} - N - WR_hW^t)(WW^t)^{-1}W^t \\ =& R_h + W^t(WW^t)^{-1}W(R_s - R_h)W^t(WW^t)^{-1}W^t \end{aligned} \qquad (VII.4.5)$$

Expressing this adaptation in the SVD basis yields

$$R_a = \begin{pmatrix} R_{oos} & R_{ouh} \\ R_{uoh} & R_{uuh} \end{pmatrix} \qquad (VII.4.6)$$

Provided that the sensor noise covariance equals its nominal value, the adaptation perfectly estimates the covariance of the observable components and leaves the other submatrices unchanged. This adaptation can thus be characterized as conservative, since it does not alter the hidden part of the statistics.

Unfortunately, being conservative with the covariance matrix adaptation is not advisable as far as the retrieval operator is concerned. Expressed in the SVD basis, the adapted retrieval operator is written as

$$D_a = \begin{pmatrix} R_{aoo}(R_{aoo} + N)^{-1} \\ R_{auo}(R_{aoo} + N)^{-1} \end{pmatrix} \qquad (VII.4.7)$$

where the covariance is calculated as above. Hence,

$$D_a = \begin{pmatrix} R_{soo}(R_{soo} + N)^{-1} \\ R_{huo}(R_{soo} + N)^{-1} \end{pmatrix} \qquad (VII.4.8)$$

Note that the retrieval operator for the unobservable components becomes dramatically altered when the sample and historic covariance matrices for the observable components are different.

This imperfection may yield disastrous effects when the eigenvalues of the sample covariance are smaller than those of the historic covariance. This situation will occur if the sample covariance is computed using few realizations of temperature profiles (characteristic of a specific air mass) and the historic covariance is compiled using numerous profiles from different climatologies.

Unfortunately, this type of situation is common in adaptation problems. When this is the case, the gain between deviation of the brightness temperatures from their a-priori mean values and unobservable components will be large. This will result in large rms errors if the observed brightness temperature has not been corrected for the errors in the radiative transfer (α, β scheme of section VI.4.1, or alike) and orographic errors. This difference between the magnitudes of the eigenvalues will result in large bias error if the true mean profile is different from the a-priori or adapted profile.

One way to correct for this problem is to consider the NAPIE rather than the full LMMSE. As discussed in Chapter V, the NAPIE sets R_{uo} to zero and sets the unobservable components equal to the unobservable a-priori mean. Adapting the NAPIE is guaranteed to reduce the retrieval error if the noise covariance matrix (which models sensor and uncertainty in the ground reflectivity) is correct.

This problem with the adaptation of the unobservable components comes from the non-positive definiteness of the adapted covariance matrix when the sample observable covariance matrix is too small entries. As a quick proof of this non-positive definiteness, let us consider the determinant of the adapted covariance

matrix. Denoting the determinant by $|.|$,

$$\begin{aligned} |R_a| &= \left| \begin{pmatrix} R_{soo} & R_{huo} \\ R_{hou} & R_{huu} \end{pmatrix} \right| \\ &= |R_{soo} - R_{hou} R_{huu}^{-1} R_{huo}| |R_{huu}| \end{aligned} \qquad (VII.4.9)$$

When the entries of R_{soo} are small, the above expression can be approximated by

$$|R_a| \approx -|R_{hou} R_{huu}^{-1} R_{huo}| |R_{huu}| \qquad (VII.4.10)$$

Since R_{huu} is a positive definite matrix, both its determinant and the determinant of a product of the form $AR_{huu}A^t$ are positive. Thus the determinant of the adapted covariance matrix is negative and the matrix is not positive semidefinite.

One must derive another form of adaptation scheme that avoids the shortcomings of the Fisher estimation. This task is performed in the next section where the adaptation problem is cast in a more general framework.

VII.4.3 Observation-Consistent Retrieval Operators: Bayesian Case

The Bayesian approach to the adaptation problem aims at incorporating additional knowledge about the statistics and notably an "average" covariance kernel which tends to overcome the shortcomings of the Fisher adaptation scheme. Bayesian adaptation attempts to provide a statistical description of the statistics. The proposed adaptation technique is based on the brightness temperature measurements and an additional description of the statistics. This description can be compiled on a monthly or yearly basis and represent the average of numerous air masses. It can also be provided by neighboring rawinsondes. The philosophy behind this technique is similar to adaptation techniques for scalar signal processing where the signal is assumed stationary over a finite window, and its statistics computed from the signal. Because of the inherent problem in vectorial signal processing of

observable/unobservable subspaces, one must include an average covariance matrix to perform the "extrapoladaptation" (extrapolation of the adaptation) from observable to unobservable.

The general philosophy can be phrased as follows. The temperature is modeled as a stochastic process whose covariance matrix (true second moment) is denoted by R. Any finite realization of this stochastic process has a sample covariance matrix S defined as a finite sum of outer product of temperature profiles with themselves (once the mean has been estimated and removed). An ergodicity assumption or a weak law of large numbers type of argument allows us to take the true covariance as the expected value of the possible sample covariance matrices.

Consider the expected value of the penalty for misspecifying the sample covariance over all realizations of sample covariance. As for the Fisher framework, the objective function to minimize is

$$e = trace\{H(CWSW^tC^t - S)^2\} \qquad (VII.4.11)$$

The difference between this framework and the previous one is that the expectation of the penalty function is performed over the ensemble of dataset sample covariance matrices. Appendix VII.C provides an expression for the above expectation in the specific case where the temperature field can be modeled as a Gaussian stochastic variable and the sample covariance matrix is replaced by a sum of outer products. Under these working assumptions, VII.4.11 becomes

$$\begin{aligned} e = {} & 2trace\{HCWRW^tC^tCWRW^tC^t\} + trace\{HCWRW^tC^t\}trace\{CWRW^tC^t\} \\ & - trace\{HRCWRW^tC^t\} + trace\{HRW^tC^tRCW\} \\ & -trace\{HCWRW^tC^tR\} + trace\{HW^tC^tRCWR\} + trace\{HCWR\}trace\{RCW\} \\ & +2trace\{HR^2\} + trace\{HR\}trace\{R\} + trace\{HRCW\}trace\{CWR\} \end{aligned} \qquad (VII.4.12)$$

This objective function is a quartic function of the determination matrix whose solution may not be analytically tractable.

Computing the gradient $\nabla_C e$ of that penalty function with respect to C and noting that the impact matrix is symmetric yields

$$\begin{aligned}\nabla_C e =& 4CHWRW^tC^tCWRW^t + trace\{CWRW^tC^tH\} \\ &+ trace\{CWRW^tC^t\}HCWRW^t - 2HCRCWRW^t - 2HRCWRW^t \\ &- 2HRW^tC^tRW^t - trace\{HCWR\}RW^t - trace\{CWR\}HRW^t\end{aligned} \tag{VII.4.13}$$

The optimal estimator is found by setting the gradient to zero.

Because the impact matrix is not full rank, the optimal estimator, if it exists, is not unique. Furthermore, it is quite difficult to solve the design equation. A somewhat inspired guess for the estimator follows the same form as that of the estimation of the temperature profile, namely

$$C = RW^t(WRW^t)^{-1} \tag{VII.4.14}$$

Substituting this estimator into (VII.4.12) yields the following value of the gradient for such a C matrix.

$$\begin{aligned}\alpha =& 4HRW^t(WRW^t)^{-1}WR^2W^t + trace\{HRW^t(WRW^t)^{-1}WR\}RW^t \\ &+ trace\{RW^t(WRW^t)^{-1}WR\}HRW^t - 2HR^2W^t - 2HRW^t(WRW^t)^{-1}WR^2W^t \\ &- trace\{HRW^t(WRW^t)^{-1}WR\}RW^t - trace\{RW^t(WRW^t)^{-1}WR\}HRW^t\end{aligned} \tag{VII.4.15}$$

After algebraic simplifications,

$$\alpha = HRW^t(WRW^t)^{-1}WR^2W^t - HR^2W^t \tag{VII.4.16}$$

Using the transfer matrices $\mathcal{T}_n$ and $\mathcal{T}$ which correspond respectively to estimation in the presence and the absence of noise, namely

$$\begin{aligned}\mathcal{T}_n =& RW^t(WRW^t + N)^{-1} \\ \mathcal{T} =& RW^t(WRW^t)^{-1}\end{aligned} \tag{VII.4.17}$$

yields

$$\alpha = \mathcal{T}_n^t(\mathcal{T}^t - I)RW^t \tag{VII.4.18}$$

Both transfer matrices share the same null and row subspaces. Since the null space of the difference $I - \mathcal{T}$ is the row space of $\mathcal{T}_n$, the product of these two matrices has the entire vectorial space of the temperature profiles as its null space. Only the null matrix possess such properties among real matrices; thus α is null and the proposed estimator is optimal, albeit not unique.

The proposed solution to the adaptation problem is the homogeneous solution to the adaptation equation, in the sense the gradient α will equal zero even if the impact matrix is full rank.

The optimal Bayesian adaptation operator ($C = RW^t(WRW^t)^{-1}$) for the covariance matrix uses the same functional form as the Bayesian estimator for the temperature profile ($D = SW^t(WSW^t + N)^{-1}$). The salient difference is that the adaptation operator operates on both the left and right sides of the covariance matrix of the brightness temperature once the noise covariance N corresponding to the noise has been removed.

To express the Fisher adaptor as a special case of the Bayesian adaptor, one must employ a constant diagonal matrix R.

Let us now examine the adaptation process and derive expressions for the adapted covariance matrix, retrieval operator, and average penalty.

Adaptation for the second moment can be cast as follows: One has two covariances matrices R_h and R_c. The historical covariance matrix R_h is characteristic of an air mass or set of temperature profiles close, from a climatology point of view, to the set of temperature being estimated. This closeness is characterized from geographical or temporal arguments. This covariance matrix is typically computed using a small number of degrees of freedom. It should have the location of the tropopause inversion and the lapse (with or without inversion) close to the ground with reasonable accuracy because these two elements are essential to good performance of the retrieval process.

The climatological covariance matrix R_c, on the other hand, is obtained by averaging numerous covariances matrices, each characteristic of specific sets. It can also be an analytical model, such as the mexican hat for the vertical covariance matrix, or the model developed in Chapter III for the three-dimensional description). Note that nothing in the expression of the estimator prevents those two matrices R_c and R_h to be identical.

Given these matricial descriptions, and given the sample covariance matrix of brightness temperature R_{TbTb} (associated with the sample temperature covariance matrix R_s), the historic covariance matrix R_h can be adapted to the covariance matrix R_a according to

$$R_a = R_h + R_c W^t (W R_c W^t)^{-1} (R_{TbTb} - N - W R_h W^t)(W R_c W^t)^{-1} W R_c \quad (VII.4.19)$$

Expressing this adaptation in the SVD basis representation and using the same indexation scheme as in Chapter V and in the section on Fisher adaptation (u for unobservable and o for observable), the adaptation operator is written as

$$C = \begin{pmatrix} I \\ R_{cuo} R_{coo}^{-1} \end{pmatrix} \quad (VII.4.20)$$

and the adapted vertical covariance matrix

$$R_a = \begin{pmatrix} R_{soo} & R_{hou} + (R_{soo} - R_{hoo}) R_{coo}^{-1} R_{cou} \\ R_{huo} + R_{coo}^{-1} R_{cuo} (R_{soo} - R_{hoo}) & R_{huu} + R_{cuo} R_{coo}^{-1} (R_{soo} - R_{hoo}) R_{coo}^{-1} R_{ou} \end{pmatrix} \quad (VII.4.21)$$

In the special case where climatological and historical covariance matrices coincide, the adapted covariance matrix takes a simpler form

$$R_a = \begin{pmatrix} R_{soo} & R_{soo} R_{coo}^{-1} R_{cou} \\ R_{cuo} R_{coo}^{-1} R_{soo} & R_{huu} + R_{cuo} R_{coo}^{-1} (R_{soo} - R_{hoo}) R_{coo}^{-1} R_{cou} \end{pmatrix} \quad (VII.4.22)$$

The Bayesian adaptation alters all submatrices of the vertical covariance matrix. As in the Fisher case, the covariance of the observable components is estimated exactly. It should be noted that, although changed by the estimation, the

covariance matrix of unobservable components with themselves is irrelevant to the sensing process. This liberty in choosing R_{auu} reveals that the optimal adaptation scheme under the consistency constraint is not unique. This assertion is confirmed by the non-full rankness of the impact matrix H in VII.18.

The adapted determination matrix expressed in the SVD basis takes the now familiar form

$$D_a = \begin{pmatrix} R_{aoo}(R_{aoo}+N)^{-1} \\ R_{auo}(R_{aoo}+N)^{-1} \end{pmatrix} \qquad (VII.4.23)$$

where the covariances are calculated above. The general adaptation scheme yields

$$D_a = \begin{pmatrix} R_{soo}(R_{soo}+N)^{-1} \\ R_{huo}(R_{soo}+N)^{-1} + R_{cuo}R_{coo}^{-1}(R_{soo}-R_{hoo})(R_{soo}+N)^{-1} \end{pmatrix} \qquad (VII.4.24)$$

This adapted covariance matrix can be written as the sum of two operators

$$D_a = \begin{pmatrix} R_{soo}(R_{soo}+N)^{-1} \\ R_{huo}(R_{soo}+N)^{-1} \end{pmatrix} + \begin{pmatrix} 0 \\ R_{cuo}R_{coo}^{-1}(R_{soo}-R_{hoo})(R_{soo}+N)^{-1} \end{pmatrix} \qquad (VII.4.25)$$

The first of these operators can be recognized as the Fisher solution to the adaptation problem. The second operator which depends on the climatological covariance matrix is a correction operator which accounts for the additional knowledge concerning the evolution of vertical covariance kernels. Should the climatological description be a diagonal covariance matrix, the submatrix R_{cuo} will be null and the Bayesian adapted retrieval operator will become the Fisher adapted retrieval operator.

In the special case where historic and climatological vertical covariance matrices coincide, the adapted retrieval operator is

$$D_a = \begin{pmatrix} R_{soo}(R_{soo}+N)^{-1} \\ R_{cuo}R_{coo}^{-1}R_{soo}(R_{soo}+N)^{-1} \end{pmatrix} \qquad (VII.4.26)$$

When the matricial signal-to-noise ratio of the sounding process is large, the retrieval operator adopts the limiting form

$$D_a = \begin{pmatrix} R_{soo}(R_{soo}+N)^{-1} \\ R_{cuo}R_{coo}^{-1} \end{pmatrix} \qquad (VII.4.27)$$

The above expression demonstrates that the retrieval operator estimating the unobservable statistics is solely based upon the historical statistics when it coincides with the climatological one. The estimation of fine features such as lapse rate or changes in the slope of the profile around the tropopause which are better, in general, characterized by the unobservable components of the temperature profile, will not be altered by the adaptation process.

Should the historical statistics (obtained by say the small set method) place the tropopause inversion at 250 *mbar* rather than the actual say 200 *mbar*, thus resulting in a large retrieval error at that location, the adaptation scheme will not reduce the estimation error if the vectorial subspace spanned by the rows of the weighting matrix does not permit a sharp discontinuity around 200 *mbar* which would correct for the incorrect statistics. This problem in adapting the weighting matrix unobservable subspace exists in cases where the knowledge about the statistics is only a single a-priori matrix.

It should be noted that in the specific case where historical and climatological descriptions coincide, the Fisher and Bayesian adaptation operators differ in the way they modify the elements of the retrieval process. The Fisher adaptation leaves the covariance submatrices related with the unobservables components R_{uo}, R_{ou}, R_{uu} unchanged whereas the Bayesian adaptation leaves determination submatrix estimating the unobservable components D_u unchanged.

The average penalty for misspecifying the covariance matrix is given by Appendix VII.C and equals

$$\begin{aligned} e =& trace\{H\mathcal{T}R\mathcal{T}R\} + 2trace\{HR^2\} + trace\{HR\}trace\{R\} \\ & - 2trace\{HR\mathcal{T}R\} - 2trace\{H\mathcal{T}^t R\mathcal{T}R\} \\ & - trace\{HR\mathcal{T}\}trace\{\mathcal{T}R\} \end{aligned} \qquad (VII.4.28)$$

where the matrix $\mathcal{T}$ is the transfer matrix for the perfect observation case associated

with the transmission R,

$$\mathcal{T} = RW^t(WRW^t)^{-1} \qquad (VII.4.29)$$

This expression for the error does not provide any worthwhile interpretation. It should be noted that in the above analysis, the number of elements used to evaluate the sample covariance matrices did not play any role. Hence, sample matrices having few degrees of freedoms (characteristic of specific air masses) and matrices having many more degrees of freedoms were treated alike.

VII.4.4 Inclusion of Multidimensionality.

When the sounding angle changes across the sounding frame, so do the weighting functions and the covariance matrix of the resulting brightness temperature. Because of the change in the weighting matrix $w(n)$ with sounding angle $\alpha(n)$ ($n = 0,\ldots,5$ for MSU), the observed covariance matrices are rotated versions of the same covariance matrix for different observable components of the temperature field. This analysis assumes that the vertical covariance of a particular set does not change with its position with respect to the track of the satellite, which can be expected.

The brightness temperature at angle $\alpha(n)$ follows the now traditional expression

$$R_{T_bT_b}(n,n) = w(n)R_{TT}w^t(n) + N \qquad (VII.4.30)$$

R_{TT} can then be adapted using the methods already proposed. However, it is not possible to group these different weighting functions into a larger one according to

$$\overline{\overline{W}} = (w^t(0),\ldots,w^t(N)) \qquad (VII.4.31)$$

This is true because the cross-covariance $R_{T_bT_b}(n,m)$ involves the isobaric correlation of temperature field when collected by the instrument and does not involve it when used in the estimation.

The solution to the adaptation problem in a multidimensional context is to use a different adaptation operator at each sensing angle. This yields an iterative algorithm. For example, the Bayesian approach is

$$R(n+1) = R(n) + R_c w^t(n)(w(n)R_c w^t(n))^{-1}(R_{TbTb}(n) - N - w(n)R(n)w^t(n))$$
$$(w(n)R_c w^t(n))^{-1} w(n) R_c \qquad (VII.4.32)$$

The final vertical covariance kernel depends upon the sequence of brightness temperatures used, except in the unlikely case where the observable components at each sounding angle are orthogonal to the observable components at the other angles. This last condition is expressed mathematically as

$$w^t(n)w(m) = \delta_{nm} \qquad (VII.4.33)$$

where δ is the kronecker symbol. In order to minimize the impact of error in the ground reflectivity on the adaptation process, it is better to put more emphasis on the "cleanest" observation equation. This is why the covariance is changed from nadir to the extreme observation angle.

VII.4.5 An Ad-Hoc Adaptation Scheme.

A possible adaptation scheme derived from heuristic considerations consists of computing the covariance matrix of the retrieved temperature profiles and using it in the computation of the retrieval operator. This adaptation scheme is relevant to this analysis because it is related to the Bayesian adaptation in one of its limiting cases.

The retrieved temperature profile is given by

$$\hat{\overline{T}} = < \overline{T}_h > + D_h W(\overline{T} - < \overline{T}_h >) + D_h \eta \qquad (VII.4.34)$$

Subtracting the sample mean of the estimated profile $< \hat{\overline{T}}_s >$ yields

$$\hat{\overline{T}} - < \hat{\overline{T}}_s > = D_h W(\overline{T} - < \overline{T}_s >) + D_h \eta \qquad (VII.4.35)$$

where the subscript h denotes the historic statistics and s the sample statistics. The covariance matrix of such a collection of profiles is

$$\begin{aligned} R_{\hat{T}\hat{T}} &= < (\hat{\overline{T}} - < \hat{\overline{T}}_s >)(\hat{\overline{T}} - < \hat{\overline{T}}_s >)^t > \\ &= D_h(W < (\overline{T} - < \overline{T}_s >) < (\overline{T} - < \overline{T}_s >) W^t + N) D_h^t \end{aligned} \qquad (VII.4.36)$$

That is, denoting by R_{TbTb_s} the sample covariance of the brightness temperature

$$R_{\hat{T}\hat{T}} = D_h R_{TbTb_s} D_h^t \qquad (VII.4.37)$$

The positive semi-definiteness of $R_{\hat{T}\hat{T}}$ is guaranteed by the positive definiteness of R_{TbTb_s}. However, $R_{\hat{T}\hat{T}}$ is not full rank since the retrieval operator D_h is not.

Expressing the adapted vertical covariance kernel in the SVD basis to compare the expression with the ones of previous adaptation schemes yields

$$R_{\hat{T}\hat{T}} = \begin{pmatrix} R_{hoo}(R_{hoo}+N')^{-1}(R_{soo}+N')(R_{hoo}+N')^{-1}R_{hoo} & R_{hoo}(R_{hoo}+N')^{-1}(R_{soo}+N') \\ R_{huo}(R_{hoo}+N')^{-1}(R_{soo}+N')(R_{hoo}+N')^{-1}R_{hoo} & R_{huo}(R_{hoo}+N')^{-1}(R_{soo}+N') \end{pmatrix} \qquad (VII.4.38)$$

As in previous cases of adaptation, this expression simplifies whenever the matricial SNR is large. In such a case,

$$R_{\hat{T}\hat{T}} = \begin{pmatrix} R_{soo} & R_{soo}R_{hoo}^{-1}R_{hou} \\ R_{huo}R_{hoo}^{-1}R_{soo} & R_{huo}R_{hoo}^{-1}R_{soo}R_{hoo}^{-1}R_{hou} \end{pmatrix} \qquad (VII.4.39)$$

Except for the covariance of the unobservable components, this adaptation scheme yields the same solution as the Bayesian adaptation in the special case where historical and climatological covariances are the same. This fact shows that the Bayesian adaptation scheme is not appropriate for adaptation when only one statistical description is available.

VII.4.6 Impact of the Adaptation

To determine it the adaptation is beneficial, one needs only to use the different expressions for the adapted retrieval operators in the expression for the penalty and compare it with the penalty for the original operator.

The historical (original) penalty for using the wrong statistics is

$$p_h = trace\{(R_{oo} + N)(D_h^t - D_s^t)(D_h - D_s)\} \qquad (VII.4.40)$$

and the penalty for the adapted operator is

$$p_a = trace\{(R_{oo} + N)(D_a^t - D_s^t)(D_a - D_s)\} \qquad (VII.4.41)$$

The improvement through adaptation can be written using the fact that for symmetric matrices A the trace of ABC equals that of AC^tB^t and

$$\begin{aligned} \Delta p =& trace\{2(R_{oo} + N)(D_a^t - D_h^t)(\frac{D_a + D_h}{2} - D_s)\} \\ =& trace\{(R_{oo} + N)(D_a^t - D_h^t)(D_a - D_s + D_h - D_s)\} \end{aligned} \qquad (VII.4.42)$$

Improvements for both the observable and unobservable components follow the same type of relationship.

For the observable components, if the sensor noise sample covariance matrix N equals its nominal value, the improvement Δp is always greater or equal to zero, and the adapted retrieval operator is optimal. This result has important consequences for sounding systems with an observable subspace that captures most of the signal energy. For such sounding systems, the adaptation benefits are controlled by the improvement in the observable components which is always positive, hence these systems (such as AMSU) have improved performances through adaptation.

This is not the case for unobservable components where the improvement can go either way. In the special case of Bayesian adaptation, the improvement can be positive or negative (that is degradation) on any specific set. However, the improvement's average value is positive if the climatological covariance matrix in (VII.4.19) equals (and by extension is close to) the covariance matrix which corresponds to the average optimal retrieval operator. Namely if

$$D(R_c) \approx \frac{1}{N} \sum_{i=1}^{N} D(R_{si}) \qquad (VII.4.43)$$

where the sum is performed over all optimal sample set retrieval operators whose average covariance matrix is supposed to be R_c. However, in general, there is no method to verify if the adaptation is working to our advantage or not. The matrices used in the adaptor must thus be chosen carefully on climatological arguments.

It should be noted that adaptation is not a bad technique, especially for newer systems such as AMSU where most of the signal energy is spanned by the observable subspace. Moreover, the lifting of the weighting function with observation angle results in changes in the fundamental subspaces of the weighting matrix. Hence use of multidimensionality in the adaptation process can notably improve its performance for systems with few channels such as MSU. Nevertheless, adaptation of multi-channel retrieval operators cannot exhibit the flexibility of scalar adaptator unless an analytical form can be found for the vertical covariance which depends upon only a few parameters. Should such an analytical model exists, the entire vertical covariance kernel could be estimated from the observable components covariance matrix. Models of this sort, including the ones discussed in Chapter III, are unlikely to accurately model the discontinuity at the tropopause. This modeling is essential to the estimation of the temperature at the tropopause.

Before concluding this Chapter with a description of the implementation of adaptation algorithms (for both simulated and real data, MSU and AMSU), this theoretical analysis concludes with another application of adaptation which is related to the maximum likelihood and Bayesian extrapolation-based assimilation techniques developed in Appendix IV.A.

Note that the weighting functions used in the observation equation are arbitrary full row-rank matrices. They could be sparse matrices with ones as the only non-zero entries. Such matrices correspond to the observation of a temperature profile at fewer pressure levels than the 33 levels used in the model. The equation used for the adaptation (Bayesian) can extrapolate the sample covariance computed

at those limited pressure levels to the entire 33 levels. Such an example of limited sampling is the set of 12 levels (ground up to 50 *mbar*) that the NMC analysis field provides.

This extrapolation (according to equation VII.4.) is identical to the Bayesian extrapolation of Chapter IV except for the covariance of unsampled pressure levels (unobservable components covariance R_{uu}).

VII.4.7 Adaptation of the a-Priori Mean Profile

Adaptation of the mean profile can be cast as estimation of the mean bias between the a-priori and the sample mean profile field. This estimation is in fact a retrieval problem which can be solved using single or multi-spot retrieval operator.

The estimator for the bias operates on the average biases of the different components of the brightness temperature. Namely,

$$\hat{\overline{b}} = D_a < (\overline{T}_b(\theta) - W\overline{T}_a(\theta)) > \qquad (VII.4.44)$$

where θ denotes the latitude. This formulation shows that, although the mean profile is not constant over the considered set, the correction is.

For the adaptation of the covariance kernel (and other second moments), one can not use the different weighting functions resulting from changes in the sounding angle without using a-priori three-dimensional statistics. This is not the case for the adaptation of the mean bias. The only assumption required to combine these different measurements is that the correction be the same at all latitudes. This assumption allows the synthesis of a sensor with more channels than the satellite really possesses. For MSU, $6 \times 3 = 18$ such channels are formed by combining the three temperature sounding channels at the six different sounding angles. This synthesis permits to put more of the bias components into the observable space and thus relies less on statistics.

Once the bias is found, it is incorporated into the a-priori profile and the re-

trieval begins. One must note that even under ideal conditions (linear RTE, perfect vertical discretization, impeccable knowledge of the topography and reflectivity), the resulting brightness temperature field will not be zero mean. That is, the mean deviation of the observed brightness temperature from the value computed using the corrected mean profile is not zero. The adapted mean profile is

$$< \hat{\overline{T}} >= \overline{T}_a + \hat{\overline{b}} = \overline{T}_a + D_a < \overline{T}_b > -D_a W \overline{T}_a \qquad (VII.4.45)$$

and the bias for the brightness temperature equals

$$< \overline{T}_b > -W < \hat{\overline{T}} >= (I - WD_a)(< \overline{T}_b > -W\overline{T}_a) \qquad (VII.4.46)$$

This quantity is not equal to zero, except when the sensor noise covariance is null. This mean brightness temperature is when the signal-to-noise ratio of the process is large.

To compensate for this difference, one can consider an iterative approach similar to the retrieval algorithm proposed by Chahine (1969) where the new estimate of the mean profile is injected as the a-priori mean profile of the next operation. That is,

$$< \hat{\overline{T}}_{n+1} >=< \hat{\overline{T}}_n > +D_a(n) < \overline{T}_b > -D_a(n) W < \hat{\overline{T}}_n > \qquad (VII.4.47)$$

The eigenvalues of the difference $(I - WD_a)$ are less than one for any valid determination matrix. Thus, this iterative algorithm converges to an "observation consistent" mean temperature profile. However, this mean profile is not optimal from a meansquare point of view. It will be nearly optimal when the signal-to-noise ratio is large. This result is the opposite of the D matrix case where the consistency with the observation is nearly optimal for bad SNR cases.

Recall that the mean profile is a vector indexed along the latitude axis (perfect for a time series representation), thus estimations techniques previously

introduced for temperature profiles may be reconsidered as adaptation schemes. One can consider Kalman filtering (see Ledsham and Staelin, 1978) where the simple Markovian assumption is applied to determine the evolution of the mean profile and three dimensional filters are used to further reduce retrieval error.

Another possibility is to merge latitudinally the a-priori mean with those retrieved profiles to produce a new mean. The weight associated with the a-priori mean can be determined from temporal persistence considerations (Lilly, 1983) and, thus yield finite memory forgetting filters (Schweppe, 1973).

These two methods are ad-hoc attempts to model the dynamic variations of the atmospheric flow. Such an implementation using a true weather prediction scheme (but not combined with three dimensional retrieval operators) was implemented with little success by Toldalagi (1980).

However the selective replacement of the mean can be thought as the identification of modes in the retrieval dataset. As exhibited in section VII.2.1, this identification in terms of the mean vector is efficient for systems with few channels. Hence, it should work for MSU. Only the replacement of mean on a latitudinal basis was considered in this dissertation in order to test the later hypothesis.

VII.5 Implementations of Adaptation Schemes

As for implementations of multidimensional retrieval operators, errors in the weighting functions around and above the tropopause and orographic effects will impair our ability to perform the adaptation. As for classical retrieval implementations of this dissertation, both simulated (for AMSU and MSU) and real implementations of adaptation schemes for MSU will be considered.

VII.5.1 Microwave Sounding Unit

* marks the operator(s) that yields the lowest rms error.
a-priori mean is interpolated US standard atmosphere.
latitude between 34 and 54, longitude between -135 and -75
latitude between 35 and 55, longitude between -10 and 30
statistics compiled using 3223 retrievals, all sounding angles

pressure	bias	rms	std		bias	rms	std		bias	rms	std	
1000 mb	-3.1	5.2	4.1	*	-4.4	6.3	4.5		-3.2	5.2	4.0	*
850 mb	-1.8	2.6	1.9		-2.6	3.9	2.9		-1.9	2.7	1.9	
700 mb	-0.7	1.6	1.4		-0.8	1.6	1.4		-0.8	1.6	1.4	
500 mb	0.6	1.7	1.6	*	1.1	2.1	1.8		0.6	1.7	1.6	*
400 mb	1.6	2.3	1.6		2.0	2.8	1.9		1.6	2.3	1.6	
300 mb	2.6	3.0	1.5		2.8	3.4	1.9		2.6	3.0	1.5	
250 mb	2.7	3.3	1.9		2.5	3.2	2.0		2.8	3.6	2.2	
200 mb	0.1	2.4	2.4		-0.4	2.1	2.1	*	0.1	2.3	2.3	
150 mb	-3.6	4.2	2.2		-3.5	4.5	2.8		-3.7	4.4	2.3	
100 mb	-2.4	3.0	1.9		-2.3	3.2	2.2		-2.5	3.1	1.9	
70 mb	0.3	0.9	0.8		0.2	0.7	0.7	*	0.1	1.4	1.4	
50 mb	2.4	3.1	1.9		2.4	3.2	2.2		2.5	3.2	2.0	
	summer Europe				summer USA				July Europe			
1000 mb	-3.7	5.3	3.8									
850 mb	-1.3	2.5	2.2	*								
700 mb	-0.6	1.4	1.3	*								
500 mb	0.8	1.7	1.5	*								
400 mb	1.6	2.2	1.5	*								
300 mb	1.7	2.4	1.7	*								
250 mb	1.7	2.4	1.7	*								
200 mb	-0.7	2.1	2.0	*								
150 mb	-2.1	3.2	2.4	*								
100 mb	-1.3	2.4	2.0	*								
70 mb	-0.3	0.8	0.7									
50 mb	1.7	2.5	1.9	*								
	summer Peoria											

Table VII.5.1: rms error for summer case using single-spot retrieval operators derived from four different statistics

The impact of the vertical covariance kernel is illustrated in Table VII.5.1 where five different climatologies have been used to derive single-spot retrieval op-

erators which is turn are used above western Europe and the USA. There is no dramatic difference in the rms errors except in the troposphere where between the best and the worse statistics the error may degrade by as much as 55 % (check the 850 *mbar* level between the operator derived from summer profiles over Europe to the one derived from summer profiles over Peoria). Note that with very few exception, the operator derived from Peoria statistics outperforms the other operators, hence its choice for the summer implementation of Chapter VI. Note that the rms error and its standard deviation vary the most close to the ground and around the tropopause inversion. In the middle troposphere and stratosphere where statistics plays a lesser role because of the large values of the weighting functions or because of the stability of the atmosphere, the rms error varies less from D matrix to D matrix (check 700, 500, 400, and 100 *mbar*).

It may appear peculiar that retrieval operators derived from the statistics of temperature profiles collected over the USA would be outperformed by operators derived from more peculiar statistics, such as the 100 profiles from Peoria. One must bear in mind that the statistics used to compute the retrieval operator is provided by NMC observational data (ADP report) as reported by Chapter IV, whereas the NMC analysis field ("ground truth") is found by assimilating data from the observational data and satellite radiances. Moreover, the alpha-beta scheme required to correct for the discrepancies between computed and observed brightness temperature alters the statistics of the temperature field under analysis.

Two simulations of the impact of adaptation of statistics for MSU are presented here. The first one (Figure VI.5.1) considers the impact of adaptation in the Fisher framework where only the covariance matrix of the observable components of the profile are changed. In the exhibited simulation, the historic statistics are for summer over the United States whereas the sample statistics are for summer over Peoria. In this case the adaptation uses only information collected at

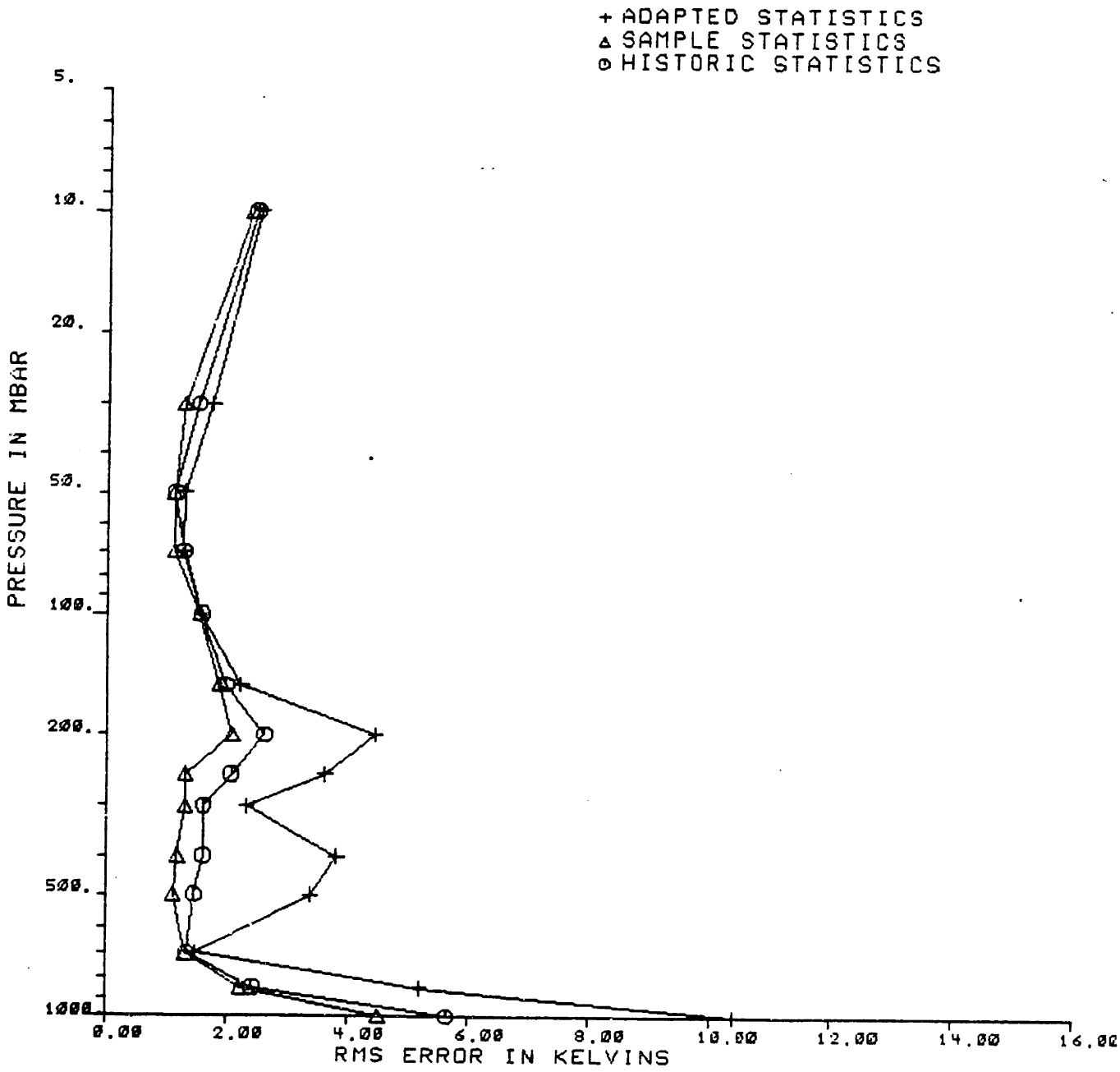


Figure VII.5.1: Fisher-type adaptation for MSU at nadir

nadir where the retrievals are assumed performed. The adapted retrieval operator performs much worse than either the optimal or original suboptimal D matrices, which shows that the Fisher adaptation is not appropriate. The second example of adaptation for MSU is set in the Bayesian framework and consider the same climatologies as above. The climatological statistics used in this adaptation (see

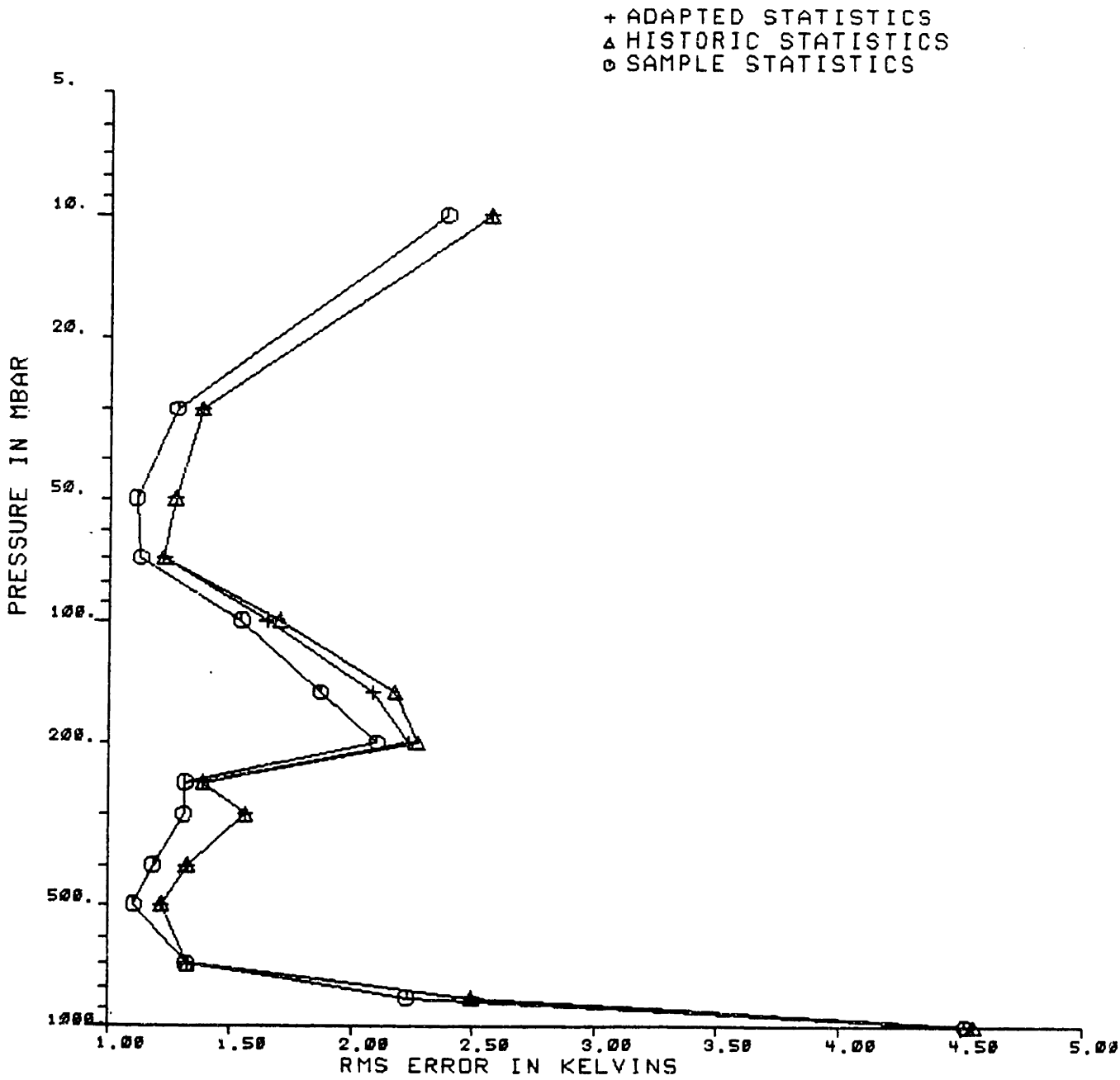


Figure VII.5.2: Bayesian-type adaptation for MSU at nadir

equation VII.4.19) are the concatenation of summer over Canada and the United States. The adaptation is beneficial in the lower stratosphere but leaves most of the retrieval operator and rms error unchanged. This result is typical of the adaptation for MSU. For the record, one must note that the author was hard pressed to find an MSU simulation where improvement through adaptation was noticeable. Based

on these simulations alone, the future of adaptation for MSU looks gloomy.

Consider now the impact of adaptation on real measurements. Adaptation of first and second moments for MSU yielded in general an improvement less or equal to the improvement obtained from the inclusion of several spots in the retrieval process. That is, starting from a single-spot retrieval operator, it is more advantageous to include more points in the retrieval process than performing the adaptation of statistics. Doing both, of course, is the best solution.

In some cases, the improvements on implementations were found larger than could be expected from the simulations, mostly because of changes in the a-priori mean vector which could not be taken into account easily for the Bayesian adaptation. As it will soon appear, this improvement is strongly dependent on the sounding angle.

To be fair in the analysis of adaptation, one must start the adaptation with a filter derived from statistics close the sample statistics. Hence for the summer case, the historic statistics are derived from the Peoria profiles. The climatological statistics used for the Bayesian adaptation were obtained by merging summer statistics over Europe and the USA.

Table VII.5.2 present the constant correction vector applied to all latitudes of the a-priori mean vector according to equation VI.4.43. Table VI.5.3 present the historic vertical covariance kernel as well as the sample covariance. Note the large changes in the power distribution close to the ground. Tables VII.5.4 and VII.5.5 compare the normalized covariance kernels for the historical, adapted and the 12-level NMC field statistics corresponding to the locations of the retrievals.

Table VII.5.7 presents the results of two different adaptation experiments using single frame (two-dimensional) retrieval operators: In the first experiment, only the mean is adapted, whereas in the second experiment both mean vector and covariance matrices are adapted. Note the mean adaptation does not provide any

```
Correction to the mean vector
pres:  1000 mbar cor:  -2.79 K
pres:  950 mbar cor:  -2.19 K
pres:  900 mbar cor:  -2.15 K
pres:  850 mbar cor:  -2.18 K
pres:  800 mbar cor:  -2.04 K
pres:  750 mbar cor:  -2.13 K
pres:  700 mbar cor:  -2.29 K
pres:  650 mbar cor:  -2.21 K
pres:  600 mbar cor:  -1.97 K
pres:  550 mbar cor:  -1.99 K
pres:  500 mbar cor:  -2.09 K
pres:  450 mbar cor:  -2.18 K
pres:  400 mbar cor:  -2.23 K
pres:  350 mbar cor:  -2.60 K
pres:  300 mbar cor:  -2.63 K
pres:  250 mbar cor:  -2.24 K
pres:  200 mbar cor:  -1.32 K
pres:  175 mbar cor:  -0.22 K
pres:  150 mbar cor:  0.86 K
pres:  125 mbar cor:  1.57 K
pres:  100 mbar cor:  0.89 K
pres:  80 mbar cor:  0.09 K
pres:  70 mbar cor:  -0.20 K
pres:  60 mbar cor:  -0.36 K
pres:  50 mbar cor:  -0.51 K
pres:  40 mbar cor:  -0.60 K
pres:  30 mbar cor:  -0.58 K
pres:  25 mbar cor:  -0.72 K
pres:  20 mbar cor:  -0.91 K
pres:  15 mbar cor:  -0.95 K
pres:  10 mbar cor:  -1.05 K
pres:  7 mbar cor:  -1.15 K
pres:  5 mbar cor:  -0.97 K
```

Table VII.5.2: Constant bias correction for summer adaptation

significant improvements in terms of bias error. Several reasons can advanced to

VERTICAL COVARIANCE KERNEL: historic covariance matrix

	1000	850	700	500	400	300	250	200	150	100	70	50
1000	33.8	15.3	9.7	7.6	6.9	6.8	4.7	-0.8	-5.8	-4.2	-2.7	-0.9
850	15.3	14.1	8.7	6.0	6.2	6.6	4.3	-1.0	-5.8	-5.0	-2.5	-0.1
700	9.7	8.7	10.6	7.2	7.4	7.7	4.5	-2.0	-6.4	-5.6	-1.8	-0.1
500	7.6	6.0	7.2	7.8	7.5	7.4	4.7	-1.7	-6.1	-5.8	-2.2	-1.0
400	6.9	6.2	7.4	7.5	8.5	8.4	5.2	-1.8	-6.9	-6.1	-2.2	-1.0
300	6.8	6.6	7.7	7.4	8.4	10.1	6.9	-0.3	-7.6	-6.5	-2.1	-0.8
250	4.7	4.3	4.5	4.7	5.2	6.9	7.3	3.2	-5.0	-4.8	-1.5	-0.4
200	-0.8	-1.0	-2.0	-1.7	-1.8	-0.3	3.2	10.8	3.2	1.1	1.2	1.5
150	-5.8	-5.8	-6.4	-6.1	-6.9	-7.6	-5.0	3.2	11.5	7.8	4.2	2.5
100	-4.2	-5.0	-5.6	-5.8	-6.1	-6.5	-4.8	1.1	7.8	10.6	4.5	3.0
70	-2.7	-2.5	-1.8	-2.2	-2.2	-2.1	-1.5	1.2	4.2	4.5	5.0	3.0
50	-0.9	-0.1	-0.1	-1.0	-1.0	-0.8	-0.4	1.5	2.5	3.0	3.0	3.7

VERTICAL COVARIANCE KERNEL: adapted covariance matrix

	1000	850	700	500	400	300	250	200	150	100	70	50
1000	82.5	42.6	29.7	17.7	15.1	10.5	-1.5	-23.7	-15.5	-3.6	-0.9	4.0
850	42.6	30.0	20.9	12.4	11.7	9.8	1.7	-13.5	-11.5	-4.5	-0.6	3.6
700	29.7	20.9	20.6	12.7	12.5	11.3	3.4	-11.0	-10.8	-5.0	0.6	3.8
500	17.7	12.4	12.7	12.1	11.7	11.8	7.5	-1.9	-8.5	-7.3	-2.0	0.0
400	15.1	11.7	12.5	11.7	12.8	13.2	8.9	-0.6	-9.1	-7.9	-2.0	-0.1
300	10.5	9.8	11.3	11.8	13.2	16.4	13.6	5.8	-9.4	-10.0	-3.2	-1.2
250	-1.5	1.7	3.4	7.5	8.9	13.6	17.5	17.5	-5.2	-10.9	-5.4	-4.2
200	-23.7	-13.5	-11.0	-1.9	-0.6	5.8	17.5	36.0	5.4	-8.3	-6.8	-7.4
150	-15.5	-11.5	-10.8	-8.5	-9.1	-9.4	-5.2	5.4	12.5	6.9	2.6	0.3
100	-3.6	-4.5	-5.0	-7.3	-7.9	-10.0	-10.9	-8.3	6.9	13.4	6.6	5.1
70	-0.9	-0.6	0.6	-2.0	-2.0	-3.2	-5.4	-6.8	2.6	6.6	7.7	5.9
50	4.0	3.6	3.8	0.0	-0.1	-1.2	-4.2	-7.4	0.3	5.1	5.9	7.0

Table VII.5.3: Historic and sample covariance sample matrices at the 12 NMC levels for summer case

explain this surprising result: (1) the bias correction has small entries, (2) this correction is constant at all latitudes and most likely explanation, (3) the $\alpha - \beta$ corrections scheme for the brightness temperature is not good enough to correct for small errors in the radiative transfer, ground reflectivity and non-linearities. The improvement in both experiments is less than 0.1 K which is not significant given

VERTICAL COVARIANCE KERNEL: historic covariance matrix

	1000	850	700	500	400	300	250	200	150	100	70	50
1000	1.0	0.7	0.5	0.5	0.4	0.4	0.3	0.0	-0.3	-0.2	-0.2	-0.1
850	0.7	1.0	0.7	0.6	0.6	0.6	0.4	-0.1	-0.5	-0.4	-0.3	0.0
700	0.5	0.7	1.0	0.8	0.8	0.7	0.5	-0.2	-0.6	-0.5	-0.2	0.0
500	0.5	0.6	0.8	1.0	0.9	0.8	0.6	-0.2	-0.6	-0.6	-0.4	-0.2
400	0.4	0.6	0.8	0.9	1.0	0.9	0.7	-0.2	-0.7	-0.6	-0.3	-0.2
300	0.4	0.6	0.7	0.8	0.9	1.0	0.8	0.0	-0.7	-0.6	-0.3	-0.1
250	0.3	0.4	0.5	0.6	0.7	0.8	1.0	0.4	-0.5	-0.5	-0.2	-0.1
200	0.0	-0.1	-0.2	-0.2	-0.2	0.0	0.4	1.0	0.3	0.1	0.2	0.2
150	-0.3	-0.5	-0.6	-0.6	-0.7	-0.7	-0.5	0.3	1.0	0.7	0.6	0.4
100	-0.2	-0.4	-0.5	-0.6	-0.6	-0.6	-0.5	0.1	0.7	1.0	0.6	0.5
70	-0.2	-0.3	-0.2	-0.4	-0.3	-0.3	-0.2	0.2	0.6	0.6	1.0	0.7
50	-0.1	0.0	0.0	-0.2	-0.2	-0.1	-0.1	0.2	0.4	0.5	0.7	1.0

VERTICAL COVARIANCE KERNEL: approximate sample covariance matrix

	1000	850	700	500	400	300	250	200	150	100	70	50
1000	1.0	0.9	0.6	0.4	0.4	0.3	0.1	-0.1	-0.2	-0.2	-0.2	-0.1
850	0.9	1.0	0.9	0.7	0.6	0.5	0.2	-0.2	-0.3	-0.3	-0.3	-0.1
700	0.6	0.9	1.0	0.8	0.8	0.5	0.1	-0.2	-0.4	-0.4	-0.3	-0.1
500	0.4	0.7	0.8	1.0	0.9	0.7	0.2	-0.2	-0.5	-0.5	-0.4	-0.1
400	0.4	0.6	0.8	0.9	1.0	0.8	0.3	-0.1	-0.4	-0.5	-0.4	-0.2
300	0.3	0.5	0.5	0.7	0.8	1.0	0.7	0.2	-0.2	-0.5	-0.4	-0.2
250	0.1	0.2	0.1	0.2	0.3	0.7	1.0	0.8	0.3	-0.2	-0.2	-0.2
200	-0.1	-0.	-0.2	-0.2	-0.1	0.2	0.8	1.0	0.8	0.2	0.1	0.0
150	-0.2	-0.3	-0.4	-0.5	-0.4	-0.2	0.3	0.8	1.0	0.7	0.5	0.3
100	-0.2	-0.3	-0.4	-0.5	-0.5	-0.5	-0.2	0.2	0.7	1.0	0.8	0.5
70	-0.2	-0.3	-0.3	-0.4	-0.4	-0.4	-0.2	0.1	0.5	0.8	1.0	0.8
50	-0.1	-0.1	-0.1	-0.1	-0.2	-0.2	-0.2	0.0	0.3	0.5	0.8	1.0

Table VII.5.4: Historic and sample normalized covariance matrices at the 12 NMC levels for summer case

the accuracy of the ground truth. Close to the ground, however, the improvement increases up to 0.2 K. Note that the rms error increase slightly at 200 and 700 *mbar*, a heuristic confirmation that adaptation is not always guaranteed to improve retrieval error. Table VII.5.7 considers the same experiments but when the sounding angle is 56°. The impact of adaptation is more important at 56° than at nadir.

VERTICAL COVARIANCE KERNEL: adapted covariance matrix

	1000	850	700	500	400	300	250	200	150	100	70	50
1000	1.0	0.9	0.7	0.6	0.5	0.3	0.0	-0.4	-0.5	-0.1	0.0	0.2
850	0.9	1.0	0.8	0.7	0.6	0.4	0.1	-0.4	-0.6	-0.2	0.0	0.2
700	0.7	0.8	1.0	0.8	0.8	0.6	0.2	-0.4	-0.7	-0.3	0.0	0.3
500	0.6	0.7	0.8	1.0	0.9	0.8	0.5	-0.1	-0.7	-0.6	-0.2	0.0
400	0.5	0.6	0.8	0.9	1.0	0.9	0.6	0.0	-0.7	-0.6	-0.2	0.0
300	0.3	0.4	0.6	0.8	0.9	1.0	0.8	0.2	-0.7	-0.7	-0.3	-0.1
250	0.0	0.1	0.2	0.5	0.6	0.8	1.0	0.7	-0.4	-0.7	-0.5	-0.4
200	-0.4	-0.4	-0.4	-0.1	0.0	0.2	0.7	1.0	0.3	-0.4	-0.4	-0.5
150	-0.5	-0.6	-0.7	-0.7	-0.7	-0.7	-0.4	0.3	1.0	0.5	0.3	0.0
100	-0.1	-0.2	-0.3	-0.6	-0.6	-0.7	-0.7	-0.4	0.5	1.0	0.7	0.5
70	0.0	0.0	0.0	-0.2	-0.2	-0.3	-0.5	-0.4	0.3	0.7	1.0	0.8
50	0.2	0.2	0.3	0.0	0.0	-0.1	-0.4	-0.5	0.0	0.5	0.8	1.0

VERTICAL COVARIANCE KERNEL: approximate sample covariance matrix

	1000	850	700	500	400	300	250	200	150	100	70	50
1000	1.0	0.9	0.6	0.4	0.4	0.3	0.1	-0.1	-0.2	-0.2	-0.2	-0.1
850	0.9	1.0	0.9	0.7	0.6	0.5	0.2	-0.2	-0.3	-0.3	-0.3	-0.1
700	0.6	0.9	1.0	0.8	0.8	0.5	0.1	-0.2	-0.4	-0.4	-0.3	-0.1
500	0.4	0.7	0.8	1.0	0.9	0.7	0.2	-0.2	-0.5	-0.5	-0.4	-0.1
400	0.4	0.6	0.8	0.9	1.0	0.8	0.3	-0.1	-0.4	-0.5	-0.4	-0.2
300	0.3	0.5	0.5	0.7	0.8	1.0	0.7	0.2	-0.2	-0.5	-0.4	-0.2
250	0.1	0.2	0.1	0.2	0.3	0.7	1.0	0.8	0.3	-0.2	-0.2	-0.2
200	-0.1	-0.	-0.2	-0.2	-0.1	0.2	0.8	1.0	0.8	0.2	0.1	0.0
150	-0.2	-0.3	-0.4	-0.5	-0.4	-0.2	0.3	0.8	1.0	0.7	0.5	0.3
100	-0.2	-0.3	-0.4	-0.5	-0.5	-0.5	-0.2	0.2	0.7	1.0	0.8	0.5
70	-0.2	-0.3	-0.3	-0.4	-0.4	-0.4	-0.2	0.1	0.5	0.8	1.0	0.8
50	-0.1	-0.1	-0.1	-0.1	-0.2	-0.2	-0.2	0.0	0.3	0.5	0.8	1.0

Table VII.5.5: Adapted and sample normalized covariance matrices at the 12 NMC levels for summer case

Compare for instance the 1000 *mbar* pressure level where the reduction in rms error is 1.2 K as compared to 0.2 K. The ground is somewhat of a limiting case because, close to the ground, the statistics (and errors in it) play a more important role than they do at nadir.

For 56°, the influence of ground on the weighting functions is greatly re-

July 2,3,4,5 and 6 1979

latitude between 34 and 54, longitude between -135 and -75

latitude between 35 and 55, longitude between -10 and 30

statistics compiled using 293 retrievals, sounding angle nadir

pressure	bias	rms	std	bias	rms	std	bias	rms	std
1000 mb	-2.7	4.0	3.0	-2.7	3.9	2.8	-2.6	3.8	2.8
850 mb	-0.7	1.5	1.3	-0.6	1.4	1.3	-0.6	1.4	1.3
700 mb	-0.1	1.1	1.1	0.0	1.2	1.2	-0.1	1.2	1.2
500 mb	0.9	1.8	1.6	0.9	1.7	1.5	0.9	1.7	1.5
400 mb	1.7	2.4	1.7	1.7	2.3	1.6	1.7	2.3	1.6
300 mb	1.7	2.3	1.6	1.7	2.2	1.5	1.6	2.2	1.5
250 mb	1.3	2.0	1.5	1.3	1.9	1.5	1.2	1.9	1.5
200 mb	-1.5	2.5	2.0	-1.7	2.6	2.0	-1.7	2.6	2.0
150 mb	-2.2	3.4	2.6	-2.3	3.3	2.3	-2.2	3.2	2.3
100 mb	-1.1	2.3	2.0	-1.0	2.2	1.9	-1.0	2.2	1.9
70 mb	-0.1	0.7	0.7	-0.1	0.7	0.6	-0.1	0.7	0.6
50 mb	1.9	3.0	2.2	2.0	2.9	2.1	1.9	2.9	2.1
	no adaptation			adaptation of mean			adaptation of mean and covariance		

Table VII.5.6: Adaptation in summer, 2 cases: first moment adaptation and first and second moments adaptation. Sounding angle is nadir

duced compared to nadir. Hence the contribution of the surface emission to the brightness temperature covariance matrix (the input to the recursion scheme) is reduced. This makes the adaptation algorithm more effective as is illustrated by the larger retrieval error's reduction. From 850 to 100 *mbar*, the improvement varies from 0.1 to 0.5 K.

For the adaptation to work at nadir or in situations where the ground contribution is important (such as over ocean), it is necessary to determine the ground contribution by means other than through the temperature sounding channels, using, for instance transparent microwave channels or infrared measurements.

July 2,3,4,5 and 6 1979

latitude between 34 and 54, longitude between -135 and -75

latitude between 35 and 55, longitude between -10 and 30

statistics compiled using 586 retrievals, sounding angle 56.24

pressure	bias	rms	std	bias	rms	std	bias	rms	std
1000 mb	-5.5	7.8	5.5	-5.1	7.6	5.5	-4.5	6.6	4.9
850 mb	-2.6	4.2	3.3	-2.5	4.1	3.3	-2.2	3.9	3.2
700 mb	-1.8	2.5	1.7	-1.7	2.4	1.7	-1.6	2.2	1.5
500 mb	0.3	1.1	1.0	0.3	1.1	1.0	0.3	1.1	1.0
400 mb	0.9	1.5	1.1	0.9	1.5	1.1	0.9	1.5	1.1
300 mb	1.4	2.1	1.6	1.4	2.1	1.6	1.4	2.1	1.6
250 mb	2.2	2.9	1.9	2.2	2.9	1.9	2.1	2.7	1.8
200 mb	0.8	2.5	2.4	0.8	2.5	2.4	0.7	2.0	1.9
150 mb	-1.4	2.3	1.9	-1.4	2.3	1.9	-1.4	2.2	1.7
100 mb	-1.5	2.7	2.2	-1.4	2.6	2.2	-1.4	2.4	1.9
70 mb	-0.8	1.2	0.9	-0.8	1.2	0.9	-0.8	1.2	0.9
50 mb	1.1	1.6	1.2	1.1	1.6	1.2	1.1	1.6	1.2
	no adaptation			adaptation of mean			adaptation of mean and covariance		

Table VII.5.7: Adaptation in summer, 2 cases: first moment adaptation and first and second moments adaptation. Sounding angle is 56°

The replacement of the mean profile on a latitudinal basis proved to be the most effective way to adapt statistics for the MSU instrument. As stated the a-priori mean vector is a function of latitude which is usually obtained by interpolating linearly in time January and July atmosphere's, and linearly in space between tropical, subtropical, temperate, and polar atmosphere. This interpolation is performed for every degree, thus yielding a mean matrix of 181×33 elements.

The replacement technique was used in parallel with the standard second moment adaptation in two different ways: in the first implementation the mean profile at a given latitude was replaced by the estimated temperature profile provided that the latter's latitude was less than two degrees away. This replacement

retrievals performed October 10 to 19 1979
Latitude between 34 and 54, longitude between -135 and -75
Latitude between 35 and 55, longitude between -10 and 30
Statistics compiled using all sounding angles and 9229 retrievals
a-priori statistics: July 1979
historic statistics: Peoria's summer

pressure	bias	rms	std	bias	rms	std	bias	rms	std
1000 mb	-2.0	7.4	7.1	-2.6	8.9	8.5	-1.1	5.6	5.4
850 mb	0.4	5.6	5.6	0.7	6.4	6.3	-1.1	4.5	4.4
700 mb	-0.2	5.0	5.0	0.0	5.5	5.5	-1.1	4.4	4.3
500 mb	-1.3	4.5	4.3	-1.4	4.5	4.3	-1.0	4.2	4.1
400 mb	-1.2	4.6	4.4	-1.2	4.6	4.5	-0.7	4.2	4.2
300 mb	-1.9	5.7	5.4	-2.1	6.4	6.1	-0.7	4.3	4.2
250 mb	-1.2	4.0	3.8	-1.3	4.2	3.9	-0.9	3.5	3.4
200 mb	-0.4	3.5	3.5	-0.2	3.9	3.9	-1.2	3.1	2.8
150 mb	0.8	4.4	4.4	0.9	4.5	4.4	0.4	3.8	3.8
100 mb	0.9	4.1	4.1	0.9	4.3	4.2	0.9	3.4	3.3
70 mb	1.0	3.1	2.9	1.1	3.2	3.0	0.7	2.5	2.4
50 mb	1.0	2.8	2.6	1.0	2.7	2.5	1.0	2.5	2.4
	no adaptation			global adaptation			segmented adaptation		

Table VII.5.8: Adaptation through mean replacement for October case over Europe and the USA:a-priori statistics summer

was performed as soon as one frame of temperature profile was estimated. In the second implementation, this replacement was segmented in the sense that the mean is replaced not only according to latitude but to retrieval regions. For instance, the mean vector for Europe will not be replaced by temperature profiles estimated over the United States and *vice versa*. Hence different mean vectors are used for different regions, and their evolutions are independent (from an implementation point of view). Likewise the adaptation of the vertical covariance kernel is performed separately for the two regions, thus yielding different retrieval operators.

Tables VII.5.8 and VII.5.9 consider this adaptation through replacement schemes. In both case, the retrievals are performed during October 1979 over the

retrievals performed October 10 to 19 1979
Latitude between 34 and 54, longitude between -135 and -75
Latitude between 35 and 55, longitude between -10 and 30
Statistics compiled using all sounding angles and 9229 retrievals
a-priori statistics: January 1979 USA
historic statistics: winter USA

pressure	bias	rms	std	bias	rms	std	bias	rms	std
1000 mb	-1.4	6.9	6.8	-2.2	8.1	7.8	-1.4	5.2	5.0
850 mb	0.7	5.1	5.0	1.9	7.3	7.1	-1.6	6.0	5.8
700 mb	-0.9	4.3	4.2	-1.0	4.5	4.4	-0.9	4.5	4.4
500 mb	-1.9	5.9	5.6	-2.1	6.1	5.7	-0.8	4.5	4.4
400 mb	-2.1	6.3	6.0	-2.5	6.9	6.4	-0.5	4.1	4.1
300 mb	-0.9	3.9	3.8	-0.8	3.9	3.8	-0.4	3.1	3.1
250 mb	0.9	3.8	3.7	1.5	5.2	5.0	-0.5	2.1	2.0
200 mb	0.2	2.9	2.9	0.6	3.1	3.1	-1.1	3.1	2.9
150 mb	-0.6	2.3	2.2	-0.9	2.9	2.7	-0.5	2.1	2.0
100 mb	0.7	2.7	2.7	0.8	2.8	2.7	0.6	2.5	2.4
70 mb	0.9	4.1	4.0	0.7	4.1	4.1	1.2	3.7	3.5
50 mb	1.5	3.6	3.3	1.7	3.9	3.6	1.2	3.1	2.9
	no adaptation			global adaptation			segmented adaptation		

Table VII.5.9: Adaptation through mean replacement for October case over Europe and the USA:a-priori statistics winter

USA and Europe: the a-priori mean was computed as described above. interpolation. The historic and climatological covariance matrices used are characteristic of either winter or summer climates. The global replacement of the mean profile degrade the retrieval error which shows that it is not an appropriate technique. However, when the means are replaced independently over each retrieval regions, the improvement was substantial. The improvement is usually greater than the one encountered using the constant bias correction scheme over the summer case. This is true for two principal reasons:

(1) The historic climatologies used are different from the sample climatologies more for the October implementation than for the summer implementation.

(2) The on-line replacement of the mean vector is related to the Kalman filtering of temperature field where the mean vector is updated though the Kalman equation (see Appendix V.A). In this thesis' implementation, the a-priori mean vector is replaced by the estimated profile rather than being an average between a-priori and estimated profiles. Also, the indexing used for the replacement is not along the track of the satellite but along the latitude axis. Otherwise, the adaptation though replacement is equivalent (for the propagation of the mean vector) to a Kalman filter with the identity as a transition matrix. As for the deviation of the new temperature profiles from the a-posteriori mean, it relies on the two-dimensional retrieval operator perpendicular to the satellite's track which uses the lifting of the weighting functions as well as the horizontal correlation to reduce rms error.

As noted, the proposed adaptation scheme can also be used to extrapolate a sample covariance matrix from a few available pressure levels to a more adequate representation. Figures VII.5.3 and VII.5.4 present the impact of this extrapolation. In this simulation, Peoria's 100 summer profiles are assumed to be the sample statistics. Two D matrices (single-spot) are used to retrieve the temperature: One is based on the correct 33 levels description and the other only on the 14 levels NMC mandatory levels. The resulting rms errors (assuming a null bias) is presented in Figure VII.5.3. The difference between both curves is small in the lower stratosphere because it is a stable region where temperatures are strongly correlated over long distances. Figure VII.5.4 superposes on the same plot the rms error associated with the D matrix found by extrapolating the 14 levels using the 33×33 covariance matrix of summer over the USA as a climatological matrix. Except for p=250 *mbar*, the error is reduced significantly which proves that this method is appropriate for MSU with limited statistics.

VII.5.2 Advanced Microwave Sounding Unit

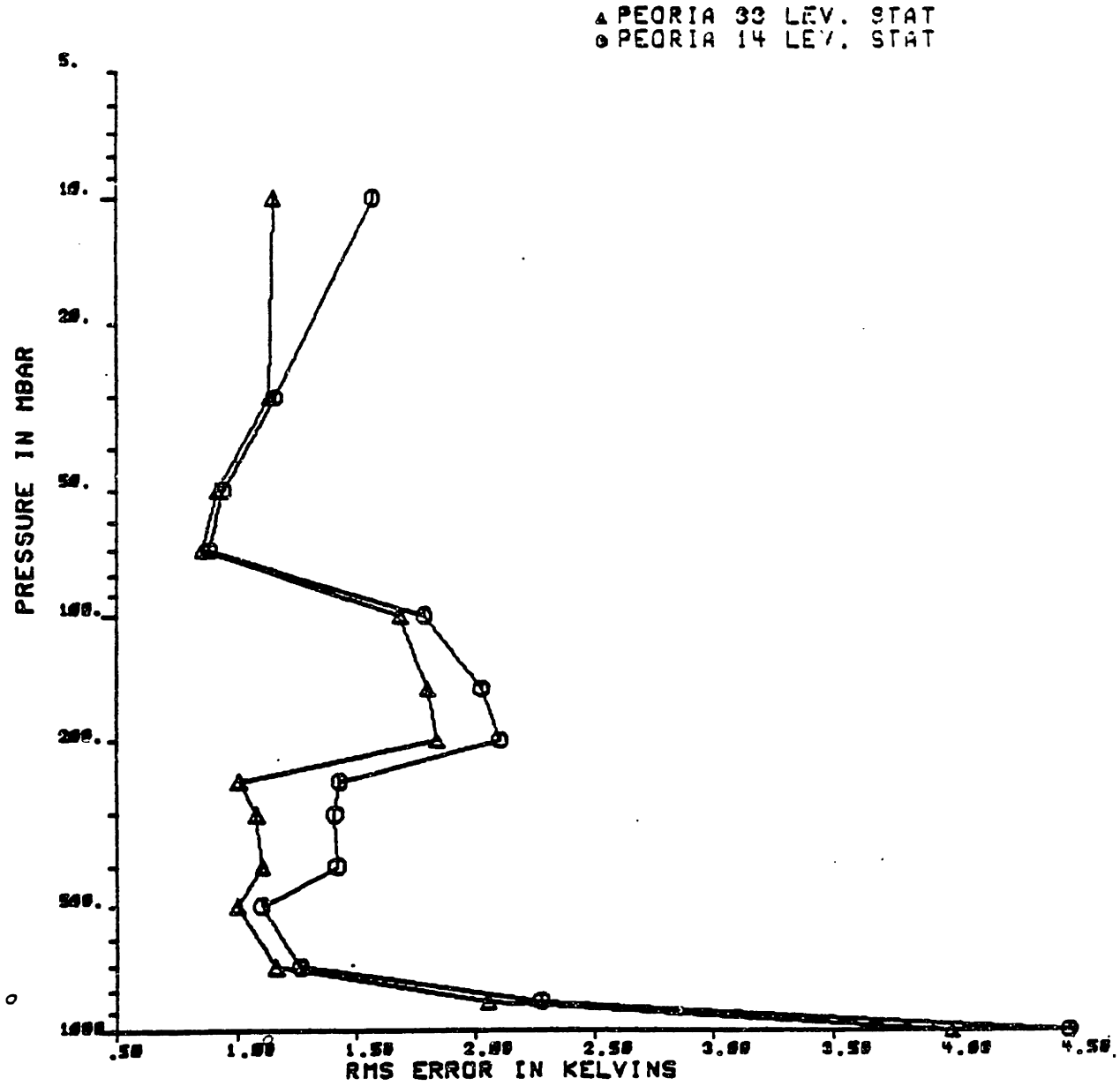


Figure VII.5.3: Expected rms error for optimal operator derived from 14 and 33 level vertical covariance matrices

As could be expected, because the AMSU observable subspace captures most of the temperature profile's energy, the adaptation works much better with this instrument than it did with MSU. For the same reason, multidimensional retrieval operators are very narrow in the spatial domain. Hence, the present analysis will be concentrated on the single-spot problem.

Figures VII.5.5, VII.5.6 and VII.5.7 presents three different cases of adap-

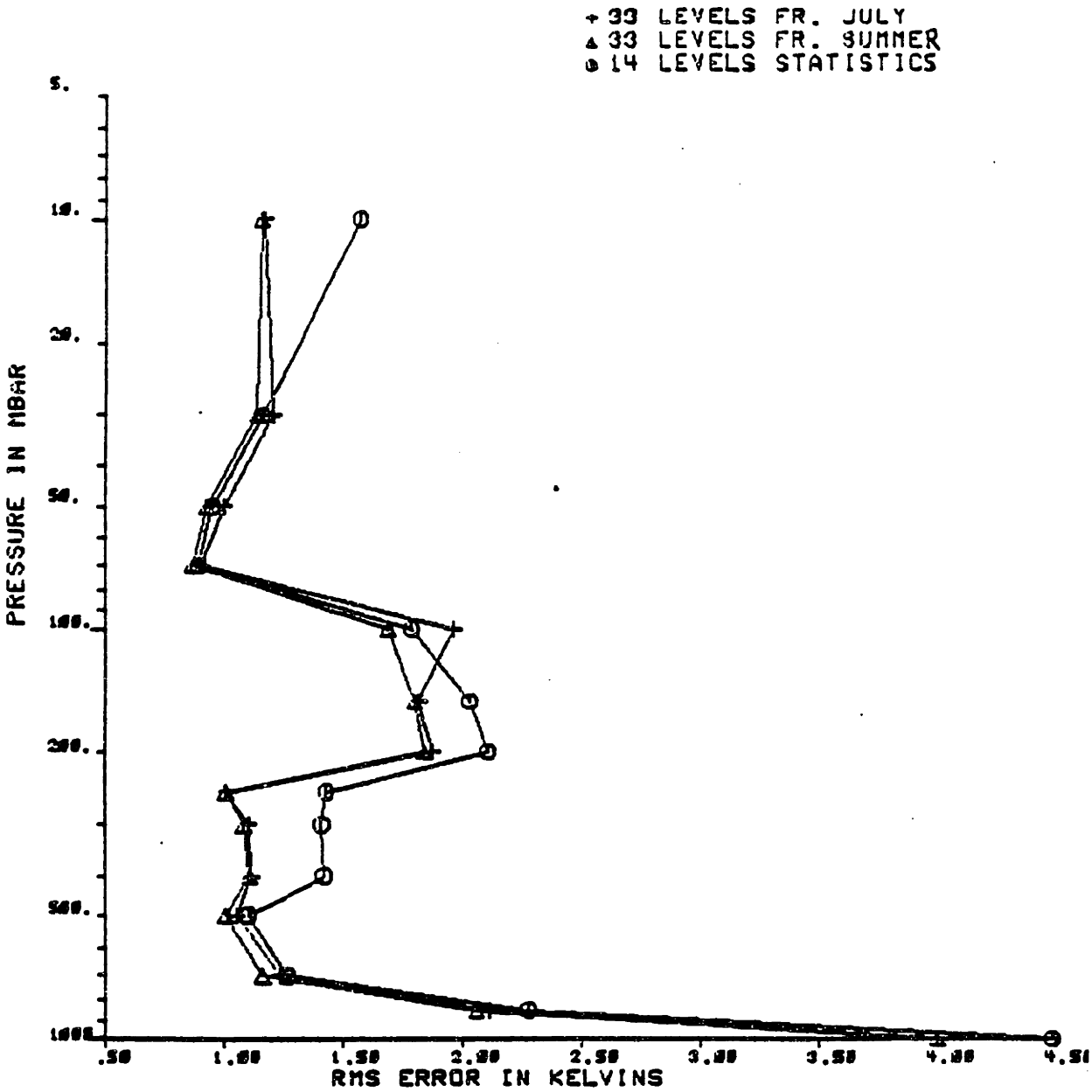


Figure VII.5.4: Expected rms error for optimal operator derived from 14 and 33 level vertical covariance matrices superposed with rms error for operator based on extrapolated 14 level covariance

tation. The differences with MSU are striking.

Figure VII.5.5 considers a case where only two covariance matrices are available: The historic covariance (equal to the climatological one) corresponds to Pt-Mugu (CA) in June whereas the sample statistics are for the brightness temperature over Peoria in summer. This adaptation correspond to the case where

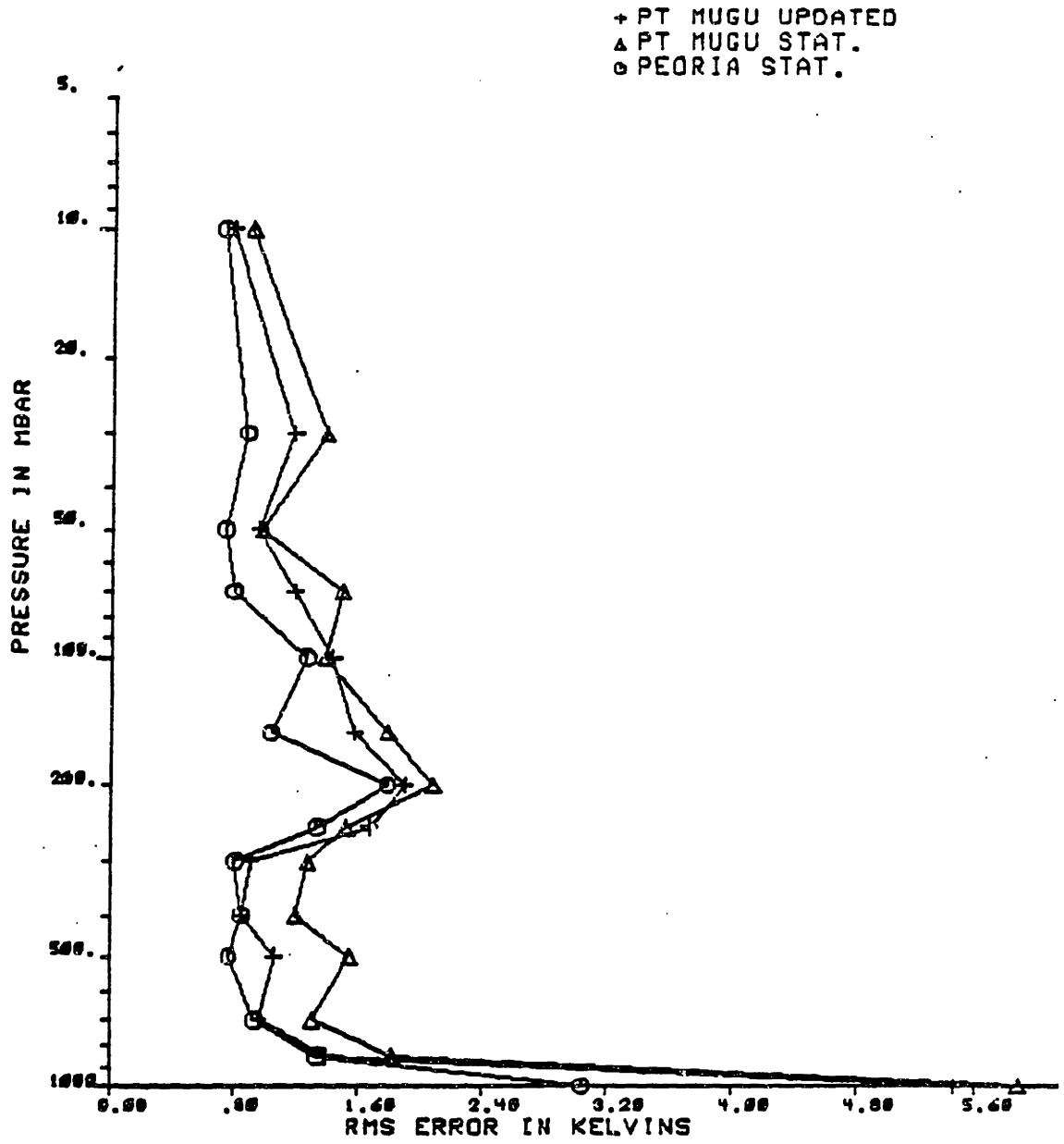


Figure VII.5.5: Adaptation of second moment statistics for AMSU:a-priori as well as historic statistics is from June at Pt-Mugu, sample statistics is summer at Peoria.

the retrieval operator for the observable components is the same for the adapted operator as for the optimal operator, whereas the operator for the unobservable

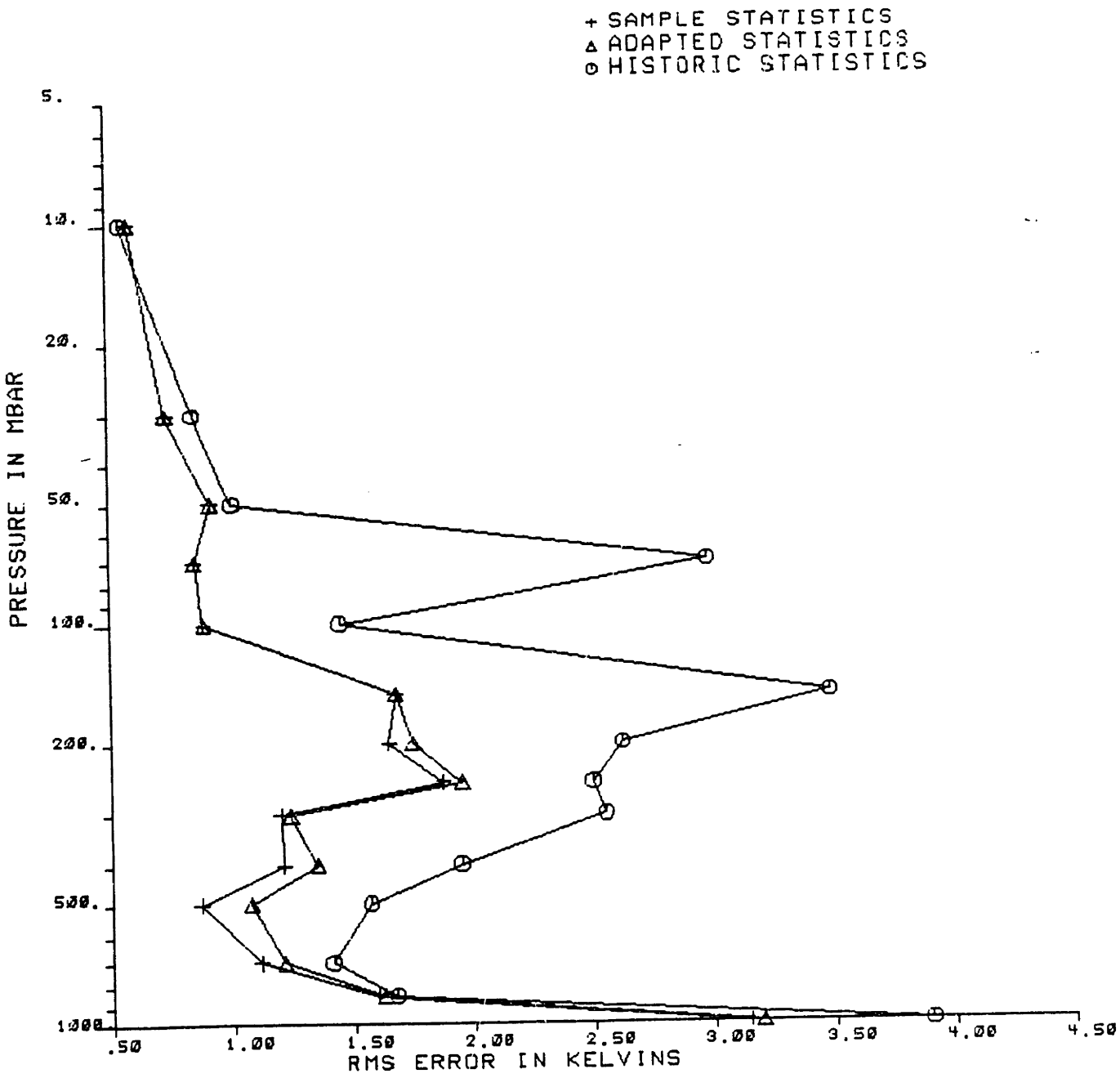


Figure VII.5.6: Adaptation of second moment statistics for AMSU:a-priori statistics is winter USA, climatological statistics summer USA, and sample statistics summer Peoria.

components remains unchanged through the adaptation. The rms error with the exception of $p = 250mbar$ is reduced through adaptation, often substantially.

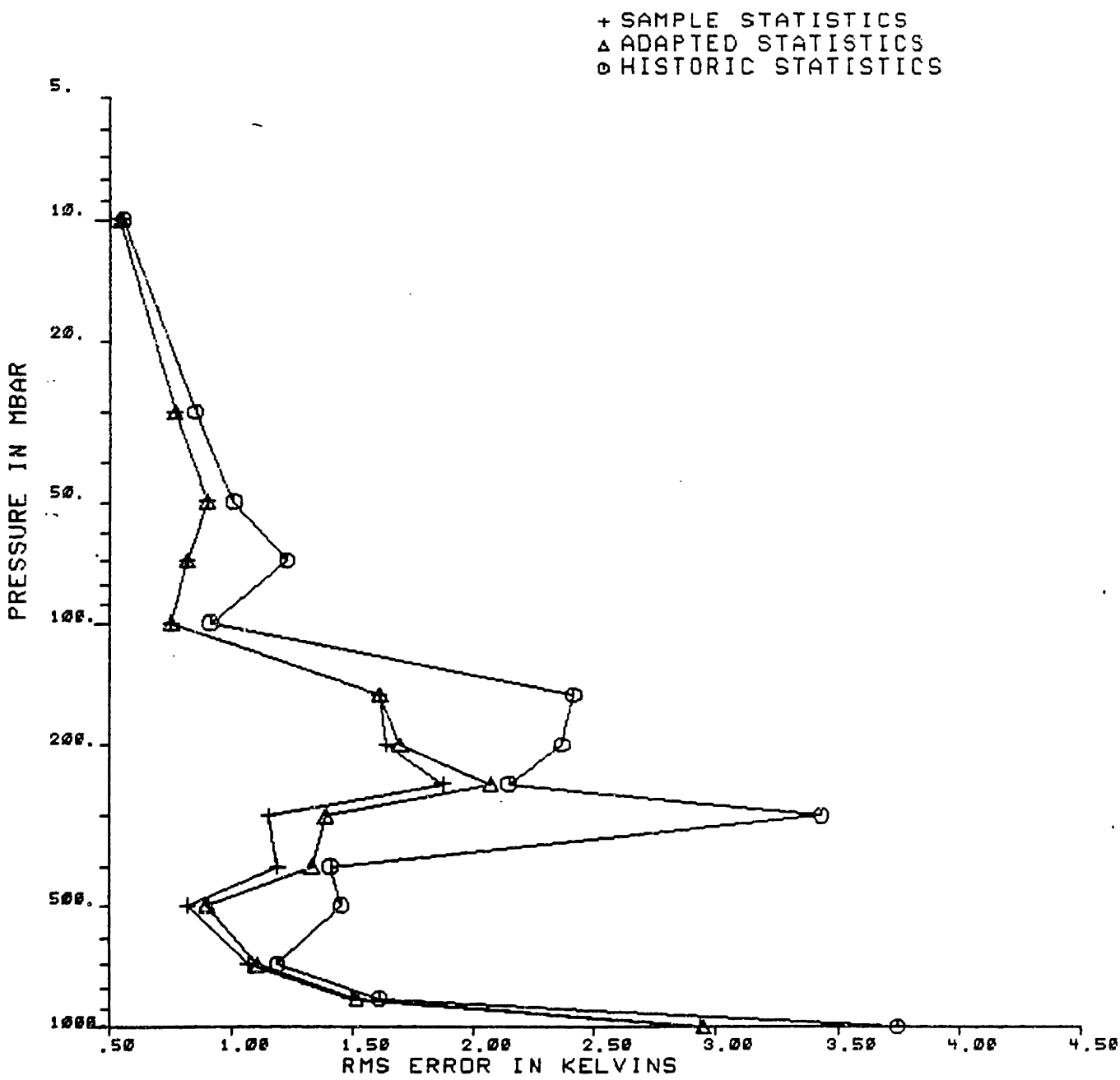


Figure VII.5.7: Adaptation of second moment statistics for AMSU: a-priori is January at Papas, sample statistics is June at Peoria, Fisher framework

Figure VII.5.6 exhibits the rms error for a case where three covariances matrices are available: the historical statistics corresponding to the winter over the

USA, the climatological statistics corresponding to the summer over the USA, and the brightness temperature covariance of June over Peoria. The adapted statistics again does as well as the optimal one. The impact of introducing a climatological covariance close to the sample covariance is evident. At high altitudes where the differences between the two summer covariance matrices are small, it impossible to distinguish between adapted and optimal operators.

The ultimate test is presented in Figure VII.5. where the adaptation from January over the western part of the Pacific Ocean to June over Peoria is performed with no climatological statistics, that is within the Fisher framework. Except at and below the tropopause, the adapted D matrix performs as well as the sample D matrix.

VII.6 Conclusions

In this chapter, the impact of using wrong and inaccurate first and second moments to derive retrieval operators was analytically and heuristically quantified. Because, the optimal adaptation operators are not computable, two suboptimal families of solutions (Fisher and Bayesian) to the adaptation problem were derived and characterized.

A measure of specificity for vertical covariance matrices was introduced. The measure of sensitivity of the optimal retrieval errors to these covariance matrices was developed, and quantified. Both measures can be used to characterize climatologies and datasets of temperature profiles to determine how pathological they are and how robust the associated retrieval errors are.

The derived adaptation operators were then implemented on simulated and real data. Because of its limited sounding capabilities, adaptation does not always work for MSU (as measured by a reduction of rms error for each dataset). Adaptation works better through reduction of the bias error, when the bias error is cor-

rected locally. Adaption works in general better for large sounding angles, where the ground contribution to the measurements is lessened.

For AMSU, which is characterized by a large obvervable subspace, the reduction of error through adaptation comes from a better representation of the second moment characteristics. Although not tested because of the lack of data, the reduction of bias error will also play an important role in the reduction of the overall error. Adaptation works for AMSU and should be used whenever possible.

Combining the Bayesian adaptation with a segmented replacement of the a-priori mean profile (or by a weighted average of the old mean with the estimated profile) appears as a good adaptation scheme for systems with few sounding channels. This result shows that for such systems, keeping the a-priori field (the baseline field) as close as possible from the true temperature field is important. This results also confirms the correctness of present assimilations techniques which use NWP-based temperature fields as a-priori guesses.

Finally, the more segmented the replacement of the a-priori mean is, the better the adaptation for the mean performs. This indicates that the a-priori mean profile should in the future be both latitude- and longitude-dependent, and that any correction to this profile should be localized.

Appendix VII.A: Some Expectations

In this appendix the mathematical derivations required to find the optimal Bayesian adaptor are derived.

To reduce the impact of misspecifying the covariance matrix for retrieval operators consistent with observations, the criterion to minimize is

$$e = trace\{H(CWSW^tC^t - S)^2\} \qquad (VII.A.1)$$

where the matrix S is a sample covariance matrix. This sample covariance matrix is characteristic of a specific finite set of realizations of the stochastic temperature profiles. S is typically varying with latitude and time of year to characterize specific air masses. In that sense, it is itself a stochastic process. The expectation of the sample covariance matrix S is R, the true covariance matrix of the underlying stochastic process. R will itself be estimated by compiling realizations of temperature profiles over different types of climatologies.

This Appendix deals with the characterization of the expectation of the above optimization criterion over all possible sample covariance matrices. When computing this expectation, the expected value of S^2 will be considered, i.e, the expected value of the fourth moment of the temperature profile. In order to readily derive an analytical expression for these expectations, one must hypothesize that the temperature profile is a Gaussian variable in order to use moment factoring.

The optimization problem can be reformulated. Let T be a vectorial stochastic process with a Gaussian pdf. Let S denote a sample covariance matrix of this process estimated by summing a series of outer products of temperature profiles all belonging to the same climatology Let R denote the true covariance matrix of the stochastic process. Under these assumptions, what is the expected value of (VII.A.1)?

To simplify the derivation, let us note by $\mathcal{T}$ the transmission matrix associated with the estimation ($\mathcal{T}$ is simply the product CW). The optimization criterion becomes after expansion

$$e = \overline{trace\{H(\mathcal{T}S\mathcal{T}^t\mathcal{T}S\mathcal{T}^t - H\mathcal{T}S\mathcal{T}^tS - HS\mathcal{T}S\mathcal{T}^t + HS^2)\}} \qquad (VII.A.2)$$

Let us consider P_{hh} the first term of this sum and replace S by its outer product counterpart ($S = xx^t$). Since only the trace of the product of matrices matters, one need only to compute the terms of the main diagonal, i.e,

$$P_{hh} = \overline{\sum_{i,k,l,m,n,p} H_{hi}\mathcal{T}_{ik}x_kx_l\mathcal{T}_{ml}\mathcal{T}_{mn}x_nx_p\mathcal{T}_{hp}} \qquad (VII.A.3)$$

Using moment factorization, this product becomes

$$\begin{aligned} P_{hh} = &\sum_{i,k,l,m,n,p} H_{hi}\mathcal{T}_{ik}\overline{x_kx_l}\mathcal{T}_{ml}\mathcal{T}_{mn}\overline{x_nx_p}\mathcal{T}_{hp} \\ &+ \sum_{i,k,l,m,n,p} H_{hi}\mathcal{T}_{ik}\overline{x_kx_n}\mathcal{T}_{mn}\mathcal{T}_{ml}\overline{x_lx_p}\mathcal{T}_{hp} \\ &+ \sum_{i,k,l,m,n,p} H_{hi}\mathcal{T}_{ik}\overline{x_kx_p}\mathcal{T}_{hp}\mathcal{T}_{ml}\overline{x_lx_n}\mathcal{T}_{mn} \end{aligned} \qquad (VII.A.4)$$

Grouping sums to recognize products of matrices yields

$$\begin{aligned} P_{hh} = &\sum_{i,m} H_{hi}(\mathcal{T}R\mathcal{T}^t)_{im}(\mathcal{T}R\mathcal{T}^t)_{mh} \\ &+ \sum_{i,m} H_{hi}(\mathcal{T}R\mathcal{T}^t)_{im}(\mathcal{T}R\mathcal{T}^t)_{mh} \\ &+ \sum_{i,m} H_{hi}(\mathcal{T}R\mathcal{T}^t)_{ih}(\mathcal{T}R\mathcal{T}^t)_{mm} \end{aligned} \qquad (VII.A.5)$$

and summing over all diagonal terms yields

$$P = 2trace\{H\mathcal{T}R\mathcal{T}^t\mathcal{T}R\mathcal{T}^t\} + trace\{H\mathcal{T}R\mathcal{T}^t\}trace\{\mathcal{T}R\mathcal{T}^t\} \qquad (VII.A.6)$$

Using the same type of moment factorization one can find the expectations of the second and third terms of VII.A.2, i.e,

$$\begin{aligned} e =& \overline{trace\{-H\mathcal{T}S\mathcal{T}S^t - HS\mathcal{T}S\mathcal{T}^t\}} \\ =& -trace\{HR\mathcal{T}R\mathcal{T}^t\} + trace\{HR\mathcal{T}^tR\mathcal{T}\} + trace\{HR\mathcal{T}\}trace\{\mathcal{T}R\} \\ &- trace\{H\mathcal{T}R\mathcal{T}^tR\} + trace\{H\mathcal{T}^tR\mathcal{T}R\} + trace\{h\mathcal{T}R\}trace\{R\mathcal{T}\} \end{aligned} \qquad (VII.A.7)$$

The last term is found by taking the expression for the first term and setting the transmission matrix to the identity. Hence,

$$P = 2trace\{HR^2\} + trace\{HR\}trace\{R\} \qquad (VII.A.8)$$

The final step is to regroup these four terms and replace the transmission matrix by its definition:

$$\begin{aligned} e = {} & 2trace\{HCWRW^tC^tCWRW^tC^t\} + trace\{HCWRW^tC^t\}trace\{CWRW^tC^t\} \\ & -trace\{HRCWRW^tC^t\} - trace\{HRW^tC^tRCW\} - trace\{HRCW\}trace\{CWR\} \\ & -trace\{HCWRW^tC^tR\} - trace\{HW^tC^tRCWR\} - trace\{HCWR\}trace\{RCW\} \\ & +2trace\{HR^2\} + trace\{HR\}trace\{R\} \end{aligned} \qquad (VII.A.9)$$

This is the desired result

In this dissertation, the optimal adaptation matrix C is found to equal

$$C = RW^t(WRW^t)^{-1} \qquad (VII.A.10)$$

Plugging the optimal adaptation operator into the error term (equation (VII.A.8)) yields

$$\begin{aligned} e = {} & 2trace\{HRW^t(WRW^t)^{-1}WR^2W^t(WRW^t)^{-1}WR\} \\ & + trace\{HRW^t(WRW^t)^{-1}WR\}trace\{RW^t(WRW^t)^{-1}WR\} \\ & + 2trace\{HR^2\} + trace\{HR\}trace\{R\} \\ & - trace\{HR^2W(WRW^t)^{-1}WR\} \\ & - trace\{HRW^t(WRW^t)^{-1}WR^3W^t(WRW^t)^{-1}W\} \\ & - trace\{HR^2W^t(WRW^t)^{-1}WR^2\}trace\{RW^t(WRW^t)^{-1}WR\} \\ & - trace\{HRW^t(WRW^t)^{-1}WR^2\} \\ & - trace\{HRW^t(WRW^t)^{-1}WR^3W^t(WRW^t)^{-1}WR\} \\ & - trace\{HRW^t(WRW^t)^{-1}WR\}trace\{R^2W^t(WRW^t)^{-1}W\} \end{aligned} \qquad (VII.A.11)$$

Introducing the transfer matrix $\mathcal{T}$ associated with the covariance matrix R (when no sensor is present), the average penalty becomes

$$\begin{aligned} e =& trace\{H\mathcal{T}R\mathcal{T}R\} + 2trace\{HR^2\} + trace\{HR\}trace\{R\} \\ & - 2trace\{HR\mathcal{T}R\} - 2trace\{H\mathcal{T}^t R\mathcal{T}R\} \\ & - trace\{HR\mathcal{T}\}trace\{\mathcal{T}R\} \end{aligned} \qquad (VII.A.12)$$

Appendix VII.B: Optimization of Statistics using the Batthacharyya distance

Adaptation of statistics, at least for the covariance matrix in the case of observation consistent statistics was performed by minimizing the mean square difference between the sample covariance matrix of the set of temperature profiles being retrieved and the matrix used to derive the retrieval operator. This mean square measure is simply the Euclidian distance between the two matrices. One might then wonder if it is possible to use a difference distance measure and obtain a different adaptation scheme. This problem is considered here in the context of the Bhattacharyya distance between random vectors.

The Bhattacharyya distance between two random vectorial processes x_1 and x_2 is given by (Bhattacharyya , 1947, Matusita, 1965)

$$B = -\ln \int [p_1(x)p_2(x)]^{1/2} dx \qquad (VII.B.1)$$

If the two probabilities densities are close to each other, they will be aligned to each other and their direction cosine will be close to one; thus its logarithm is close to zero. As the angle between the pdf approaches a right angle, the cosine tends towards zero, and the distance increases to infinity.

The Bhattacharyya distance is also related to bit rate and probability of error in information theory (Shannon, 1967, Schweppe, 1973). It corresponds to the opposite of the moment factor used in the computation of the Chernoff bound for the classical binary detection theory (Van Trees, 1963).

This measure takes a simpler form for Gaussian probability distributions. Denoting by m_i and R_i the mean and covariance matrix of x_i, the Bhattacharyya distance becomes

$$B = \frac{1}{2}\ln|\frac{1}{2}(R_1 + R_2)| - \ln|R_1| - \ln|R_2| + (m_1 - m_2)^t(R_1 + R_2)^{-1}(m_1 - m_2) \qquad (VII.B.2)$$

When the two statistical distributions have the same mean vector, it becomes

$$4B = \ln \frac{1}{4} + \ln |R_1 + R_2| + \ln |R_1^{-1} + R_2^{-1}| \qquad (VII.B.3)$$

One can now cast the adaptation problem as follows: assuming that the temperature profiles are Gaussian variables, and given both an a-priori (historical) covariance R_a and an incomplete observation of the sample covariance matrix X, what is the matrix closest (in the Bhattacharyya sense) to the a-priori matrix that matches the observed covariance matrix?

The Bhattacharyya distance between a-priori and sample covariance matrix is given by (ignoring irrelevant scaling factors)

$$C(X, R_a) = \ln |R_a + X| + \ln |R_a^{-1} + X^{-1}| \qquad (VII.B.4)$$

The observed covariance matrix for the brightness temperature is given (once the covariance matrix for the sensor noise is removed) by

$$S = WXW^t \qquad (VII.B.5)$$

To solve this constrained estimation problem, let us use the Lagrange multiplier method. The functional to be minimized becomes

$$E(X, R_a, S, W) = \ln |R_a + X| + \ln |R_a^{-1} + X^{-1}| + trace\{\Lambda(WXW^t - S)\} \quad (VII.B.6)$$

where Λ is a full rank square matrix whose size is the number of the observation channels. Moreover, using an alternate expression for the distance measure, the objective can be rewritten as

$$E(X, R_a, S, W) = 2 \ln |R_a + X| - \ln |R_a| - \ln |X| + trace\{\Lambda(WXW^t - S)\} \quad (VII.B.7)$$

The gradient of the objective function with respect to X is found using readily available matrix derivative formulas (Athans and Schweppe, 1963). Namely

$$\frac{\partial E}{\partial X} = 2(R_a + X)^{-1} - X^{-1} + W^t \Lambda W \qquad (VII.B.8)$$

The set of equations to be solved is, thus,

$$\begin{cases} 2(R_a + X)^{-1} - X^{-1} + W\Lambda W^t = 0 \\ WXW^t = S \end{cases} \qquad (VII.B.9)$$

Eliminating the inverse in the first line by multiplication yields

$$\begin{cases} X - R_a + XW\Lambda W^t R_a + XW^t\Lambda X = 0 \\ WXW^t = S \end{cases} \qquad (VII.B.10)$$

To solve for the Lagrange multiplier matrix, let us multiply the first line on the left by W and on the right by its transpose W^t. Hence

$$\begin{cases} X - R_a + XW\Lambda W^t R_a + XW^t\Lambda X = 0 \\ S - S_a + S\Lambda S_a + S\Lambda S = 0 \\ WXW^t = S \end{cases} \qquad (VII.B.11)$$

where S_a is defined as the product WR_aW^t. This yields

$$\begin{cases} \Lambda = S^{-1}(S_a - S)(S_a + S)^{-1} \\ \Lambda = (S^{-1}S_a - I)(S_a + S)^{-1} \\ 0 = X - R_a + RW^t\Lambda WR_a + XW^t\Lambda WX \end{cases} \qquad (VII.B.12)$$

Noting the three matrices

$$\begin{aligned} A &= W^tS^{-1}(S_a - S)(S + S_a)^{-1}W \\ B &= I + W^tS^{-1}(S_a - S)(S + S_a)^{-1}WR_a \\ C &= -R_a \end{aligned} \qquad (VII.B.13)$$

The optimal matrix X is given by the equation

$$XAX + XB + C = 0 \qquad (VII.B.14)$$

The solution to such an quadratic equation is not obvious without some clever manipulations.

The key to the solution of (VII.B.14) is to consider, once more, the singular value representation of the observation matrix and project the equation onto that basis. The SVD decomposition is defined as

$$W = \mathcal{W}_l^t \begin{pmatrix} \Sigma & 0 \end{pmatrix} \mathcal{W}_r \qquad (VII.B.15)$$

Denoting by the subscripts o and u the components in the observable and unobservable subspaces of the weighting matrix, the three matrices defining the quadratic equation become

$$A = \begin{pmatrix} -X_{oo}(X_{oo} - R_{aoo})(X_{oo} + R_{aoo})^{-1} & 0 \\ 0 & 0 \end{pmatrix} \qquad (VII.B.16)$$

Also,

$$B = \begin{pmatrix} I + X_{oo}^{-1}(R_{aoo} + X_{oo})(X_{oo} + R_{aoo})^{-1}R_{aoo} & X_{oo}^{-1}(R_{aoo} - X_{oo})(R_{aoo} + X_{oo})^{-1}R_{aou} \\ 0 & I \end{pmatrix} \qquad (VII.B.17)$$

and

$$C = \begin{pmatrix} -R_{aoo} & -Raou \\ -R_{auo} & -R_{auu} \end{pmatrix} \qquad (VII.B.18)$$

This decomposition allows us to rewrite equation (VII.B.14) into four separate matrix equations. Let us solve those four quadratic matricial equations for the four different submatrices (oo, ou, uo, and uu)

- *covariance of observable components with themselves oo*

$$\begin{aligned}(XAX + XB + C)_{oo} =& - X_{oo}X_{oo}^{-1}(X_{oo} - R_{oo})^{-1}(X_{oo} + R_{oo})^{-1}X_{oo} \\ & - X_{oo}X_{oo}^{-1}(X_{oo} + R_{co})^{-1}R_{oo} - R_{oo} \\ =&(X_{oo} - R_{oo})[I - (X_{oo} + R_{oo})^{-1}(X_{oo} + R_{oo})] \\ =&0\end{aligned} \qquad (VII.B.19)$$

which is thus always satisfied. This is true because of the constraint condition.

- *covariance of observable components with unobservable components uo*

$$\begin{aligned}0 =&X_{uo}X_{oo}(R_{oo} - X_{oo})(X_{oo} + R_{oo})^{-1}X_{oo} + X_{uo} \\ &+ X_{uo}X_{oo}^{-1}(R_{oo} - X_{oo})(X_{oo} + R_{oo})^{-1}R_{oo} - R_{auo} \\ 0 =&X_{uo}X_{oo}^{-1}(R_{oo} - X_{oo}) + X_{uo} - R_{auo} \\ 0 =&X_{uo}X_{oo}^{-1}R_{oo} + X_{uo} + X_{uo} - R_{auo}\end{aligned} \qquad (VII.B.20)$$

which yields, for optimal estimate,

$$X_{ou} = R_{auo}R_{aoo}^{-1}X_{oo} \qquad (VII.B.21)$$

- *covariance of unobservable components with observable components ou*

$$0 = (R_{oo} - X_{oo})(R_{oo} + X_{oo})^{-1}(X_{ou} + R_{aou}) + (X_{ou} - R_{aou}) \qquad (VII.B.22)$$

Using the transpose of the solution (VII.B.21) into (VII.B.14) satisfies the equality, thus the optimal covariance matrix is

$$X_{ou} = X_{oo}R_{aoo}^{-1}R_{auo} \qquad (VII.B.23)$$

- *unobservable components with themselves uu*

The equation becomes

$$\begin{aligned} 0 =& R_{uo}X_{oo}^{-1}(R_{oo} - X_{oo})(X_{oo} + X_{oo})^{-1} + X_{uu} - R_{auu} \\ &+ X_{uo}X_{oo}^{-1}(R_{oo} - X_{oo})(X_{oo} + X_{oo})^{-1}R_{aou} \qquad (VII.B.24) \\ =& R_{auo}[R_{oo}(I - X_{oo}R_{oo}^{-1})]R_{aou} + X_{uu} - R_{auu} \end{aligned}$$

This yields the optimal estimate:

$$X_{uu} = R_{auu} + R_{auo}R_{aoo}^{-1}X_{oo}R_{aoo}^{-1}R_{aou} - R_{auo}R_{aoo}^{-1}R_{aou} \qquad (VII.B.25)$$

Comparing this optimal estimate of the covariance matrix using the Bhattacharyya distance with the ones derived using the adaptation operator in section VII.4.3, it appears that it is the same one that the operator derived using the Bayesian adaptation operator in the case where historical and climatological covariances are identical.

Hence the advantages and shortcomings of the Bayesian adaptation developed in Chapter VII also apply when the criterion used is the Bhattacharyya distance. Because of the connections between the Bhattacharyya distance and detection theory, one can conjecture that the proposed adaptation schemes can be extended to M-ary type detection schemes and that they will be optimal for such detections.

CHAPTER VIII
CONSTRAINED, MIXED ESTIMATION AND OTHER VARIATIONS ON THE LMMSE THEME

"Fifth, Chinese Food should be served hot (unless, of course, it's a cold dish). Like a souffle, it waits for no one." Rhoda Yee, in Chinese Village Cookbook, A Practical Guide to Cantonese Country Cooking.

VIII.1 Introduction

The simplest way to change the form of the LMMSE is to change the criterion it optimizes. This is easier to do than changing the form of the estimator by allowing non linear processing of the observations. This is true because non-linear estimators require knowledge of the higher moments of the pdf which characterizes the temperature profiles. New criteria can be obtained by casting the problem in a new framework or by adding new terms to the minimum mean square error criterion already developed. This latter technique is referred to in this dissertation as objective penalty optimization, because the terms added to the original optimization criterion change the estimated temperature profiles to match a "user-derived" objective. The final estimate is compromise between a desired form and mathematics. A typical example of such a penalty is an integrated measure of the smoothness of the temperature profile. Supplementary information, such as bounds on the range of possible temperatures and/or lapse rates, can also be introduced.

In this chapter, both methods will be considered and when possible, characterized. The intended purpose of this analysis is NOT an exhaustive investigation of these different techniques, but rather a tentative study of new avenues to do remote sensing whenever auxiliary information about the temperature profiles is available. It also attempts to bridge some of the numerous linear methods to perform the inversion of radiances.

VIII.2 Changing the Optimization Criteria

The first change considered involves the recasting of the sensing process as a communication problem between temperature and brightness temperature. This analysis does not produce any new estimator, but is nevertheless relevant because it extends our understanding of remote sensing.

VIII.2.1 An Information theoretic Approach to Remote Sensing

In this section, the sounding and retrieval processes are analyzed from an information theoretic point of view. The purpose of this analysis is to relate the concepts of mutual information, uncertainty and entropy developed in information theory to the concept of mean square error developed in estimation theory. Such a connection has already been drawn for the purpose of coding using the concept of rate distortion theory (Gallager,1968, Davison,1972). However, it has not been applied to remote sensing for the purpose characterizing the performances of specific retrieval systems. Recently, information theory concepts have been advanced (Husain *et al.*, 1986) for optimal objective analysis as an alternative to the minimum mean square criterion introduced by Gandin (1963) and analyzed (partially) in chapter IV. This analysis will start with a brief review of the relevant concepts of information theory, then apply them to the static estimation problem (single spot retrieval).

The most important element in the characterization of the informational content of a remote sensing experiment is the joint probability density function $p(T, T_b)$ between the temperature T and the brightness temperature T_b (the overline signs are not written in the derivation but all the temperatures are vectors). The mutual information between those two variables is defined as (Shannon and Weaver, 1949)

$$\tilde{I}(T;T_b) = \log \frac{p(T,T_b)}{p(T)p(T_b)} \qquad (VIII.2.1)$$

This mutual information depends upon the specific realization of the two stochastic processes and is thus random. Its average value is

$$I(T;T_b) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \log \frac{p(T,T_b)}{p(T)p(T_b)} dT dT_b \qquad (VIII.2.2)$$

In the above equation, the symbols T and T_b refer to the complete distribution of temperature and brightness vectors respectively. The mutual information $I(T;T_b)$ can be thought of as the information transmitted by the sounding process from the temperature profile T to the brightness T_b. It can be optimized as a function of the number of channels and weighting matrices (Pokrosky, 1969). However, it does not fully characterize the retrieval process. To do so, one must introduce the concept of entropy (or uncertainty) of stochastic variables. Although defined rigorously for discrete stochastic processes (Gallager, 1968), entropy is defined as the average of its self information. That is,

$$H(T) = \int_{-\infty}^{\infty} p(T) \log \frac{1}{p(T)} dT \qquad (VIII.2.3)$$

(A similar definition applies to T_b). Conditional entropy is likewise defined as

$$H(T|T_b) = \int_{-\infty}^{\infty} p(T,T_b) \log \frac{1}{p(T|T_b)} dT \qquad (VIII.2.4)$$

The sounding process defines the following relationships

$$\begin{aligned} I(T;T_b) &= H(T) - H(T|T_b) \\ &= H(T_b) - H(T_b|T) \end{aligned} \qquad (VIII.2.5)$$

These equations permit us to interpret $I(T;T_b)$ as the average amount of uncertainty about the temperature profile resolved by the observation of T_b.

The conditional entropy $H(T|T_b)$ characterizes the average remaining uncertainty (ARU) about the temperature after the sounding is performed. In most cases of interest, the number of channels used is greater than the number of degrees of freedom (number of pressure or slabs) in the atmospheric field. Thus, the retrieval

operator or D matrix which operates on the brightness temperature is not full rank. Denoting by $\hat{T}$ the estimate of the temperature ($\hat{T} = DT_b$), the brightness vector can exactly be reconstructed from the estimated profile according to

$$T_b = (D^t D)^{-1} D^t \hat{T} \qquad (VIII.2.6)$$

From an information theoretical point of view, no information is lost or gained by applying the D matrix to T_b as long as it is a full rank matrix. This result is known as the data processing theorem (Gallager, 1968). Note that the ARU is not changed if the matrix D changes. These results demonstrate that information theory does not provide us with an optimal estimator. It characterizes the informational content of the brightness temperature. The conditional entropy $H(T|T_b)$ characterizes the entire retrieval process.

This quantity can be evaluated using the equivalent of a Bayes' rule. Namely,

$$H(T|T_b) = H(T) - H(T_b) + H(T_b|T) \qquad (VIII.2.7)$$

Note that the quantities introduced above have the unit of bits. Let us compute this expression in the context of single spot remote sensing (the concepts and results extend directly to multidimensional cases). The sensing equation is as always

$$T_b = WT + \eta \qquad (VIII.2.8)$$

where η is a zero mean Gaussian noise with covariance N, T is modeled as Gaussian with mean $< T >$ and covariance matrix R. The relevant probability density functions are

$$P(T) = \frac{1}{2\pi^{\frac{L}{2}}} \frac{1}{|R|} \exp\{-\frac{1}{2}(T - < T >)^t R^{-1} (T - < T >)\} \qquad (VIII.2.9)$$

where L is the number of pressure levels in the profile

$$P(T_b|T) = \frac{1}{2\pi^{\frac{M}{2}}} \frac{1}{|N|} \exp\{-\frac{1}{2}(T_b - WT)^t N^{-1} (T_b - WT)\} \qquad (VIII.2.10)$$

(M is the number of channels) and

$$P(T_b) = \frac{1}{2\pi^{\frac{M}{2}}} \frac{1}{|WRW^t + N|} \exp\{-\frac{1}{2}(T_b^t - < T_b >^t)(WRW^t + N)^{-1}(T_b - < T_b >)\} \quad (VIII.2.11)$$

In the above equations, $|.|$ denotes the determinant. Simple integration yields the diverse entropies needed for the computation,

$$\begin{aligned} H(T) &= \log \sqrt{e} 2\pi^{\frac{L}{2}} |R| \\ H(T_b|T) &= \log \sqrt{e} 2\pi^{\frac{M}{2}} |N| \\ H(T_b) &= \log \sqrt{e} 2\pi^{\frac{M}{2}} |WRW^t + N| \end{aligned} \quad (VIII.2.12)$$

The ARU is then

$$H(T|T_b) = \log \sqrt{e} 2\pi^{\frac{L}{2}} + \log |R| + \log |N| - \log |N + WRW^t| \quad (VIII.2.13)$$

This expression is valid only for non-degenerate cases where all covariance matrices are full rank so that the above mentioned Gaussian pdf's are defined. Should the noise covariance become zero, the uncertainty becomes minus infinity. This is true because a linear combination of the components of the temperature profile is measured exactly, when the noise covariance is not full rank. When this is the case, the uncertainty goes to minus infinity because the underlying pdf is assumed continuous, (Gallager,1968). This limit reflects the fact that an infinite number of bits are required to represent such a linear combination. Those readers uncomfortable with this notion should take note that quantum mechanics ensures that no measurement of temperature can be performed in a finite amount of time with infinite accuracy.

Note that the entropy of the random variable T can be expressed as a function of the latent roots ρ_i of R as

$$H(T) = \log \sqrt{e} 2\pi^{\frac{L}{2}} + \sum_i \log \rho_i \quad (VIII.2.14)$$

This expression is different from the one developed in chapter VII and is aimed at characterizing the specificity of a set of temperature profiles from the corresponding covariance kernel.

To gain more insight into the above sensing process and its limitations, consider the singular value decomposition of the weighting matrix (observable and unobservable components). This decomposition is, using previously introduced nomenclature,

$$W = \mathcal{W}_l \begin{pmatrix} \Sigma & 0 \end{pmatrix} \mathcal{W}_r \qquad (VIII.2.15)$$

The covariance matrix for the temperature profile is also projected along the two observation subspaces to yield

$$H(T|T_b) = \log \sqrt{e} 2\pi^{\frac{L}{2}} + \log \left| \begin{pmatrix} R_{oo} & R_{ou} \\ R_{uo} & R_{uu} \end{pmatrix} \right| + log|N| - \log |\mathcal{W}_l^t N \mathcal{W}_l + \Sigma R \Sigma^t| \qquad (VIII.2.16)$$

The determinant of the temperature covariance matrix can be expressed as the product of two determinants. Hence,

$$\begin{aligned} H(T|T_b) = &\log \sqrt{e} 2\pi^{\frac{L}{2}} + \log |R_{oo}| + \log |R_{uu} - R_{uo} R_{oo}^{-1} R_{ou}| \\ &+ log|N| - \log |\mathcal{W}_l^t N \mathcal{W}_l + \Sigma R \Sigma^t| \end{aligned} \qquad (VIII.2.17)$$

Introducing the normalized noise covariance matrix $\tilde{N}$

$$\tilde{N} = \Sigma^{-1} \mathcal{W}_r^t N \mathcal{W}_l \Sigma^{-1} \qquad (VIII.2.18)$$

the remaining uncertainty about the temperature profile after sounding is

$$H(T|T_b) = \log \sqrt{e} 2\pi^{\frac{L}{2}} + \log |R_{oo}| + \log |R_{uu} - R_{uo} R_{oo}^{-1} R_{ou}| + log|\tilde{N}| - \log |\tilde{N} + R| \qquad (VIII.2.19)$$

and after a final factorization

$$\begin{aligned} H(R, W, N) = H(T|T_b) = &\log \sqrt{e} 2\pi^{\frac{L}{2}} + \log |R_{uu} - R_{uo} R_{oo}^{-1} R_{ou}| \\ &+ \log |\tilde{N} (R_{oo} + \tilde{N})^{-1} \tilde{R}_{oo}| \end{aligned} \qquad (VIII.2.20)$$

The amount of information sensed during the sounding is

$$\begin{aligned} I(T; T_b) &= \log |\tilde{N} + R_{oo}| - \log |\tilde{N}| \\ &= \log |1 + R_{oo} \tilde{N}^{-1}| \end{aligned} \qquad (VIII.2.21)$$

This quantity is bounded above by the capacity of the channel associated with the sounding equation.

These two quantities allow analysis of the sensing process and permit the optimization of a sounding unit through the choice of weighting functions (controlled by the operating frequencies) and choice of receiver noise specification.

Note that the information content in the retrieval process is solely characterized by the radiative transfer equation and the first two moments of the temperature profile's pdf. This characterization is somewhat similar to that for mean square estimation in that the lower bound on the retrieval error is directly computable from W, N and R. However, it is more explicit that the mean square case since it does not require the calculation of a retrieval operator.

This information theoretical approach allows one to appreciate and compare different sounding systems and permit the allocation of sounding frequencies. A series of analyses pertaining to the impact of different parameters can be used to illustrate this claim.

The sensor noise's impact can be drastic, especially in cases of good SNR because of the singularity for perfect observations. Analytically, variations of the ARU with respect to the sensor noise level are difficult to track because of the term $\log|R_{oo} + N|$. The variable is a diagonal matrix (in the natural representation). This difficulty is not present in the case for the MSE. However, it is apparent from equation (VIII.2.22) that the ARU becomes independent of the noise covariance when the eigenvalues of N are large compared to those of R_{oo} When this is the case, the observations are useless and the ARU becomes the temperature a-priori uncertainty.

Figure VIII.2.1 presents the variations of the ARU in the case of MSU sounding above land for different sounding angles as a function of noise power. The noise power is normalized to the MSU nominal sensor noise temperature. That is denoting

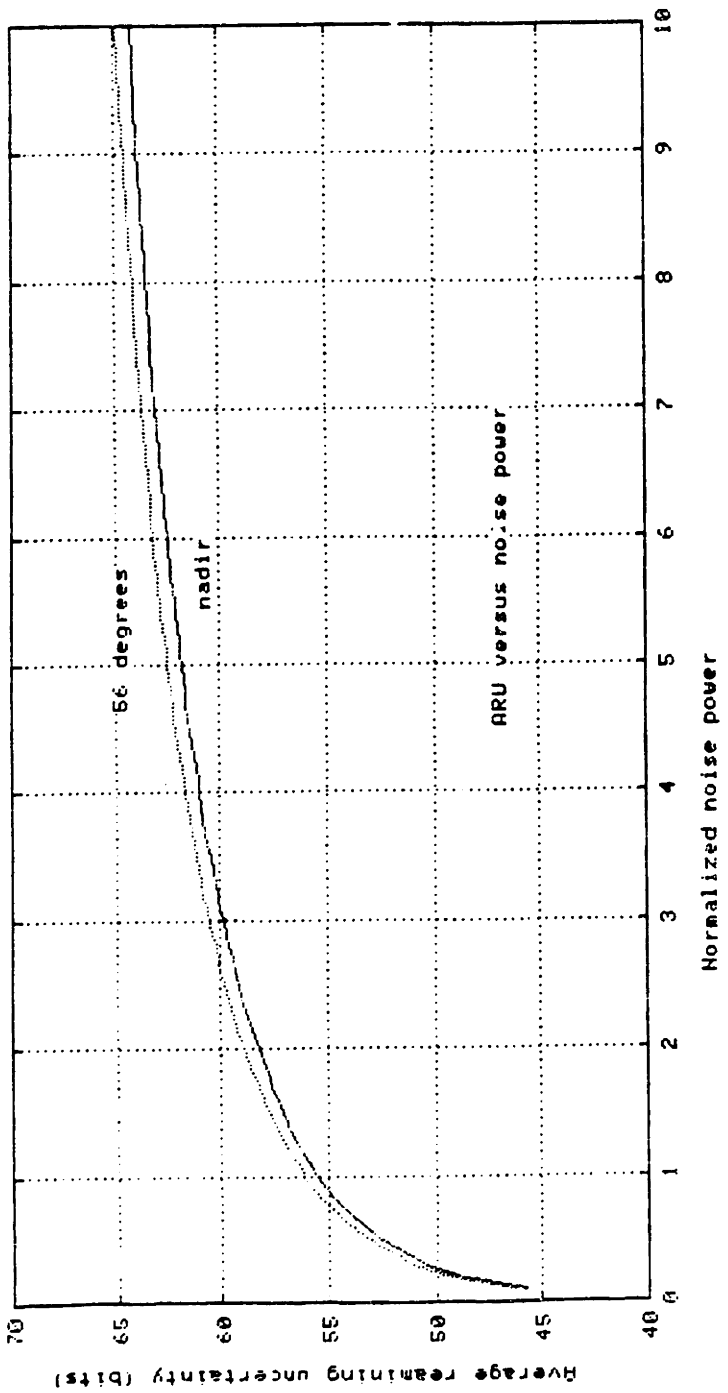


Figure VIII.2.11: ARU for different sounding angles of the MSU system as a function of the noise power normalized to MSU nominal values.

the noise covariance by $N(\epsilon)$,

$$N(\epsilon) = \begin{pmatrix} 0.21^2\epsilon & 0 & 0 \\ 0 & 0.18^2\epsilon & 0 \\ 0 & 0 & 0.26^2\epsilon \end{pmatrix} \qquad (VIII.2.22)$$

The soundings are assumed performed over one hundred temperature profiles characteristic of summer in the northern hemisphere. As could be expected, the ARU follows a somewhat logarithmic curve. The MSU nominal noise level ($\epsilon = 1$) is located at the knee of that curve. This implies the penalty in entropy from increasing the noise level is not as large as the benefits to be gained from reducing it.

As for the mean square error, the ARU depends on the statistical description associated with a specific set of temperature profiles. The ARU will vary depending on the amount of correlation between temperature at different pressure levels and adjacent points.

In terms of designing a remote sounding system, the most important factor is, in terms of impact, the number and spectral location of the sensing channels (through the corresponding weighting matrices).

Figure VIII.2.2 performs such an analysis by plotting the variations of the ARU for the AMSU system as the number of channels increases from one to twelve. The channels are added in an increasing or decreasing height probing order as measured by the location and amplitude of the largest entry of the discretized weighting functions. In other words, in one case, the first channel probes closest to the ground and the last channel probes high in the atmosphere and vice versa for the second case. As expected the ARU is a monotonically decreasing function of the number of channels. The specific form of the ARU curve depends upon the order used to add new channels, i.e., it matters if the new channels used in the retrieval process are probing the atmosphere from the ground up or from the top of the atmosphere down. More information about the temperature profile is added by the lower probing channels thus resulting in a lower ARU which is as expected

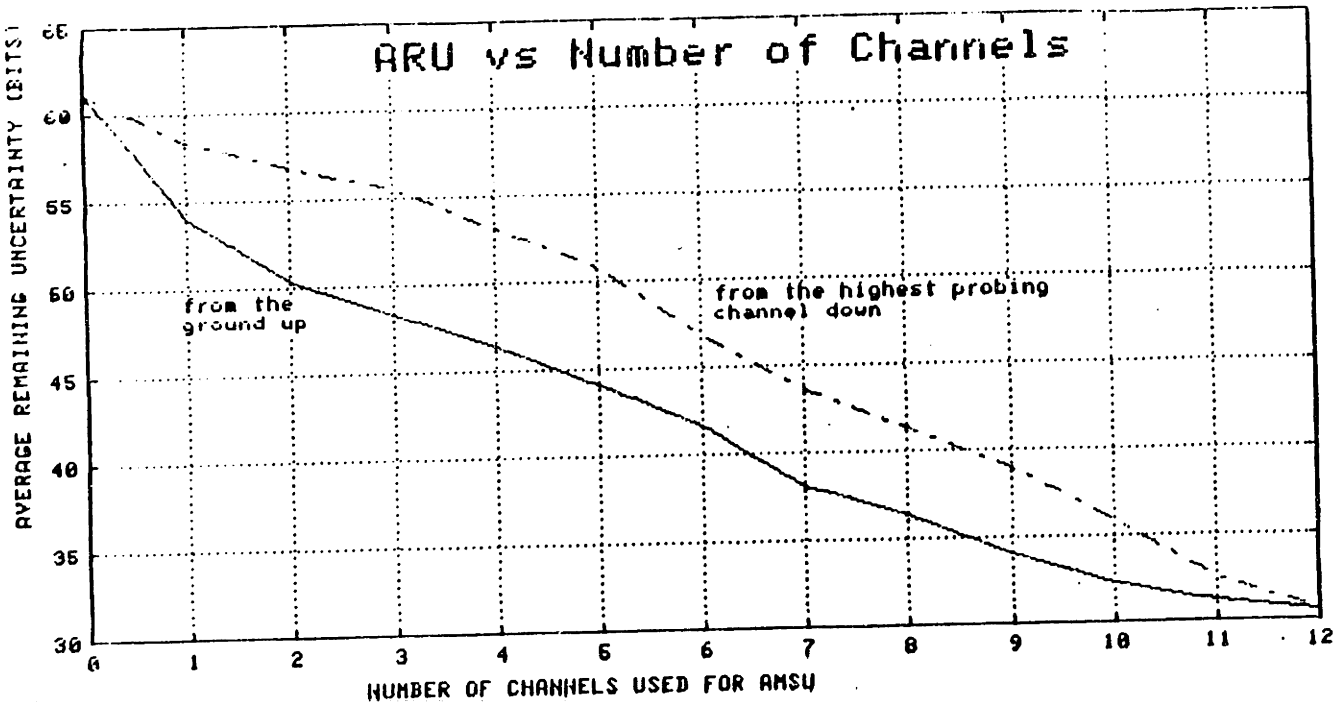


Figure VIII.2.2: Variations of the ARU for the AMSU system as the number of weighting functions increases from 1 to 12.

because of the larger variances at those pressure levels.

Finally, let us compare mean square error with average remaining uncertainty, more precisely their variations with the sounding parameters. The mean square error was computed in chapter V and found equal to

$$E(W,R,N) = trace\{R - RW^t(WRW^t + N)^{-1}R\} \qquad (VIII.2.23)$$

Projected in the SVD basis and using the same notation than in (VIII.2.19), this error is written as

$$E(W,R,N) = trace\{\tilde{N}(R_{oo} + \tilde{N})^{-1}R_{oo}\} + trace\{R_{uu} - R_{uo}(R_{oo} + \tilde{N})^{-1}R_{ou}\} \qquad (VIII.2.24)$$

When subtracting out the term representing the vertical representation of temperature profile of the average remaining uncertainty ($\log \sqrt{e} 2\pi^{\frac{L}{2}}$), one obtains

$$\tilde{H}(R,W,N) = \tilde{H}(T|T_b) = \log|R_{uu} - R_{uo}R_{oo}^{-1}R_{ou}| + \log|\tilde{N}(R_{oo}+\tilde{N})^{-1}\tilde{R}_{oo}| \qquad (VIII.2.25)$$

The resemblance between the two criteria is striking! To further appreciate the similitude between the two expressions, let us express in a different way the log of the determinant and the trace of a symmetric positive definite matrix. Let A be an $N \times N$ matrix and let $\lambda_{Ai}, i = 0, \ldots, N-1$ denote its set of eigenvalues (all positive and real). One then has

$$trace\{A\} = \sum_{0}^{N-1} \lambda_{Ai} \qquad (VIII.2.26)$$

whereas

$$\log|A| = \log \prod_{0}^{N-1} \lambda_{Ai} = \sum_{0}^{N-1} \log \lambda_{Ai} \qquad (VIII.2.27)$$

Thus, if the retrieval operator performs well from a mean square point of view, it will perform well from a average remaining uncertainty point of view. This is especially true if the matricial signal-to-noise ratio between the observable components and the normalized sensor noise is large and the matrices R_{oo} and $R_{oo} + \tilde{N}$ are close to one another. However, one cannot derive a one-on-one relationship between the number of bits of the ARU and the mean square of the LMMSE. Such a mapping can only be performed for each eigenvector of the matrices $R_{uu} - R_{uo}R_{oo}^{-1}R_{ou}$ and $\tilde{N}(R_{oo}+\tilde{N})^{-1}R_{oo}$.

Remember that, in the above analysis, the temperature profile has been modeled as a Gaussian random variable. For such a variable, the LMMSE provides the optimal solution. For a Gaussian variable, the informational content is found in the first and second moments which form the sufficient statistics and which also fully characterizes the linear retrieval process.

One noteworthy characteristic of the information theoretic approach to re-

mote sensing is that it characterizes the entire operation using the joint pdf between the physical temperature profile and brightness temperature measurements. As long as the estimation process is reversible, that is T_b can be reconstructed from $\hat{T}$, this characterization can be extended to any pdf. Thus, this is a perfect candidate for remote sensing problems such as the observation of water vapor (Rosenkranz *et al.*, 1982) where the microwave sensing is not a linear function of the geophysical parameters. In such an instance, the design of the optimal mean square estimator may be not practical, but the computation of the ARU might be possible. Because the ARU does not depend upon on the retrieval operator, it allows the characterization and the analysis of the retrieval process (noise level, radiative transfer equation, channel allocation, a priori statistics) without having to compute an operator.

VIII.2.2 Optimal Retrieval Operators Based on Different Criteria for Observable and Unobservable Components of Retrieved Temperature Fields

As discussed in section V.4.3, when optimized using the mean square error criterion, estimates of the observable components of the temperature profile are independent of any statistics related to the unobservable components. Thus, they can be completely separated from the estimation of the unobservable components. The estimation of the observable components from the brightness temperature is a associated with a complete observation problem, neither under- nor over-specified, where the parameters to be estimated are seen through a ful-rank invertible matrix and corrupted by an independent additive noise. The estimation of the unobservable components is now performed in two steps: (1) one estimates the observable components, (2) one maps these observable components onto the unobservable ones according to the mean square criterion.

Let us now examine the possibility of mixing criteria for observable and

unobservable components, in particular by changing the criterion optimized by the estimated observable components. Two different criteria, maximum likelihood and maximum power, are considered. In each case, the unobservable components are found by taking the minimum mean square estimate of the unobservable components given the observed ones.

The sounding equation for the observable components is (expressed in the SVD basis)

$$\overline{T}'_b = \Sigma \overline{T}_o + \eta' \qquad (VIII.2.28)$$

where the matrix Σ is square and diagonal. The observation noise η' has for covariance matrix N'.

• *Maximum Likelihood*

The maximum likelihood (ML) estimate is found by finding the observable components that maximize the conditional probability $p(\overline{T}'_b|\overline{T}_u)$ given the measured brightness temperature. Modeling the additive noise as Gaussian yields the conditional probability

$$p(\overline{T}_b|\overline{T}_o) = \frac{1}{(2\pi)^k |N'|^{-1/2}} \exp\left(-\frac{1}{2}(\overline{T}'_b - \Sigma\overline{T}_o)^t N'^{-1}(\overline{T}'_b - \Sigma\overline{T}_o)\right) \qquad (VIII.2.29)$$

The optimal estimate of the observable components is given by (Schweppe, 1973)

$$\hat{\overline{T}}_o = (\Sigma N'^{-1}\Sigma)^{-1}\Sigma N'^{-1}\overline{T}'_b \qquad (VIII.2.30)$$

This estimate takes a simplified form because the observation matrix is invertible and the noise covariance matrix is full-rank.

$$\hat{\overline{T}}_o = \Sigma^{-1}\overline{T}'_b \qquad (VIII.2.31)$$

The solution to the maximum likelihood estimation is similar to the Fisher estimation where the observable components are assumed to be completely unknown.

Since mapping from the observable to the unobservable relies on the statistics of the temperature profile, such a framework is not appropriate. This is true because this framework would take the observable components to be Fisher estimate (no statistics available) and the unobservable components to be Bayesian (some statistics available). From the maximum likelihood estimation point of view, the nature of $\overline{T}$ is irrelevant and the two criteria can be used simultaneously.

The unobservable components are found using the traditional formula

$$\hat{\overline{T}}_u = R_{T_uT_o} R_{T_oT_o}^{-1} \hat{\overline{T}}_o \qquad (VIII.2.32)$$

Combining the two estimates yields in the natural representation

$$\hat{\overline{T}} = RW^t(WRW^t + N')^{-1}\overline{T}_b \qquad (VIII.2.33)$$

which is the LMMSE when the observations are perfect. Profiles retrieved using this estimator will display more high frequency components than the traditional noisy LMMSE. This is true because the noise covariance matrix in the LMMSE tends to filter high frequency components down.

- *Maximum Power*

Maximum power (MP) estimation tends to incorporate as much energy in the estimated profile as possible. That is, it maximizes the quadratic form

$$E_o = < trace\{(\hat{\overline{T}}_o - \overline{m}_o)(\hat{\overline{T}}_o - \overline{m}_o)^t\} > \qquad (VIII.2.34)$$

where $< . >$ symbolizes the expectation over all temperatures and sensor noises. MP estimation was introduced by Trussell (1981) to restore X-ray spectra. Its purpose is to avoid low pass filtering of the estimate by the estimator. This method can be extended to vectorial estimation but only to the observable components, otherwise it would set the unobservable components to infinity.

Because such an optimization yields unbounded observable components, it must be constrained. The constraint used pertains to the observation noise. Namely,

$$< trace\{(\Sigma\hat{\overline{T}}_o - \overline{T}_b)(\Sigma\hat{\overline{T}}_o - \overline{T}_b)^t\} >= trace\{< \eta'\eta'^t >\} = N' \qquad (VIII.2.35)$$

The obvious solution method is to use Lagrange multipliers, thus forming the objective function

$$E_o =< trace\{(\hat{\overline{T}}_o - \overline{m}_o)(\hat{\overline{T}}_o - \overline{m}_o)^t\} > +\lambda < trace\{(\Sigma\hat{\overline{T}}_o - \overline{T}_b)(\Sigma\hat{\overline{T}}_o - \overline{T}_b)^t - N'\} \qquad (VIII.2.36)$$

Finding the optimal determination matrix by differentiation yields (after taking all appropriate expectations)

$$D(\Sigma R_{oo}\Sigma^t + N') + \lambda\Sigma^2 D(\Sigma R_{oo}\Sigma^t + N') = \lambda\Sigma(\Sigma R_{oo}\Sigma + N') \qquad (VIII.2.37)$$

which yields

$$D = (I + \lambda\Sigma^2)^{-1}\lambda\Sigma \qquad (VIII.2.38)$$

The Lagrange multiplier λ is found by plugging VIII.2.38 into VIII.2.37

$$trace\{(\Sigma(I+\lambda\Sigma^2)^{-1} - I)(\Sigma R_{oo}\Sigma^t + N')(\Sigma(I+\lambda\Sigma^2)^{-1} - I)^t - N'\} = 0 \quad (VIII.2.39)$$

This equation can be reduced by the introduction of σ_i, the singular values of the weighting matrix, r_{ooi} the variances of the observable components, and n_i the noise variances.

$$\sum_i \left(\frac{\sigma_i^2\lambda}{1+\lambda\sigma_i^2} - 1\right)(\sigma_i r_{ooi}\sigma_i n_i)\left(\frac{\sigma_i^2\lambda}{1+\lambda\sigma_i^2} - 1\right) = \sum_i n_i \qquad (VIII.2.40)$$

After manipulation, it reduces to

$$\sum_i n_i\left(\frac{1 + \frac{r_{ooi}}{n_i}\sigma_i^2}{1+\lambda\sigma_i^2}\right) = \sum_i n_i \qquad (VIII.2.41)$$

The Lagrange multiplier can be interpreted as an average of the SNR for each observable component. Hence, the optimal MP estimator depends upon the statistics

of the signal and sensor noise. However, each estimated observable component is determined only by its noise-perturbed measurement. Thus, in the SVD basis, the MP estimator is a diagonal matrix.

The retrieval operator's form as well as the interpretation of the Lagrange multiplier as a SNR makes the MP solution similar to the zeroth order Tikhonov Regularization (Tikhonov and Arsenin, 1977, Fymat, 1977). This method is used in the context of underspecified Fisher type problems where there are no available statistics about the temperature profile. It guarantees the existence of a unique and continuous estimate for the profile given the measurements. The zeroth order regularized solution is found by minimizing the functional relationship

$$|\Sigma\overline{T}_o - \overline{T}_b'|^2 + \alpha|L\overline{T}_c| \qquad (VIII.2.42)$$

where the functional L is a regularizing term and α is a constant which controls the trade-off between closeness to the measured brightness temperature and well conditioning of the inversion problem. The solution to such a estimation problem is of the form

$$\hat{\overline{T}}_o = (\Sigma^2 + \alpha H)^{-1}\Sigma\overline{T}_b' \qquad (VIII.2.43)$$

where H is a full rank matrix. The case where α equals zero is the case of perfect fidelity to the measurements. This is the minimum information solution where the estimation corresponds to the division of the brightness temperature $\overline{T}_b'$ by the appropriate gains.

A special case of such regularized estimation is one where H is diagonal. The retrieval becomes

$$\hat{\overline{T}}_o(i) = \frac{\sigma(i)}{\sigma(i)^2 + \alpha h_i}\overline{T}_b(i) \qquad (VIII.2.44)$$

Each observable component only depends upon one brightness temperature measurement reading, as is the case for the MP estimator. More on Tikhonov Regularization and other objective penalty function will be examined in the following section

In either case (ML or MP), the unobservable components are found using the mean square mapping from the observable to the unobservable

$$\hat{\overline{T}}_u = R_{T_u T_o} R_{T_o T_o}^{-1} \hat{\overline{T}}_o \qquad (VIII.2.45)$$

The ability to extend optimization criteria, developed in the context of scalar processing and overspecified (incomplete observation of the signal) vectorial processing, can be extended to underspecified processing by separating the estimation of observable and unobservable components. However, the resulting retrieval operators are not notably different from the LMMSE. This discovery tends to make their derivations more academic than practical. However, it opens the possibility for using complicated processing (signature recognition, pattern analysis, etc) developed in the context of image analysis to remote sensing from satellite by disregarding the unobservable components. The unobservables components are incorporated at the last moment of the estimation using a simple mapping between observable and unobservable components.

VIII.2.3 Objective Penalty Functions

Until now, the design of estimators has been guided by engineering considerations where an error measure (often related to the mean square error) between the true profiles and estimates is minimized. The only control left to the user is of a parametric nature where statistics of signal and of noise are tuned. However, a profile optimal from a mean square point of view may not be optimal to the final user of temperature estimates, persons doing numerical weather prediction and data assimilation. This is true because the impact of estimation error on the numerical prediction depends upon the spectral content of the error (both in the vertical and horizontal direction) and upon the spectral content on the true temperature field (sometimes referred to as the baseline field).

Because the equations describing atmospheric flow are not linear (check chapter III), growth of error in numerical weather predictions is not linear either. Thus, exact error bounds cannot be computed past a certain span of time. This creates problems in terms of data assimilation because the relative weights to give to measurements and to the computer generated field are unknown. In general, high frequency terms (short wavelengths) are the most nonlinear, and saturate the in weather prediction the fastest (Fiocco, 1970, Lilly, 1983a, Rosenkranz and Staelin, 1984). Errors in the temperature fields will result, through the geostrophic conditions, in errors in the wind fields. Errors in the balance of wind fields result in high frequency gravity waves (Kasahara, 1977, Toldalagi, 1980) that must be filtered out before running general circulation models. Thus, it can be seen that, at times, reducing the high frequency content of the estimation is beneficial.

The most natural thing to reduce the high frequency content of the estimate (in this section, we will concentrate on the vertical description of temperature fields) is to weigh the MSE more heavily for the high frequency components of the estimate. That is use a modified optimization criterion

$$E = trace\{\mathcal{B}(\hat{\overline{T}} - \overline{T})(\hat{\overline{T}} - \overline{T})^t\} \qquad (VIII.2.46)$$

However for non-singular weight matrices $\mathcal{B}$, ie with trivial null space, the optimum determination matrix does not depend upon the weight matrix (Luenberger, 1976, Malvar, 1986).

The correct method adds an extra term to the MSE to reduce the high frequency content. This operation is performed using a concept known as roughness.

The roughness M_R of a temperature profile is defined as the integrated value of the second derivative squared of the temperature with respect to height or isobaric displacement (Wahba, 1981). The integration is performed over the relevant

domain. Thus, for roughness in the vertical direction

$$M_R = \int_0^\infty \left(\frac{\partial^2 T(h,x,y)}{\partial h^2} \right)^2 dh \qquad (VIII.2.47)$$

and in the horizontal direction

$$M_R = \int_{x,y\in\Omega} \left(\frac{\partial^2 T(h,x,y)}{\partial x^2} \right)^2 + \left(\frac{\partial^2 T(h,x,y)}{\partial y^2} \right)^2 dxdy \qquad (VIII.2.48)$$

This dissertation will concentrate on the control of the vertical roughness.

The criterion to optimize becomes

$$E = < trace\{(\hat{\overline{T}} - \overline{T})(\hat{\overline{T}} - \overline{T})^t\} > + \lambda M_R \qquad (VIII.2.49)$$

where λ controls the tradeoff between MSE and roughness. Optimal determination of the parameter λ is the subject of intensive research (Wahba and Wendelberger, 1980, Wahba, 1984, Backus and Gilbert, 1970). The second term on the right hand side of the above equation is often referred to as the regularizing functional. The overall criterion is called an objective penalty function. λ is usually positive in order to smooth the estimated profile, or at least the deviation of the mean profile from its mean. Nothing prevents λ from being negative to accentuate the high frequency components of the signal. However, as it will soon become apparent, a negative λ can result in ill-conditioned design equations.

Because temperature is not estimated as a continuous variable but rather as a vector, computation of the smoothness (or roughness) is intricate. Integration must be replaced by a discrete summation. Several techniques are available to perform this task: all are based on some sort of interpolation of the vector to a continuous signal.

Because the design and implementation of retrieval operators is done discretely, a quadratic expression for the roughness in terms of the temperature profile $\overline{T}$ must be derived.

Recall that, in chapter V, the temperature profiles are assumed to be piecewise linear (in the height or log-pressure domain) between sampled pressure levels. This particular interpolation method was used because it facilitates the computation of the discrete weighting matrix. This interpolation is not appropriate for the intended purpose of this derivation since, at each point of the interpolated profile, the radius of curvature is either zero or infinite.

Since sampling of the profile was determined to avoid aliasing as much as possible, let us use the sampling theorem (Papoulis, 1965) to compute the exact integrated roughness and then derive a simple (and good) approximation.

Assume that the sampling raster is (near) regular and the discrete profile is an infinite time series. The continuous profile is then expressed as

$$T(h) = \sum_i sinc(\pi \frac{h - Hi}{H}) T(Hi) \qquad (VIII.2.50)$$

where H is the sampling distance and $sinc$ the cardinal sinus ($\sin x/x$). The second derivative is computed by differentiating each term in the series. Hence

$$T''(h) = \sum_i (\frac{\pi}{H})^2 s(\pi \frac{h - Hi}{H}) T(Hi) \qquad (VIII.2.51)$$

where

$$s(x) = -\frac{\sin x}{x} + 2\frac{\sin x}{x^3} - 2\frac{\cos x}{x^2} \qquad (VIII.2.52)$$

The Taylor expansion of the interpolation function $s(x)$ around 0 is

$$s(x) \sim -1/3 \qquad (VIII.2.53)$$

and around $k\pi$ is

$$s(x) \sim \frac{2(-1)^{k+1}}{k^2\pi^2} \qquad (VIII.2.54)$$

Hence, at the sampling points

$$T''(iH) = -\frac{1}{3} T(iH)(\frac{\pi}{H})^2 + \sum_{j \neq 0} \frac{2(-1)^{j+1}}{j^2\pi^2} T(iH + jH)(\frac{\pi}{H})^2 \qquad (VIII.2.55)$$

This expression can be written as the inner product of the temperature profile with the vector J where

$$\begin{aligned} J^t &= \frac{\pi^2}{H^2}(\ldots, -\frac{2}{4\pi^2}, \frac{2}{2\pi^2}, -\frac{1}{3}, \frac{2}{2\pi^2}, -\frac{2}{4\pi^2}, \ldots) \\ &= \frac{\pi^2}{3H^2}(\ldots, -\frac{6}{4\pi^2}, \frac{3}{\pi^2}, -1, \frac{3}{\pi^2}, -\frac{2}{6\pi^2}, \ldots) \end{aligned} \qquad (VIII.2.56)$$

Within a scaling factor, this vector can be approximated by the vector associated with the second difference in the time series $T(iH)$, namely

$$J'^t = (\ldots, 0, 0, \frac{1}{2}, -1, \frac{1}{2}, 0, 0, \ldots) \qquad (VIII.2.57)$$

When the sampling is not more regular, the interpolation function is no longer sinus cardinal. While it is still possible to interpolate the discrete profile, using say a Lagrange polynomial, exact computation of the second derivative becomes intricate and sensitive to the sampled pressure level positions. Thus, the approximation using the second difference vector is adequate. Before the second difference is computed, the profile must first be interpolated on a regular grid. Curvature at the i^{th} pressure level (i increases from the ground up) is computed by taking the inner product of the profile $\overline{T}$ with J_i where

$$J_i = (\ldots, 0, 0, 1 - \frac{\ln p(i-1) - \ln p(i)}{\ln p(i) - \ln p(i-1)}, -2 + \frac{\ln p(i+1) - \ln p(i)}{\ln p(i) - \ln p(i-1)}, 1, 0, 0, \ldots) \qquad (VIII.2.58)$$

The vectors J_i are then concatenated to yield an operator M which changes the temperature profile into its second derivative. The optimization criterion thus becomes

$$E = trace\{(\hat{\overline{T}} - \overline{T})(\hat{\overline{T}} - \overline{T})^t\} + \lambda trace\{M\hat{\overline{T}}\hat{\overline{T}}^t M^t\} \qquad (VIII.2.59)$$

which can be rewritten after taking all expectations into account and introducing the determination matrix D

$$E = trace\{(DW - I)R(W^tD^t - I) + DND^t + \lambda M^tMD(WRW^t - N)D^t\} \qquad (VIII.2.60)$$

where R and N are as before.

The optimal retrieval operator is found by differentiation

$$D = (I + \lambda MM^t)^{-1} RW^t (WRW^t + N)^{-1} \qquad (VIII.2.61)$$

The impact of the smoothness constraint is found to be the multiplication of the LMMSE by the inverse of $(I + \lambda MM^T)$. For positive λ (reduction of roughness), this matrix is positive definite and has all eigenvalues larger than one. Hence, the objective-derived retrieval operator provides an estimated profile between the a priori mean and the LMMSE estimate. The estimator reduces the roughness by reducing the amplitude (defined as the deviation from the mean profile) of the estimate. In that sense, it is equivalent to an estimator operating on noisier measurements. The components most scaled down from the LMMSE estimate are those that correspond to the largest eigenvalues of MM^t. These components are associated with high frequency components. These components are close to high order eigenvectors of the vertical covariance matrix.

Setting the vertical covariance kernel R to an infinite diagonal yields the minimum information solution

$$D = (I + \lambda MM^T)^{-1} W^T (WW^t)^{-1} \qquad (VIII.2.62)$$

which is the Tikhonov regularized solution (Fymat, 1977, Gill and Murray, 1974). When the statistical description is not provided, sharp discontinuities such as the tropopause inversions are not reproduced.

Negative values of λ correspond to cases where the high frequency content is amplified thereby emphasizing sharp discontinuities. The eigenvalues of $I + \lambda MM^t$ are no longer greater than one and can even become negative. When this is the case, the existence of a unique determination matrix D is not guaranteed. This ill-behavior of the operator is caused by the tentative estimation of high frequency

components from low pass measurements of the temperature profiles. The information lost in the sensing cannot be recovered without degeneracy.

The optimal retrieval operator for an objective penalty function is thus found to be the LMMSE multiplied by a matrix that is simple to evaluate and to invert. The product of matrix MM^t can be replaced by any positive semidefinite matrix and the retrieval operator can be altered without recomputing the LMMSE. For instance, extreme smoothing of the profile around the tropopause reduces the impact of misspecifying the height of the tropopause through the covariance matrix R. Smoothing of the profile at the tropopause inversion corresponds to a loss in vertical resolution. This behavior is another case of tradeoff between resolution and noise-sensitivity (Rodgers, 1976, Backus and Gilbert, 1970).

By smoothing the profile around the tropopause inversion, the retrieval operator does not utilize the a-priori information R. Another way of achieving the same effect is to use a partial minimum information solution around the tropopause by setting the interlevel correlation to zero. Tuning the smoothness is a better method to achieve this result, because it does not require the computation and inversion of $WRW^t + N$. Moreover, it can be implemented on temperature profiles already estimated using the LMMSE.

As indicated, any positive definite matrix Q can be used according to

$$D = (I + Q)^{-1} RW^t (WRW^t - N)^{-1} \qquad (VIII.2.63)$$

to reduce the amplitude of the estimated components of $\overline{T}$.

Consider the Karhunen-Loeve (KL) expansion of the vertical covariance matrix R. That is, R can be written as (Pratt, 1978)

$$R = P\mathcal{R}P^t \qquad (VIII.2.64)$$

where $\mathcal{R}$ is a diagonal matrix whose entries are decreasing along the diagonal and P an unitary matrix formed by the eigenvectors of R that satisfies $PP^t = I$. Typically,

only a few eigenvalues of R are significant (see chapter VI). This is a direct result of the energy compaction properties of the KL transform (Van Trees, 1963, Watanabe, 1965).

Consider the diagonal matrix whose entries are zero for the first ℓ entries and are equal to Θ, a large number for the $33-\ell$ last entries. Finally, define Q as the product

$$Q = P \begin{array}{c} \begin{array}{ccccccc} \leftarrow & \ell & \rightarrow & \leftarrow & 33-\ell & \rightarrow & \end{array} \\ \begin{pmatrix} 0 & & & & & & 0 \\ & \ddots & & & & & \\ & & 0 & & & & \\ & & & \Theta & & & \\ & & & & \ddots & & \\ 0 & & & & & & \Theta \end{pmatrix} \end{array} P^t \qquad (VIII.2.65)$$

Using (VIII.2.65) into (VIII.2.63) yields the same expansion based estimator developed by Smith and Wolf (1976) which estimates components of the temperature profile along the eigenvectors with the largest eigenvalues and sets the others to zero.

As mentioned, retrieval operators derived from an objective penalty function will be outperformed in terms of rms error by the LMMSE. Hence it is difficult to derive an heuristic demonstration of the importance of these operators. As discussed, from a MSE point of view, the contrained retrieval operators will outperform the LMMSE in cases where the later is not truly the optimal operators because of errors either in the signal statistics (as discussed in great detail in chapter VII) or in the signal receiver noise.

Consider then the case where the retrieval operator underestimates the sensor noise. This case can also occur in instance where unmodeled errors in the sounder gain take place, or if the calibration process on-board the satellite is not working

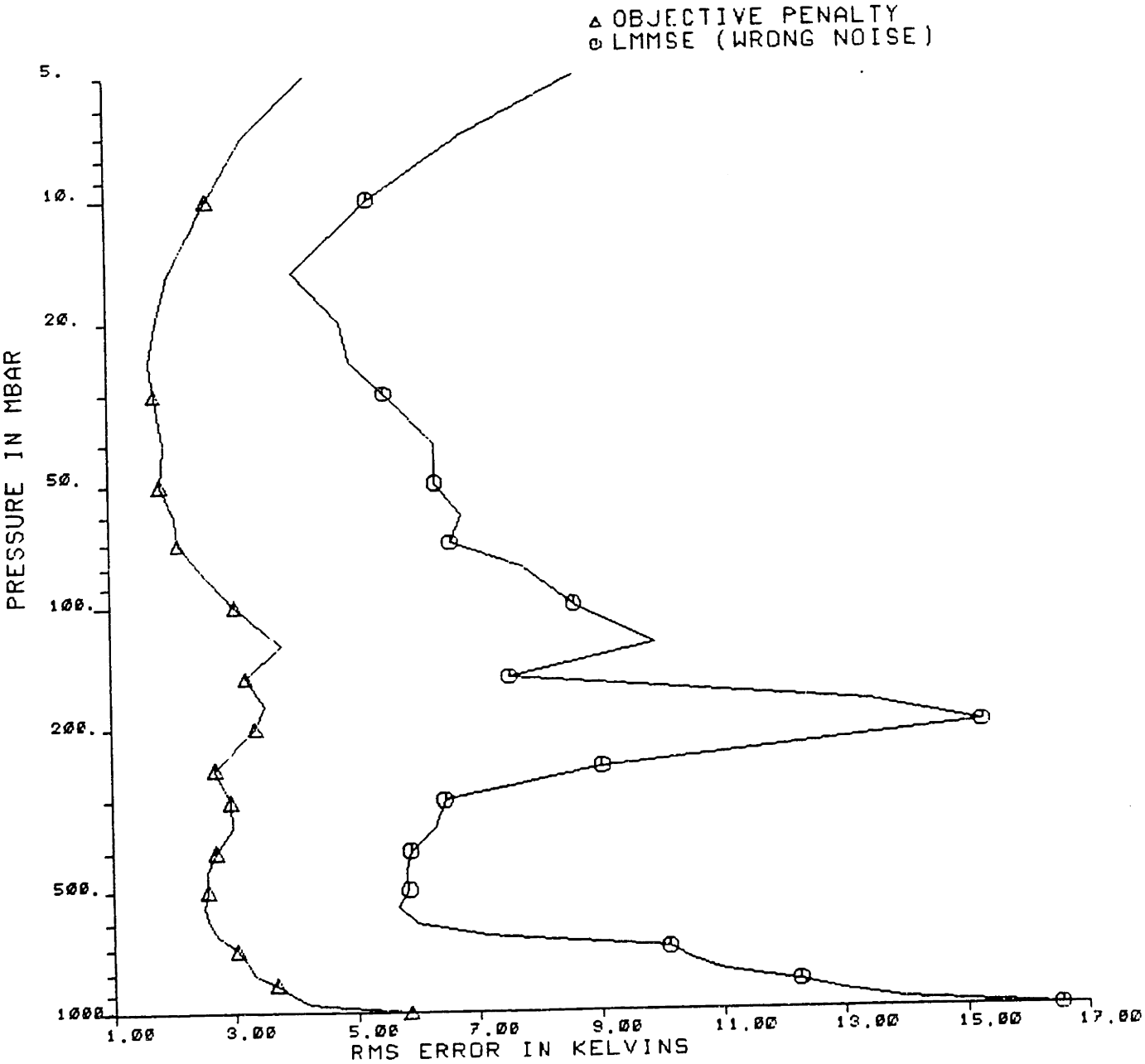


Figure VII.2.3: Rms error for constrained LMMSE and unconstrained LMMSE for MSU at nadir over Summer, noise covariance underestimated by a factor 30

properly. In the simulation presented by Figure VIII.2.1, the true noise rms of the receiver is 30 times the value used to derive the LMMSE. The retrievals are

performed for the Peoria summer dataset. The matrix M used for the smoothing is the one given by equation VIII.2.58, and the smoothing coefficient λ equals 10. Needless to say that the operator forcing the smoothing does a better job than the LMMSE which does not use the correct signal-to-noise. Needless to say that this simulation is somewhat extreme. This goal of the above analysis is to show that one can incorporate objective criteria in the Bayesian framework of estimation, and that the LMMSE can be tailored to specific cases by being multiplied on the left by a correction matrix dictated by user-objectives.

VIII.3 Constrained Estimation: The SMILE

In this section, the impact of bounds on the values of the temperature profiles and on the values of the lapse rate are considered.

There are other types of restrictions on the estimated temperature profiles, but they are usually not of an analytical nature. These methods are usually implemented in the form of Landweber type iterative restoration methods with alternate projections onto the observation (brightness temperature) and signal spaces (temperature profile) (Youla, 1978, Schafer *et al.*, 1981, Trusell and Civanlar, 1984). These methods incorporate complicated information about the profiles such as spectral bandlimitations, positivity constraints (deviation from the a-priori mean), bound constraints (range of possible values for the deviations), finite support constraints (regions where the profile is equal to its a-priori mean), etc.

These methods can be combined with a Chahine type inversion method (Chahine, 1969, 1970) to reduce retrieval error over specific classes of temperature profiles. However, with new sounding systems such as AMSU (with its 12 channels), the gains of these methods over the LMMSE, which implicitly assumes that the pdf of the profiles are Gaussian, are not clear. This is true because the power in the observable components of AMSU represents a significant percentage of

the power to be recovered by the estimation. As mentioned, most of the improvement provided by constraints and multidimensional retrieval is in the unobservable subspace.

Moreover, because of the non stationarities of atmospheric temperature fields and their temporal evolution, these constraints must be constantly tested for validity.

Hence, it is preferable to use constraints which alter the LMMSE rather than change the method of estimation to an iterative framework. This is especially true if multi-spot retrieval schemes are considered.

As stated, the contrained estimation considered in this section deals with cases where the set of temperature profiles is bounded by minimum temperature profile $\overline{T}_m$ and a maximum temperature profile $\overline{T}_M$. This case is of interest to remote sensing because such a characterization is easy to determine.

The thrust of the analysis is to combine this bound information with the traditional first and second moment characterization to derive a better linear estimate of the temperature profile. Before deriving the criteria associated with this problem and solving for the associated optimal retrieval operator, one must examine the best method of representing this bound information and reexamine the relationship between Fisher (MIE) and Bayesian estimation (LMMSE) theories.

The restriction on the temperature profile's range can be written for each component as

$$T_m(i) \leq T(i) \leq T_M(i) \qquad i = 1, \ldots, K = 33 \qquad (VIII.3.1)$$

This equation defines a cuboid in $\mathbf{R}^k$. Figure VIII.3.1 presents the minimum and maximum temperature profiles for the Peoria dataset used to derive the retrieval operators in Chapter VI and VII.

A cuboid is not practical to incorporate in the estimation because it cannot

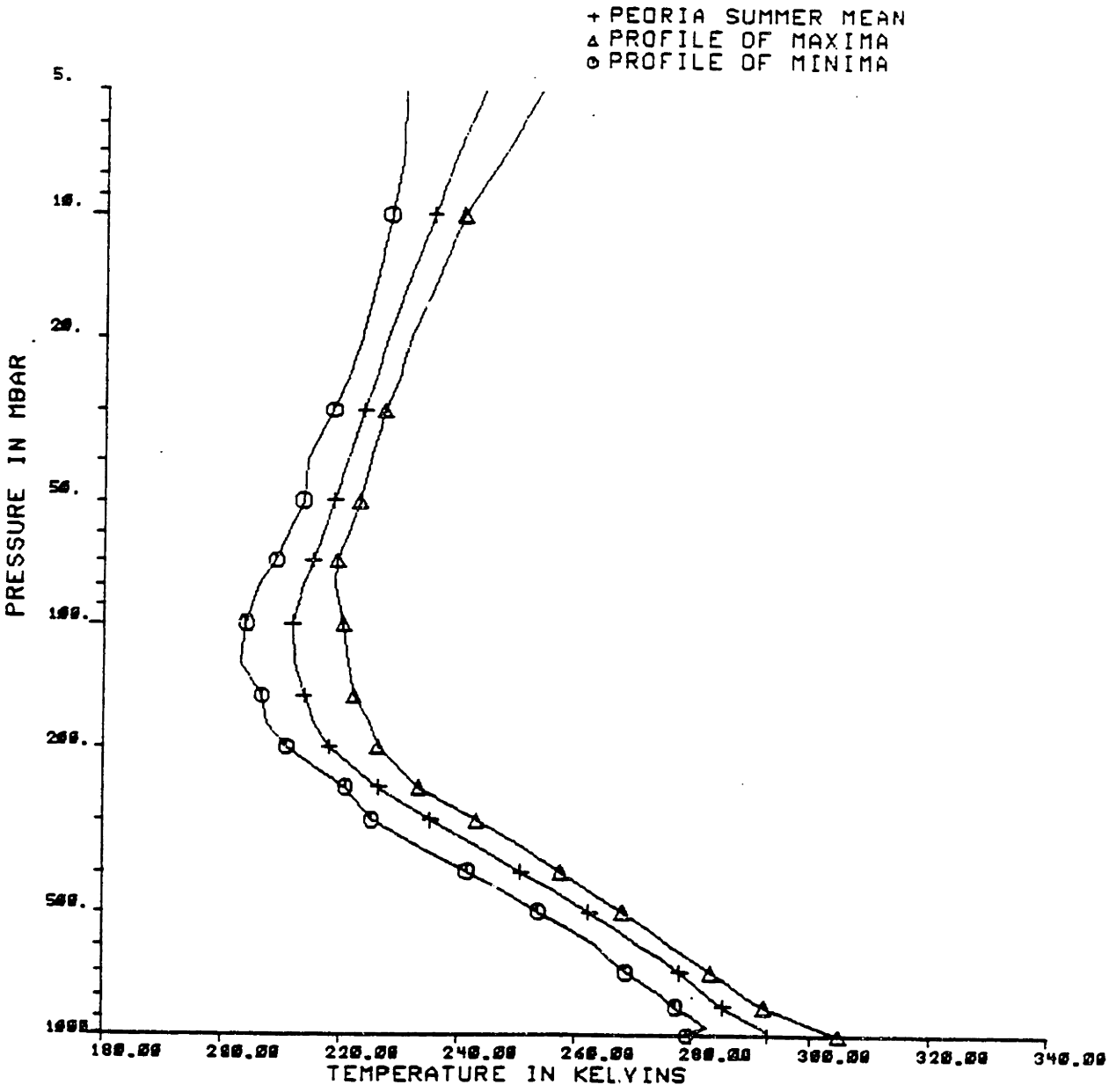


Figure VIII.3.1: Mean temperature profile, profile of minima, and profile of maxima over Peoria dataset.

be expressed in terms of the entire profile T (Schweppe, 1973). For this reason, one must find an ellipsoid of the form

$$(\overline{T} - \overline{m})^t \mathcal{E} (\overline{T} - \overline{m}) = 1 \qquad (VIII.3.2)$$

which contains the cuboid. $\mathcal{E}$ is a symmetric matrix which will be referred to as the extension matrix associated with the ellipsoid. Because the purpose of introducing

the ellipsoid is to restrict the range of the estimate, its volume must be as small as possible. For this reason, its center must be at the same location as the cuboid, that is

$$\overline{m} = \begin{pmatrix} \frac{T_m(1)+T_M(1)}{2} \\ \vdots \\ \frac{T_m(K)+T_M(K)}{2} \end{pmatrix} \qquad (VIII.3.3)$$

Moreover, the ellipsoid contain all cornerpoints of the cuboid and is aligned with its axes. The ellipsoid satisfying all these properties is characterized by the extension matrix (Toutenburg, 1982)

$$\mathcal{E} = \frac{4}{K} \begin{pmatrix} \frac{1}{(T_m(1)-T_M(1))^2} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{(T_m(K)-T_M(K))^2} \end{pmatrix} \qquad (VIII.3.4)$$

By this point, the constraint on the temperature profile has been weakened to the form of

$$(\overline{T} - \overline{m})^t \mathcal{E} (\overline{T} - \overline{m}) \leq 1 \qquad (VIII.3.5)$$

Note that for pressure levels where no constraint is known, the corresponding entry in the extension matrix $\mathcal{E}$ is null.

Using extension matrices which are not diagonal, it is possible to represent different classes of profiles; several extend ellipsoids can be combined by taking a larger ellipsoid that includes all others. The eigenvectors and corresponding eigenvalues of such extension matrices determine the orientation and width of these classes (Parks and Meier, 1971). For instance, in the troposphere the lapse rate can be assumed to be within two fractions of the adiabatic lapse rate. Hence,

$$l_m(\ln p - \ln p') \leq T(p) - T(p') \leq l_M(\ln p - \ln p') \qquad (VIII.2.6)$$

where l_m and l_M are the lower and upper bounds on the lapse rate. The extension

matrix for such a constraint is

$$\mathcal{E} = \frac{1}{(l_m - l_M)^2(\ln p - \ln p')^2} \begin{pmatrix} \ddots & & & & & \\ & 0 & & & & \\ & & 1 & -1 & & \\ & & -1 & 1 & & \\ & & & & 0 & \\ & & & & & \ddots \end{pmatrix} \qquad (VIII.2.7)$$

where 1 and -1 are located at the pressure levels where the lapse rate bounds are taken (they are assumed to be adjacent in the above example). Extend matrices can also be used in the isobaric domain to restrict, for instance, the displacement in the height of the tropopause from point to point.

In chapter V, it was demonstrated how the Fisher estimate (MIE) is a limiting form of the Bayesian estimate (LMMSE) when the signal becomes diagonal with infinite entries. The reverse operation (from Fisher to Bayesian) is possible using the mean of the temperature profile as a pseudomeasurement. The original observation equation is

$$\overline{T}_b = W\overline{T} + \overline{\eta} \qquad (VIII.3.8)$$

and the pseudomeasurement is

$$\overline{m} = \overline{T} + \overline{\epsilon} \qquad (VIII.3.9)$$

where the covariance matrix of the deviation ϵ is R, the covariance of the temperature profile $\overline{T}$. Combining the two equations together yield

$$\begin{pmatrix} \overline{T}_b \\ \overline{m} \end{pmatrix} = \begin{pmatrix} I \\ W \end{pmatrix} \overline{T} + \begin{pmatrix} \overline{\eta} \\ \overline{\epsilon} \end{pmatrix} \qquad (VIII.3.10)$$

where the "observation" noise has for covariance matrix Ψ

$$\Psi = \begin{pmatrix} N & 0 \\ 0 & R \end{pmatrix} \qquad (VIII.3.11)$$

Note that the resulting observation equation is overspecified and thus the noise covariance matrix enters into account in the MIE (Rao, 1973). This is true because

observation noise guarantees that all observations are compatible. The MIE for this modified observation equation is the traditional LMMSE (Schweppe, 1979) if the a-priori mean is correct (a fact that will be assumed in the remaining portion of this section).

Introducing the mean as a pseudomeasurement serves two purposes: (1) to introduce statistics and (2) to transform an underspecified problem into an overspecified one. This last result is important in terms of developing estimators in the Fisher framework because most of these algorithms depend on overspecification (Rao, 1973). Note that the pseudomeasurements are characterized by a relatively high noise as compared with the sensor noise, but with a higher resolution (Rodgers, 1976).

Let us now develop the Statistical MInimax Linear Estimator (SMILE) which combines bound information and first and second moment characterization. This estimator is an extension to the Bayesian framework of the minimax linear estimation introduced by Kuks and Olman (1971, 1972). The observation equation is

$$\overline{T}_b' = H\overline{T} + \overline{\eta}' \qquad (VIII.3.12)$$

Restriction on the profiles is of the form

$$(\overline{T} - \overline{m})^t \mathcal{E} (\overline{T} - \overline{m}) \leq 1 \Longleftrightarrow \overline{T} \in \Omega_T \qquad (VIII.3.13)$$

the estimation equation is

$$\hat{\overline{T}} = D^* \overline{T}_b' \qquad (VIII.3.14)$$

and the square error is

$$E(T) = trace\{(\hat{\overline{T}} - \overline{T})(\hat{\overline{T}} - \overline{T})^t\} \qquad (VIII.3.15)$$

The square error is unbounded over $\mathbf{R}^k$, although its mean value is finite. However, the error is bounded if $\overline{T} \in \Omega_T$.

The estimate $\hat{\overline{T}}^*$ is said to be a statistical minimax linear estimator of $\overline{T}$ if

$$\min_{\hat{\overline{T}}} \max_{\overline{T} \in \Omega_T} E(\hat{\overline{T}}) = E(\hat{\overline{T}}^*) \qquad (VIII.3.16)$$

The SMILE minimizes the expected value of the maximum square error that can be reached on the set of valid temperature profiles. Note that the worst square error criterion, whose appropriateness for bounded estimation is evident, as recently been established as being relevant in image processing (Tom, 1986). The estimator depends on the constraint ellipsoid Ω_T, i.e. the extension matrix $\mathcal{E}$.

Taking the expectation of the square error over all possible noise $\overline{\eta}'$ yields

$$E(T) = trace\{WNW^t\} + trace\{(\overline{T}-\overline{m})^t(DW-I)^t(DW-I)(\overline{T}-\overline{m})^t\} \quad (VIII.3.17)$$

Introducing the matrix A defined as

$$A = \mathcal{E}^{-1/2}(DW-I)^t(DW-I)\mathcal{E}^{-1/2} \qquad (VIII.3.18)$$

the square error becomes

$$E(T) = trace\{WNW^t\} + trace\{(\overline{T}-\overline{m})^t\mathcal{E}^{1/2}A\mathcal{E}^{1/2}(\overline{T}-\overline{m})\} \qquad (VIII.3.19)$$

The maximum value of the error over the set Ω_T is found by maximizing the Rayleigh quotient

$$\Xi = \frac{(\overline{T}-\overline{m})^t(DW-I)^t(DW-I)(\overline{T}-\overline{m})}{(\overline{T}-\overline{m})^t\mathcal{E}(\overline{T}-\overline{m})} \qquad (VIII.3.20)$$

or equivalently

$$\Xi = \frac{(\overline{T}-\overline{m})^t\mathcal{E}^{1/2}A\mathcal{E}^{1/2}(\overline{T}-\overline{m})}{(\overline{T}-\overline{m})^t\mathcal{E}^{1/2}\mathcal{E}^{1/2}(\overline{T}-\overline{m})} \qquad (VIII.3.21)$$

This quotient is maximized when $(\overline{T}-\overline{m})\mathcal{E}^{1/2}$ is aligned with the eigenvector associated with the largest eigenvalue $\lambda_{max}(A)$ of A (Strang, 1980) and the maximum square error is

$$\begin{aligned} \underset{T \in \Omega_T}{E_{max}}(T) &= trace\{WNW^t\} + \lambda_{max}(A) \\ &= trace\{WNW^t\} + \lambda_{max}(\mathcal{E}^{-1/2}(DW-I)^t(DW-I)\mathcal{E}^{-1/2}) \end{aligned} \qquad (VIII.3.22)$$

The introduction of $\mathcal{E}^{-1/2}$ is not too surprising since it corresponds to the conformal mapping of the constraint ellipsoid into an hypercircle in $\mathbf{R}^k$.

The SMILE is obtained by finding a D that maximizes VIII.3.22. Because variations of A's eigenvalues as a function of D are not easy to characterize (although there are continuous), the solution to the SMILE is intricate. Lauter (1975) solved the MInimax Linear Estimation (MILE) problem. Its solution can readily be extended to this problem.

Key to the solution is the introduction of an intermediate matrix connecting the statistical information with the conformal mapping operator $\mathcal{E}^{1/2}$. The matrix

$$S = W^t N^{-1} W \qquad (VIII.3.23)$$

corresponds to the impact of the receiver noise on the estimate. $\tilde{S}$ is the previous matrix operated upon by the conformal mapping $\mathcal{E}^{-1/2}$, namely

$$\tilde{S} = \mathcal{E}^{-1/2} S \mathcal{E}^{-1/2} = \mathcal{E}^{-1/2} W^t N^{-1} W \mathcal{E}^{-1/2} \qquad (VIII.3.24)$$

Lauter proved that there exists a positive definite perturbation matrix V and a positive scalar v such that

$$\begin{aligned} \frac{1}{v}(\mathcal{E}^{1/2} S^{-2} \mathcal{E}^{1/2} + V) &\geq \tilde{S}^{-2} \\ v(\mathcal{E}^{1/2} S^{-2} \mathcal{E}^{1/2} + V)^{-1/2} V &= \tilde{S}^{-2} V \\ trace\{(\mathcal{E}^{1/2} S^{-2} \mathcal{E}^{1/2} + V)^{1/2}\} &= v(1 + trace\{\tilde{S}^{-1}\}) \end{aligned} \qquad (VIII.3.25)$$

Using these quantities,one defines then an equivalent signal covariance matrix R_e (different from the original R) as

$$R_e = \mathcal{E}^{1/2}[\frac{1}{\sqrt{v}}(\mathcal{E}^{1/2} S^{-2} \mathcal{E}^{1/2} + V)^{1/2} - \tilde{S}^{-1}]\mathcal{E}^{1/2} \qquad (VIII.3.26)$$

The covariance R_e can be thought of as a compromise between the second moment characterization and the bounds. The SMILE is obtained by finding the LMMSE

to the modified observation equation where the temperature profile T has for its covariance matrix R_e. Hence

$$D^* = R_e H^t (H R_e H^t + \Psi)^{-1} \qquad (VIII.3.27)$$

The computation of V is simplified if the extension matrix has constant entries, i.e. the ellipsoid is a hypercircle. This corresponds to an ellipsoid fitted to bounds which are the same at all pressure levels. This matrix can always be obtained by choosing the widest of these bounds. Taking $\mathcal{E}$ to be γI yields

$$D^* = \frac{\gamma^{-1}}{\gamma^{-1} + trace\{W N^{-1} W^t\}} R W^t (W R W^t + N)^{-1} \qquad (VIII.3.28)$$

The SMILE is a scaled version of the LMMSE. This result one more time demonstrates that the LMMSE is the kernel to linear estimation even when new information about the profiles is available This behavior is similar to the LMMSE derived using auxiliary objective penalties. In that sense, the term $trace\{\gamma R^{-1}\}$ can be thought of as a tuning element which controls the confidence one has in the bounds; it plays the role of the product $\lambda M M^t$ in objective penalty criteria.

As γ tends to zero (loose bounds), the SMILE tends towards the LMMSE which can thus be interpreted as an unrestricted SMILE. As γ tends towards infinity, the SMILE tends towards zero and the estimated profile becomes the mean profile as expected. The behavior of the estimator for bad SNR ($|W R W^t| \ll |N|$) is, with regards to our previous experience, unusual. Under these conditions, the SMILE becomes

$$D^* = \frac{\gamma^{-1}}{\gamma^{-1} + trace\{W^t N^{-1} W\}} R W^t N^{-1} = \frac{1}{1 + trace\{\gamma W^t N^{-1} W\}} R W^t N^{-1} \qquad (VIII.3.29)$$

If the bounds are loose compared to the measurement noise ($|\gamma| << |N|$), the operator tends towards zero as does the LMMSE. Conversely, for tight bounds, the operator becomes

$$D^* = R W^t \frac{N^{-1}}{\gamma trace\{W^t N^{-1} W\}} \qquad (VIII.3.30)$$

Except if the weighting matrix is not full row-rank, which is never the case in remote sensing, this expression does not tend towards zero as the measurement noise diverges. The limiting form of the estimator depends on the distribution of noise across the sensors and on the singular values of the weighting matrix.

Expressing the determination matrix in the SVD basis yields

$$D^* = \begin{pmatrix} R_{oo}\Sigma \\ R_{uo}\Sigma \end{pmatrix} \frac{N'}{\gamma trace\{\Sigma N'\Sigma\}} = \begin{pmatrix} R_{oo}\Sigma \\ R_{uo}\Sigma \end{pmatrix} S \qquad (VIII.3.31)$$

where $\tilde{N}$ is the noise covariance rotated in the SVD basis. Denoting the elements of that matrix as n_{ij} and σ_i the singular value of W, the elements of the matrix S in (VIII.3.31) are

$$S_{ij} = \frac{n_{ij}}{\gamma \sum_i \sigma_i^2 n_{ii}} \qquad (VIII.3.32)$$

For a fixed γ, the coefficients of the retrieval operator depend upon the relative values of the elements of the rotated matrix N'. As for the LMMSE, estimation of the temperature profile can be interpreted as a two step process: (1) the estimation of the observable components,(2) the mean square mapping from observable to unobservable components. This property holds for all SMILEs where the extension matrix is diagonal, regardless of the SNR and relative values of γ and $trace\{R^{-1}\}$. However, for arbitrary extension matrices $\mathcal{E}$, this is not true.

For an homoscedastic sounding systems (all channels having the same nominal rms error in the measurements), the off-diagonal terms of N' are zero. This is true because a diagonal matrix with constant entries does not vary under rotation. The estimation of the observable components itself becomes a two step process. Each noise-corrupted brightness temperature (expressed in the SVD basis) is scaled by

$$\varpi_i = \frac{\sigma_i n_{ii}}{\gamma \sum_i \sigma_i^2 n_{ii}} \qquad (VIII.3.33)$$

These multiplications are similar to the case of the Tikhonov regularization (equation VIII.2.44) where the smoothing matrix is diagonal. The second step in the

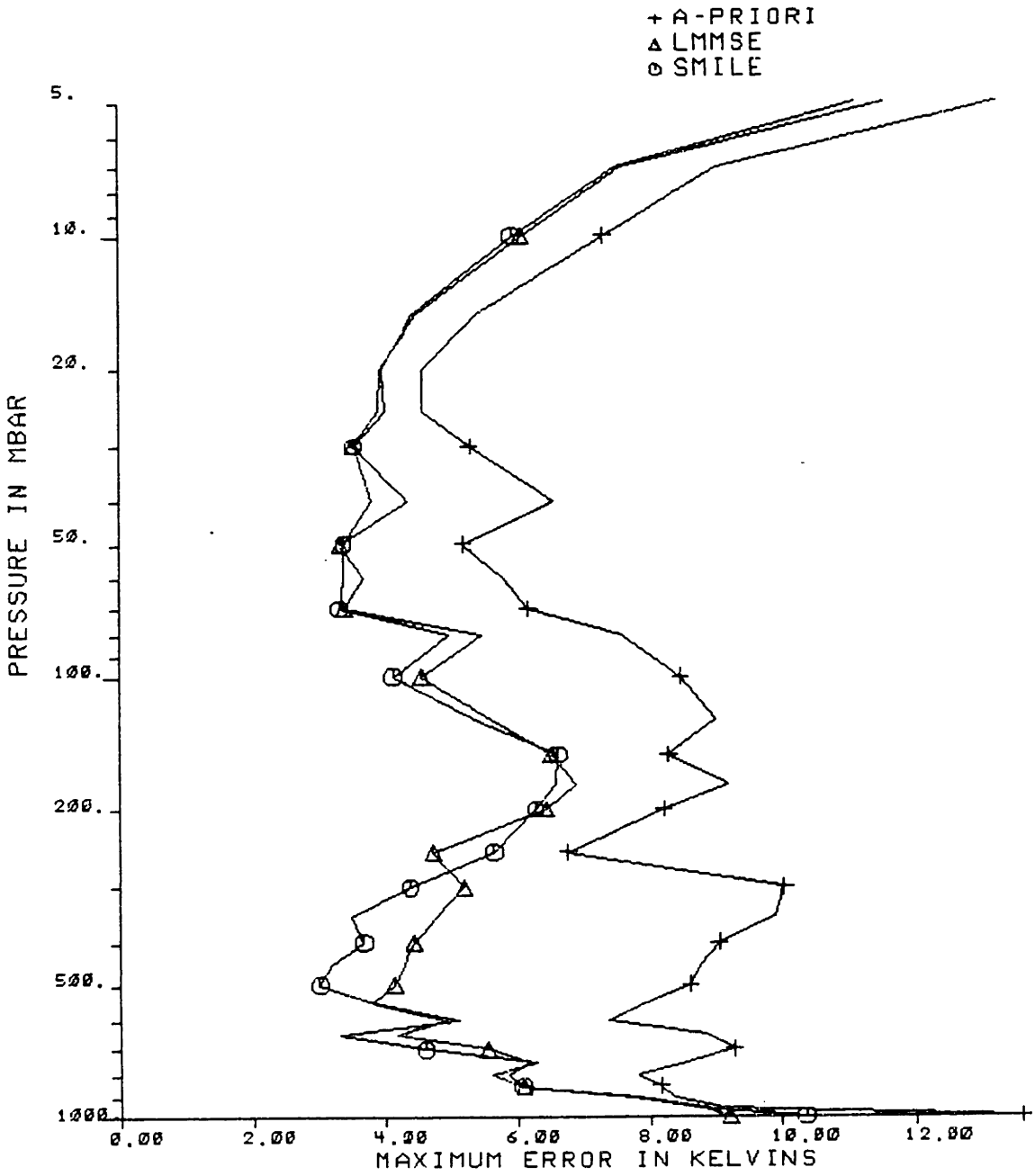


Figure VIII.3.2: Worse error for the LMMSE, the SMILE, and the a-priori distribution of that error

estimation of the observable components is the multiplication of these estimates by R_{oo}, the covariance matrix of the observable components.

In cases of bad SNR, this estimation shows that the inclusion of bound information significantly alters the behavior of the retrieval operator. This is true

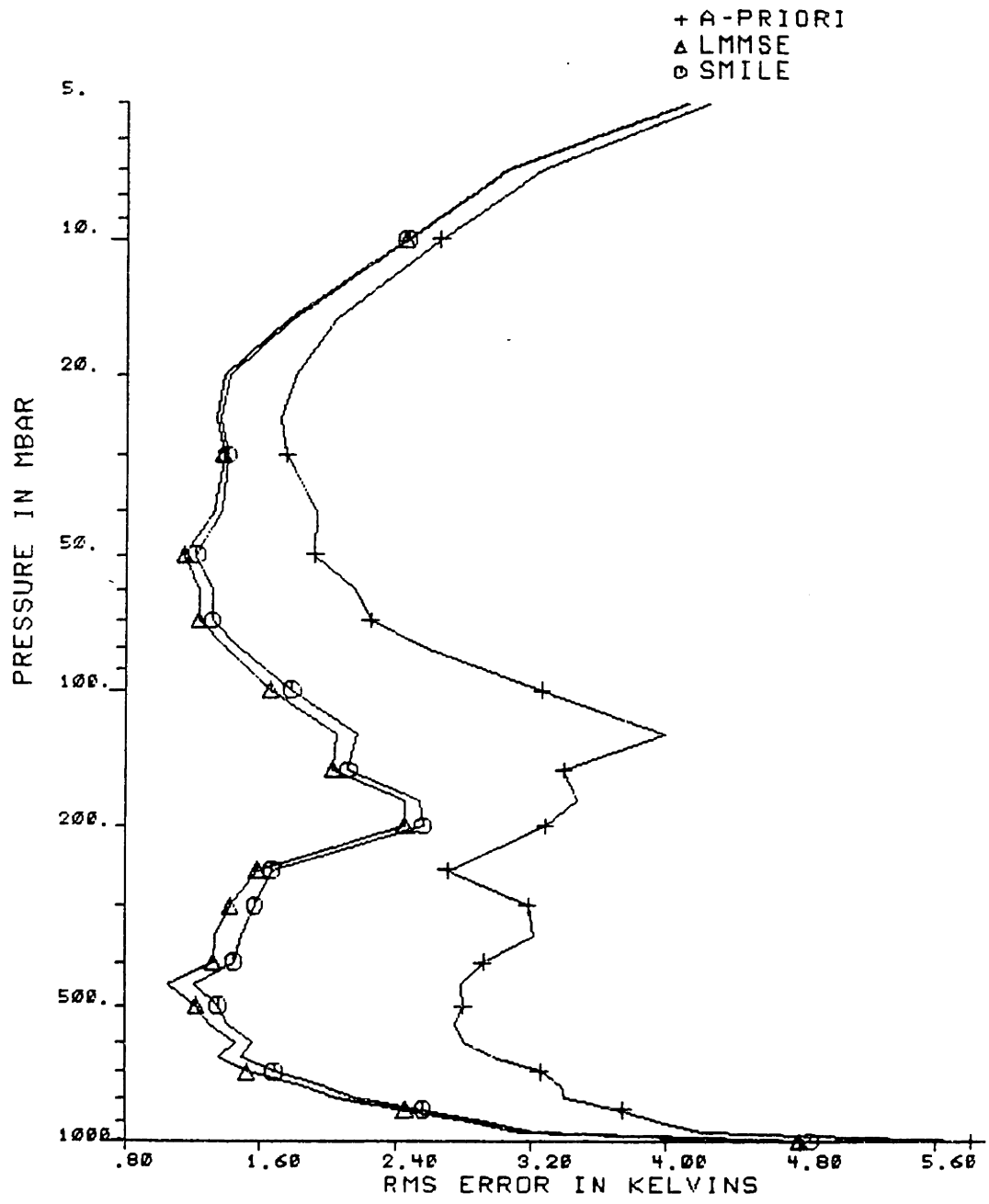


Figure VIII.3.3: Rms error for the LMMSE, the SMILE, and the a-priori distribution of that error

for two reasons: (1) the criterion optimized is not the mean square error but the worst square error, (2) bound information, although not equivalent to information about higher order moments of the pdf characterizing temperature profiles, is a significant auxiliary information about the pdf in addition to the first and second

moments.

Let us test the SMILE for MSU over the Peoria dataset. As noted, the SMILE will outperform the LMMSE in cases where the signal to noise ratio is bad, that is in cases where the receiver noise power is large. Using the MSU instrument with its nominal noise level does not provide any improvement. However, if the rms noise level is multiplied by 5, the usage of SMILE over the Peoria dataset makes a difference. Figure VIII.3.2 presents the worse case errors as a function of pressure for the LMMSE, SMILE as well as the a-priori worst error. The brightness temperatures used as inputs for these two linear filters are computed from the Peoria dataset 100 temperature profiles using weighting matrices computed above land for the US standard atmosphere. A Gaussian random noise is added to these measurements to simulate the MSU sensor noise (with the increased power). These retrievals are performed 20 times (with different noise vectors). The SMILE outperforms the LMMSE at all pressure levels, by up to 1 K at 500 *mbar* (out of a worst error of 4 K for the LMMSE). The only exception is at 250 *mbar*, around the tropopause inversion, where the LMMSE is better by about 0.8 K.

Figure VIII.3.3 compares the mean square error for both retrieval operators. As expected, the LMMSE outperforms the SMILE.

In cases where the measurements are of limited quality, the SMILE appears as a adequate alternative to the LMMSE. As for the other linear estimators derived in this chapter, the new estimator is computed as the product of the LMMSE by a scaling coefficient on its left hand side (VIII.3.28). This fact means that a series of temperature profiles estimated under the mean square error criterion can be re-estimated under the worst square error criterion (provided that the a-priori mean profile is still available after estimation).

One may argue that, because the SMILE is a simple scaled version of the LMMSE, its application to remote sensing is limited. However there are instances

where the worst error is an adequate criterion. When inserting satellite data into forecasts, the difference between the model forecast values and the radiance inverted values for the temperature profiles is used as input for an acceptance test for the satellite data (Halem *et al.*, 1978). Should the difference between those two values be larger than a given threshold (pressure dependent), the satellite data is discarded. In instances where the signal-to-noise ratio is bad, the usage of the SMILE to perform the radiance inversions with the bounds of the forecast field to compute γ in equation VIII.3.28 will reduce the expected value of the worst error, thus reducing the number of profiles being rejected.

Another field of application of the SMILE is in cases of non-linear observation equations where the bounds on the values of the geophysical parameters can eliminate divergences in the retrieval. Such a case is the remote sensing of atmospheric water vapor in opaque spectral bands (Komichak, 1982).

VIII.4 Conclusions

In this chapter several linear retrieval operators were introduced and compared with each other. In most cases, these filters are the product of the LMMSE (or its limiting form in the case of non information the MIE) with a correction matrix or a correction scalar. This form of estimators shows that the LMMSE forms the kernel of most retrieval operators derived in a Bayesian framework. Other estimates for the temperature based on different criterions or auxiliary information can be derived from the LMMSE-estimated field. Hence, its is possible to use multidimensional constrained retrieval operators, by first performing the multidimensional retrieval based on the Wiener filter (LMMSE), and then post-process the estimated temperature profiles.

Three concepts introduced in this chapter should be extended to the non-linear remote sensing of water vapor: usage of information theory concepts, the introduction of roughness measure, and the bounds on the range of possible values

of the geophysical parameters.

CHAPTER IX
CONCLUSIONS

"Free at last, free at last, thank God almighty, I am free at last." Reverend Martin Luther King., Jr.

IX.1 Summary of Main Results and Contributions

This section presents, chapter by chapter, the major results presented in this dissertation. It is an exhaustive recapitulation and expansion of the conclusions for each chapter. Section IX.2 extracts from these contributions the most significant ones and assesses their relevance to present future work in remote sensing and multidimensional, multichannel signal processing. Section IX.3 suggests possible extensions to this work.

Chapter II

In this chapter the principles of radiative transfer, which relate measured brightness temperature and vertical distribution of temperature were briefly exposed for the sake of completeness. The impact of non-linearities in the radiative transfer equation were examined from a forward-problem point of view and a correction method proposed by Rosenkranz wastested. Using bounds criterion on the error (typically 0.3 K) in computing the brightness temperature, deviations up to 10 K between the profile used to compute the weighting functions and the actual profile were found to be acceptable.

Chapter III

The proposed model for the three-dimensional covariance developed in this chapter is novel, although it relies on previous work in data assimilation techniques and geofluid dynamics. Incorporated in the model's development were horizontal covariance and its corresponding spectrum, geostrophic conditions, and mathematical conditions such as positive definiteness and moment constraints. The redundancy

of the geostrophic conditions, essential for estimating the wind field from the temperature field, as well as their incompatibility with modeled spectral decay laws were proven.

The vertical covariance kernel could not be analytically modeled because of non-stationarities across the tropopause inversion, and it required discrete temperature profiles in order to use a covariance matrix for representation. A Nyquist rate criterion was used to determine that 33 levels or so are required to represent profiles between the ground and 5 *mbar* which keep the aliasing energy to less than 1 % of the total energy. This number of levels is much larger than the NMC 14 mandatory pressure levels used in previous work.

The proposed model is the convolutional combination of a model for the vertical covariance kernel (based on a sum of outer products) and of a series of isobaric covariance kernels modeled as damped cosine functions. In terms of decay rate, the isobaric spectrum associated with the isobaric model is a compromise, between geostrophic and inertial spectral subranges.

Chapter IV

Several regression methods were derived to characterize the isobaric poles characterizing the horizontal correlation and the vertical covariance kernel. The estimation of the correlation structure proved to be not very robust, mostly because the near-field correlation was nearly isotropic. The isobaric correlation was modeled as isotropic from then on.

The model's ability to fit observed covariances was measured using normalized χ^2 tests. The range of validity of the model was found to be typically a pillbox of 2500 to 3000 *km* radius and up to 10 *mbar* in height.

In order to use optimally the NMC observational data, which typically possess gaps in the vertical coverage of temperature profile, a novel unbiased estimator guaranteeing the positive definiteness of the estimated covariance was introduced

for vectorial random process observed under conditions varying from sample to sample. This estimator was characterized and compared to Bayesian-type estimators, and a Picard-type iterative algorithm was introduced to solve it. This iterative method was proven to converge to the optimal solution for the mean vector and the covariance matrix.

Chapter V

The main goal of this chapter was to understand the impact of measurements and of atmospheric statistics on the retrieval error. This feat was accomplished by explicitly solving the two-point retrieval problem and by projecting temperature profiles and their estimates along the row subspace ("observable"), and kernel subspace ("unobservable") associated with the weighting matrices. This decomposition was designated as the "SVD decomposition" because of its intimate relationship with the singular value decomposition of the weighting matrices.

The principal contributions and results of this analysis are presented below:

For the single-spot problem, the estimation can be analyzed as a two-step process: (1) estimating the observable components (this estimation depends on the SNR), then (2) mapping these components onto the unobservable components (this mapping is independent of the SNR).

As the SNR increases, the lower bound for the observable components is given by the sensor noise scaled by the weighting matrix singular values, whereas the lower bound for the unobservable components depends solely on the vertical covariance kernel.

The improvement in the retrieval error using a multi-spot rather than a single-spot retrieval operator was found to be of the form $trace\{CKC^t\}$, where K is a matrix that is a function solely of the cross-covariance of observed components, and C is a coupling matrix which differs for observable and unobservable components.

The matrix K is a quadratic function of the cross-covariance between points, and is independent of the single-spot retrieval error for cases of weak correlation between points, but varies inversely with that error for strongly correlated fields. This analysis shows that for systems with fine sampling grids (such as AMSU, or airborne systems), the reduction of the error for the single-spot case is essential to improvements through the incorporation of horizontal correlation.

The coupling matrix for the observable components is a matricial SNR for the single-spot sounding process and goes to zero in cases of perfect measurements.

The coupling matrix for the unobservable components depends upon the barotropicity of the atmospheric field being estimated and the inner product between the SVD basis vectors. It goes to zero for purely two-dimensional flow with characteristics that do not vary in the vertical direction. Use of measurements performed at different sounding angles was also characterized, and it was found to improve for systems with few sounding channels (such as MSU) and systems for barotropic climatologies. This analysis also proved that limb correction methods for such limited instruments necessarily degrade performance.

The analysis demonstrated that retrieval operators are low-pass filters for the observable components reducing retrieval error by smoothing sensor noise, and that they are all-pass filters for the unobservable components. For a given region of support, retrieval operators spread more in the spatial domain for the unobservable components than for the observable components.

The design equations (Yule-Walker equations) for three-dimensional retrieval operators were analyzed as an extended eigenvalue problem which was solved by introducing a block Karhunen-Loève transformation. This analysis proved that the limited-support retrieval operators tend towards the type of Wiener filter operators introduced by Rosenkranz when the support along the track of the satellite becomes long compared to the atmospheric scale of fluctuations. This analysis also provided a novel method for designing retrieval operators for sounders with fine sampling

grids.

Appendix V.A examined the Kalman filter in the SVD basis and in the modal representation associated with the filter transition matrix. Ledsham's implementation of Kalman filtering, with different transition matrices for the off-line computation of the Kalman gains and for the on-line recursion of a-priori mean, was examined in that context: his specific implementation considered only the estimation of observable components and those unobservable components with the slowest variations in the vertical direction, thus avoiding unsolvable system identification problems and instabilities, while capturing most of the improvement provided by horizontal correlation.

Chapter VI

Chapter VI present the quantitative analysis of the impact of multidimensionality on the retrieval operators and on the retrieval errors. These analyses were performed on simulated and real data (after correction).

The improvement for MSU is provided by a reduction in the unobservable space retrieval errors and makes good use of the different sounding conditions in the direction perpendicular to the satellite's track. Retrieval errors computed from simulated and real data showed that improvements in the retrieval were concentrated in the lower troposphere (up to 20 %) and in the lower stratosphere (up to 15 %). All things considered, this improvement was found to be larger when the spots used were in the direction perpendicular to the satellite rather than in the parallel one. Most of the improvement was obtained by going from one- to two-dimensional retrievals. For implementation on real data, a good fraction of this improvement was provided by a reduction in the bias error. Implementation on real data proved that the use of multidimensional operators is not always profitable when the statistics used to derive the filters are different from the sample statistics.

For AMSU, the improvement was found to be much smaller (up to 8 %)

and was mostly the result of low-pass filtering in the observable subspace of the instrument. For AMSU the inclusion of only a few points in the retrieval process was found to provide most of the improvement.

The main conclusion of these simulations and implementations is that three-dimensional retrievals are useful for MSU and can include entire frames of data, whereas for AMSU, it suffices for them to be implemented with small regions of support (3×3 or 5×5). This result makes use of "generic retrievals" perfect for AMSU (as far as temperature is concerned).

Chapter VII

As was the case for chapter V, this chapter is analytical in nature and, for the most part, novel. Its main thrust was to characterize the impact of error in the retrieval operator on the retrieval error. Most of this analysis was performed in the single-spot (D-matrix) framework. The penalty for misspecifying the D-matrix was found to be proportional to the Euclidian distance between the exact and the assumed D-matrices. Because this penalty function is difficult to minimize, "sample-consistent" retrieval operators were considered. These operators are such that the brightness temperature covariance matrix associated with the temperature covariance matrix they are derived from equals the observed brightness temperature covariance matrix. For such operators, the penalty function was found to be proportional to the Euclidian distance between the datasets' temperature covariance matrices and the covariance matrices used to compute the retrieval operators.

Optimal adaptation operators were derived for sample-consistent retrieval operators for both the Fisher and Bayesian frameworks Adaptation was found always to be beneficial for the observable components, but adaptation sometimes degraded performance for the unobservable components. A condition for the average improvement of rms error for the unobservable components was given. No condition could be found to guarantee error improvement for these components for each dataset considered. The adaptation operators were characterized, simulated

and implemented on MSU, and simulated for AMSU. Adaptation of the D-matrix alone was found to have little effect for MSU, even when using all available sounding angles, and provided improvements up to 10 % for AMSU, which is characterized by a large observable subspace. This result makes adaptation a viable solution for AMSU.

For MSU, adaptation worked best through the replacement of the a-priori mean on a latitudinal and longitudinal basis. It also performed better for large sounding angles, where the ground contribution to the measurements is lessened.

In conjunction with this analysis, two measures were introduced in order to characterize how specific to a certain air mass, or how general, retrieval errors are A measure of specificity for vertical covariance matrices can be used to characterize climatologies. A measure of robustness of the optimal retrieval errors to these covariance matrices was introduced which can be computed solely using the retrieval operator.

Chapter VIII

This chapter investigated the question: what if linear retrieval operators were not optimized using the minimization of the mean-square error as a criterion? The first approach recasts the remote sensing equation as a communication problem through a noisy and limited channel. A analytical relationship between average remaining uncertainty and mean-square error was found.

Several linear retrieval operators were introduced, using criteria such as maximum estimated power or maximum likelihood estimation of observable components, as well as operators performing constraint estimation with bounds on the roughness of the profiles or their range of values. An estimator minimizing the expected maximum square error was introduced. In most cases, these filters are the product of the LMMSE with a correction matrix or a correction scalar.

IX.2 Moral and Consequences of this research

The results of this research fall in four major categories:

(1) Atmospheric Statistical Model based on Physics and Observations

(2) Multidimensional Retrievals

(3) Adaptation of Operators

(4) Auxiliary Information

For each one of these categories, we will present the main points derived in this work. The relevant sections or appendixes associated with each one of these points is given in italics.

Model derived from Physics and Observations

- This model was essential to ascertain the quantitative impact of horizontal correlation (*IV.6.1, VI.3*). It should be used for the design of multidimensional retrieval operators and for data assimilation (*IV.6.3*).
- The background information about the spectral behavior of isobaric temperature fields (*III.2.2*) and geostrophic conditions (*III.2.4*) is quite complete and worth reading, although it is mostly of academic interest.
- The vertical discretization (*III.3.1*) should be re-analyzed in the context of numerical weather prediction because NWP models use only about 10 levels.
- The maximum likelihood estimate for the first and second moment from incomplete and sample varying observations (*IV.B*) could profitably be incorporated in present data assimilation systems and extended to other fields of signal processing (array processing, tracking systems).

Multidimensional Retrievals

- Systems with few channels (MSU) should incorporate horizontal correlation as soon as possible (*VI.3, VI.4.3*). Operators should be extended in the direction perpendicular to the satellite's track to make use of the lifting of the weighting functions with view angle (*V.3.5, VI.4.3*) and should be incorporate limb correction techniques (*V.3.1*). The three-dimensional statistics

used to derive these operators should be derived from observations, not from ad-hoc modeling (*V.3.4, VII.2.1*). Improvements up to 15 to 20 percent should be expected, except at the tropopause.

- For systems with numerous channels (AMSU), multidimensional retrievals are worth using only when the channel bandwidths are small and the measurements noisy. Retrieval operators with small regions of support should provide most of the improvement (*VI.5*).

Adaptation

- For systems with numerous channels, adaptation of the second moment based on observed brightness temperature covariance matrices and climatological descriptions can be trusted, and should be part of any operational system (*VII.5.2*).
- For systems with few channels, the adaptation should concentrate on the a-priori mean (*VII.4.7*) and perform bias correction on a localized basis. A-priori mean profiles should be a function of both longitude and latitude (*VII.5.1*).
- Whenever performing adaptation, the variance statistics used should be derived from sets with similar numbers of degrees of freedom (*VII.4.2, VII.4.3, VII.5.1*).
- The Fisher and Bayesian analysis of the adaptation problem, as well as the penalty for using wrong operators (*VII.3*), should be extended to other fields dealing with non-stationary processes (adaptative filtering of images, speech, etc).

Auxiliary Information and Criteria

- Bound information (*VIII.3.*) should be used for non-linear observation problems, such as water vapor.
- Because the LMMSE is still the kernel of most constrained linear estimators,

these estimators will not produce extraordinary results (*VIII.2.2, VIII.2.3, VIII.3*). This result makes them more of academic interest than potential candidates for implementation.

IX.3 Suggestions for further work

A better method should be developed for modeling correlation structure across specific weather features (fronts and storms) , in order for multidimensional operators to perform better around such phenomena. Such a model should be developed for water vapor.

Multidimensional retrieval operators should be tested by means of airborne radiometers with numerous channels having weighting functions peaking around and at the tropopause. In this dissertation, when the tropopause was resolved, it was only through statistics. Systems that place this inversion more in the observable space will provide excellent information about how well it can be estimated. Since airborne systems typically have better spatial resolution and finer sampling grid, they are perfect candidates for this quantitative study. Retrieval operators which adapt to the topography and the presence of clouds should be implemented. These implementations are feasible for systems with small region of support. Multidimensional operators should also be investigated for water-vapor retrievals.

The impact of errors in the weighting functions - typically caused by nonlinearities, ground elevation, water vapor content, and ground emissivity- on the retrieval process should be studied.

Provided that the error behavior of numerically predicted weather is known, the direct merging of satellite and weather forecast data should be implemented and assessed in situations where the multidimensional retrieval schemes are known to perform badly (tropopause, inversion close to the ground). The possibility of analyzing weather prediction fields in order to flag for potentially critical fields

that could introduce large errors in inversion of brightness temperature should be investigated. In brief, the remote sensing of temperature should be more closely associated with the data assimilation process in order to better adapt to changing statistics.

An interesting avenue of research is to use the decomposition of the retrieval process into observable and unobservable subspaces and to extend it (if possible) to non-linear processing. The difficulties in expanding identification and detection techniques developed in scalar signal processing (such as speech and images) to vectorial processes may be dramatically reduced by distinguishing between observable and unobservable spaces. This approach would permit one to adapt retrieval operators in the horizontal as well as in the vertical direction. This decomposition can also be used for spectral analysis purposes, extraction of features, detection of waves in the temperature field, etc...

Another interesting possibility is to expand the numerous constrained estimation techniques developed originally in the Fisher framework to the Bayesian through the use of the mean as a pseudomeasurement.

Three concepts introduced in Chapter VIII should be investigated relative to the non-linear remote sensing of water vapor: usage of information theory concepts, the introduction of roughness measure, and the bounds on the range of possible values of the geophysical parameters.

In general, one should expand remote sensing past the linear mean-square framework if one desires to obtain further substantial improvements. Systems such as AMSU may provide enough measurements to allow expert systems techniques to be introduced into remote sensing. More channels also means that data can be checked for inconsistencies.

Last and not least, any further work should generate a thesis shorter than this one.

owner	acb
printer	charmin
host	newtowne-variety
Printed	Fri Aug 22 14:24:58 EDT 1986
name	/tmp/20759dvi

```
language imPRESS
jamresistance on
pagereversal on
jobheader on

IMAGEN Printing System, Version 2.2, Serial #84:6:27
Page images processed: 11
Pages printed: 11

Paper size (width, height):
  2560, 3328
Document length:
  51844 bytes
```

REFERENCES

A, N.Q., 1984: 'On the Uniqueness of the Maximum-Likelihood Estimate of Structure Covariance Matrices,' *IEEE Trans. Acoust., Speech, Signal Processing*, Vol 32, pp. 1249-1251.

Ahmed, N. and M. Flickner, 1982: 'Some Considerations of the Discrete Cosine Transform,' *16th Asilomar Conf. on Circuits, Systems, and Computers.*

Allan, T.D., 1983: *Satellite Microwave Remote Sensing*, John Wiley & Sons, New York.

Anderson, B.D.O. and J.B. Moore, 1979: *Optimal Filtering*, Prentice-Hall, Englewoods Cliffs, NJ.

Athans, M. and F.C. Schweppe, 1965: 'Gradient Matrices and Matrix Calculations,' Technical note 1965-53, Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, MA.

Arato, M., 1982: *Linear Stochastic Systems with Constant Coefficients* (Lectures Notes in Control and Information Sciences), Springer-Verlag, Berlin.

Baedem, A.P.M., W.J.A. Kuipers, and J. Reiff, 1976: 'An Interpolation Method, Based on Double Fourier Expansion, for the Calculation of Atmospheric Kinetic Energy Spectra, Compared with Bilinear Interpolation,' *J. Atmos. Sci.*, Vol. 33, pp. 764-767.

Backus, G. and F. Gilbert, 1970: 'Uniqueness in the Inversion of Gross Earth Data,' *Philos. Trans. R. Soc. Lond., Serv. A*, Vol. 266, pp. 123-192.

Baer, F., 1972; 'An alternate Scale Representation of Atmospheric Energy Spectra,' *J. Atmos. Sci.*, Vol. 29, pp. 649-664.

Baer, F., 1981: 'Three-Dimensional Scaling and Structure of Atmospheric Energetics,' *J. Atmos. Sci.*, Vol. 38, pp. 53-68.

Bauman, W.T., 1980: 'Linear Statistical Retrieval as Applied to Remote Sensing of Atmospheric Temperatures Profiles,' S.M. thesis, Massachusetts Institute of Technology, E.E.C.S. Department.

Basawa, I.V. and D.J. Scott, 1983: *Asymptotic Optimal Inference for Non-Ergodic Models*, Springer-Verlag.

Basili, P., P. Ciotii and D. Solimini, 1981: 'Inversion of Ground Based Radiometric Data by Kalman filtering,' *Radio Sci.*, pp. 83-91.

Batchelor, G.K., 1953: *Homogeneous Turbulence*, Cambridge University Press.

Beck, F.B, 1975: 'Antenna Pattern Correction to Microwave Radiometer Temperature calculations,' *Radio Science*, Vol. 10, pp. 839-845.

Beckenbach, E.F., and R. Bellman, 1965: *Inequalities*, Springer Verlag, Berlin.

Bekefi, G. and A.H. Barrett, 1977: *Electromagnetic, Vibrations, Waves, and Radiations*, The MIT Press, Cambridge, MA.

Bellman, R, 1960: *Introduction to Matrix Theory*, McGraw-Hill, New York.

Bengtsson, L., M. Ghil, and E. Kallen, (Editors), 1981: ' Dynamic Meteorology: Data Assimilation Method,' *Applied Mathematical Sciences*, Vol 367, Springer-Verlag,

Ben-Israel, A. and T.N.E. Greville, 1974: *Generalized Inverses: Theory and Applications*, Wiley and Sons, New-York.

Berg, C., J.P.R. Christensen, and P. Ressel, 1984 : *Harmonic Analysis on Semigroups. Theory of Positive Definite and Related Functions*, Springer-Verlag, New York.

Bergman, K.H., 1979: 'Multivariate Analysis of Temperature and Wind Using optimum Interpolation,' *Mon. Wea. Rev.*, Vol. 107, pp. 1423-1444.

Bertsekas, D.P, 1976: *Dynamic Programming and Stochastic Control*, Academic Press, new York.

Berztiss, A.T., 1964:, 'Least Square Fitting of Polynomial to irregularly spaced data,'*S.I.A.M. Rev.*, Vol 6, pp. 203-227.

Bevington, P.R, 1974: *Data Reduction and Error Analysis for the Physical Sciences*, McGraw-Hill, New York.

Bhat, B.R., 1974: 'On the Method of Maximum Likelihood for Dependent Observations,' *Jour. Royal Stat. Soc., Ser. B*, Vol. 36, pp. 48-53.

Bhattacharyya, A., 1947: 'On a Measure of Divergence between Two Statistical Population Defined by their Probability Densities,' *Bull. Math. Soc.*, Vol. 35, pp 99-109.

Bitmead, R.R. and B.D.O. Anderson, 1980: 'Asymptotically Fast Solution of Toeplitz and Related Systems of Linear Equations,' *Linear Algebra Appl.*, Vol. 34, pp. 103-116, 1980.

Blackmon, M.L., 1976: 'A Climatological spectral Study of the 500 mb Geopotential Height of the Northern Hemisphere,' *Jour. Atmos. Sci.*, Vol 33., pp 1607-1623.

Blackmon., M.L., R.A. Madden, J.M. Wallace, and S.L. Mullen, 1979: 'Geographical Variations in the Vertical Structure of geopotential Height Fluctuations,' *Jour. Atmos. Sci.*, Vol. 36, pp. 2450-2466.

Boer, G.J. and T.G. Sheperd, 1983: ' Large-Scale Two-Dimensional Turbulence in the Atmosphere', *Jour. Atmos. Sci.*, Vol. 40, pp. 164-184, January.

Boley, D., 1984: 'Computing the Kalman Decomposition: An Optimal Method', *IEEE Trans. Automat. Contr.*, Vol. 29, pp. 51-53.

Boltenkov, V.P., 1966: 'Some characteristics of the three-dimensional macrostructure of atmospheric temperatures,' *Trudy Glavnoi Geofizicheskoi Observatorii, imeni A. I. Voeikova*, Leningrad, No 191, pp. 47-57, (In Russian).

Brewer, J.W., 1978: 'Kronecker products, and matrix calculus in system theory, ' *IEEE Trans. Circuits and Sys.*, Vol. 25, pp. 772-781.

Brillinger, D.R., 1975: *Time Series: Data Analysis and Theory* Holt, Rinehart and Winston, New York.

Burg, J.P., D.G. Luenberger, and D.L. Wenger, 1982:' Estimation of Structured Covariance Matrices,' *Proceedings of the IEEE*, Vol. 70, pp. 963-974.

Buel, C.E., 1972: 'Correlation Functions for Wind and Geopotential of Isobaric Surfaces,' *Jour. Appl. Meteor.*, Vol 12, pp. 51-59.

Bulmer, M. G., 1967: *Principles of Statistics*, Dover Inc., New York.

Cadzow, J.A., 1979: 'An Extrapolation Procedure for Band-Limited Signals,' *IEEE Trans. Acoust., Speech and Signal Processing*, Vol. 27, pp. 4-21, Feb.

Carlson, D.H. and B.N. Datta, 1979:'The Lyapunov Matrix Equation SA + A*S = S*B* BS,' *Linear Algebra and Appl.*, Vol. 28, pp. 43-52.

Carnahan, B., H.A. Luther, and J.O. Wilkes, 1969 : *Applied Numerical Methods*, John Wiley & Sons, New York.

Chahine, M.T., 1969: 'Determination of the Temperature Profile in an Atmosphere from its Outgoing Radiance,' *J. Optical Soc. Amer.*, Vol. 58, pp. 1634-1637.

Chahine, M.T., 1970: 'Inserve Problem in Radiative Transfer: Determination of Atmospheric Parameters,' *J. Atmos. Sci.*, Vol. 27, pp. 960-967.

Chandrasekhar, S., 1950: *Radiative Transfer*, Dover Inc., New York.

Charney, J.G., 1971: 'Geostrophic Turbulence,' *Jour. Atmos. Sci.*, Vol. 28, pp. 1087-1095.

Charney, J.G., 1973: in *Dynamic Meteorology*, P. Morel (Ed.), D. Reidel Publishing Co., Boston.

Chen, T.C. and A. Wiin-Nielsen, 1978: 'Non-linear Cascades of Atmospheric Energy and Enstrophy in a Two Dimensional Spectral Index,' *Tellus*, Vol. 30, pp. 313-322.

Chernoff, H., 1962: 'A measure of Asymptotic Efficiency for Tests of a Hypothesis based on a sum of Observations,' *Annals. Math. Stat.*, Vol. 23, pp. 493-507.

Chernov, L.A., 1960: *Wave Propagation in a Random Medium*, Dover Publication, New York.

Childers, D.G., 1978: *Modern Spectral Analysis*, IEEE Press, New York.

Clarke, R.J, 1981: 'On the Relation between the Karhunen Loeve and Cosine Transforms,' *IEE Proc. F, Commun., Radar & Signal Process.*, Vol. 128, pp. 359-360.

Clarke, R.J. 1983: 'Performance of Karhunen-Loeve and Discrete Cosine Transforms for Data Having Widely Varying Values of Intersample Correlation Coefficient,' *Elec. Let.*, Vol. 19, pp. 251-252.

Colavita, M.M., 1985: 'Atmospheric Limitations of a Two-Color Astrometric Interferometer,' Ph.D. Thesis, Massachusetts Institute of Technology, EE&CS Department.

Commenges, D., 1984: 'The Deconvolution Problem: Fast Algorithms Including the Preconditioned Conjugate-Gradient to Compute a MAP Estimator,' *IEEE Trans. Auto. Cont.*, Vol.29, pp. 229-243.

Crowder, M.J., 1976: 'Maximum Likelihood Estimation for Dependent Observations,' *Jour. Royal Stat. Soc.*, Ser. B, Vol. 38, pp. 45-53.

Dartt, D.G., 1972: 'Automated Streamline Analysis Utilizing Optimum Interpolation,' *Jour. Appl. Meteor.*, Vol. 11, pp. 901-908.

Davisson, L.D., 1972: 'Rate-Distortion Theory and Application,' *Proc. IEEE*, Vol. 60., pp. 800-808.

Dempster, A.P., N.M. Laird, and D.B. Durbin, 1977: 'Maximum Likelihood from Incomplete Data via the EM algorithm', *J. Roy. Stat. Soc., Series B*, Vol. 39, pp. 1-38.

Desmarais, A., S. Tracton, R. McPherson, and R. Van Haaren, 1978: The NMC Report on the Data Systems Test, NOAA, National Meteorological Center.

Delsarte, P. and Y. Genin, 1985: 'The Split Levinson Algorithm,' Philips Research Laboratory, Brussels, Belgium.

Dongarra, J.J., C.B. Moler, J.R. Bunch, G.W. Stewart, 1979: *LINPACK Users' Guide*, Society for Industrial and Applied Mathematics. Philadelphia, PN.

Durbin, J., 1953: 'A Note on Regression when there is Extraneous Information about one of the Coefficients,'*Jour. Amer. Statist. Assoc.*, Vol. 48, pp. 799-808.

Erdey-Gruz, T., 1974: *Transport Phenomena in Aqueous Solutions*, John Wiley & Sons, New York.

Eskinazi, S., 1975: *Fluid Mechanics and Thermodynamics of our Environment*, Academic Press, New York.

Fiocco, G. (Editor), 1970: 'Mesospheric Models and Related Experiments,' *Proceedings of the Fourth Esrin-Eslab Symposium held in Frascati*, Italy, 6-10 July, D. Reidel Publishing Company, Dordrecht-Holland.

Frazer, R.A., W.J. Duncan, and A.R. Collar, 1947 : *Elementary Matrices and Some Applications to Dynamics and Differential Equations*, MacMillan Company, New York.

Fried, D.L., 1978: 'The Nature of Atmospheric Turbulence Effects On Imaging and Pseudo-Imaging Systems, and its Quantification,' *Proceedings of Colloquium no. 50 'High Angular Resolution Stellar Interferometry,'* International Astronomical Union, University of Maryland, College park, ML.

Frielander, B. and M. Morf, 1980: 'Efficient inversion Formulas for Sums of Products of Toeplitz and Hankel Matrices', *18 th annual Allerton Conference*, October 8-10.

Fymat, A.L., 1977: *Inversion Methods in Atmosphere Radiation Research, Volume I: Temperature Sounsings* International Association of Meteorology and Atmospheric Physics, Radiation Commission, JPL.

Gage, K.S., 1979: 'Evidence of a $k^{-5/3}$ Law Inertial Range in Mesoscale Two-Dimensional Turbulence,' *J. Atmos. Sci.*, Vol. 36, pp. 1950-954.

Gandin, L.S., 1965: *Objective Analysis of Meteorological Fields, Gidrometeor. Isdat, Leningrad.* (Israel Program for Scientific Translations, Jerusalem).

Gantmacher, F.R., 1959 : *The Theory of Matrices*, Vol. 1, Chelsea Publishing Company, New York.

Gautier, D. and I. Revah, 1975: 'Sounding of Atmospheric Atmospheres: A Fourier Analysis of the Radiative Transfer Equation,' *Jour. Atmos. Sci.*, Vol. 32, pp. 881-892.

Gelb, A. (Editor), 1974: *Applied Optimal Estimation*, Massachusetts Institute of Technology. Press, Cambridge, Ma.

Gill, P.E. and F.W. Murray, 1974: *Numerical Methods for Constrained Estimation*, Academic Press, New York.

Gnanadesikan, R., 1977: *Methods for data Analysis of Multivariate Observations*, John Wiley & Sons, Inc., New York.

Gradshteyn, I.S., and I.M. Ryzhik, 1980: *Tables of Integrals, Series, and Products*, Academic Press, New York.

Grody, N.C., D.G. Gray, C.S. Novak, J.S. Prasad, M. Peigrass, and C.A. Dean, 1985: 'Temperature Soundings From the DMSP Microwave Sounder,' *Interactive Workshop on Advances in Remote Sensing Retrieval Method*, Williamsburg, VA.

Gutzler, D.S. and K.C. Mo, 1983: 'Autocorrelation of Northern Hemisphere Geopotential Heights', *Mon. Wea. Rev.*, Vol. 111, pp. 155-164.

Halem, M., M. Ghil, R. Atlas, J. Susskind, and W.J. Quirk, 1978: GISS Sounding Temperature Impact Test, NASA Technical Memorandum 78063, Goddard Space Flight Center.

Hearon, J.Z., 1977: 'Nonsingular Solutions of TA-BT=C,' *Linear Algebra and Appl.*, Vol. 16, pp. 57-63.

Henderson, T.L., 1981: 'Geometric Methods for Determining the System Poles from Transient Response,' *IEEE Trans. Acoust., Speech, Signal Proc.*, Vol. 29, October.

Herring, J. R., 1980: 'Statistical Theory of Quasi-geostrophic Turbulence',*J. Atmos. Sci.*, Vol. 37, pp. 969-977, May.

Hoerl, A.E. and R.W. Kennard, 1970: 'Ridge Regression: Biased Estimation for Nonorthogonal Problems,' *Technometrics*, Vol. 12, pp. 55-67.

Holton, J.R., 1975: *An introduction to Dynamic Meteorology*, Academic Press, New York.

Holton, J.R., 1975: 'The Dynamic Meteorology of the Stratosphere and Mesosphere,' *Meteorological monographs*, Vol. 15, Number 37, American Meteorological Society.

Hunt, B.R., 1972: 'A Theorem of the Difficulty of Numerical Deconvolution,' *IEEE Trans. Audio Electroacoutst.*, Vol. 20, pp. 94-95.

Husain, T. M.A. Ukayli and H.U. Khan, 1986: 'Meteorological Network Expansion Using Information Decay Concept,' *J. Atmos. and Ocea. Tech*, Vol. 3, pp. 27-35.

Ibragimov, I.A. and R.Z. Has'minskii, 1981: *Statistical Estimation: Asymptotic Theory*, Springer-Verlag, Berlin.

Ibragimov, I.A. and Y.A. Rozanov, 1978: *Gaussian Random Processes*, Springer-Verlag, Berlin.

Jain, A.K., 1976: 'A Fast Karhunen Loeve Transform for a Class of Stochastic Processes,' *IEEE Trans. Commun.*, Vol. 24, pp. 1023-1029.

Jain, A.K., 1979: 'A Sinusoidal family of Unitary Transforms,' *IEEE Trans. on Pattern Analysis and Machine Intelligence,'* Vol. 1, pp. 356-365.

Jain, A.K. and S. Rangantath, 1981: 'Extrapolation Algorithms for Discrete Signals with Applications to Spectral Estimation', *IEEE Trans. Acoust., Speech and Signal Processing*, Vol. 29, pp. 830-845, Aug.

Jayant, N.S. and P. Noll, 1984: *Digital Coding of Waveforms Principles and Applications to Speech and Video*, Prentice-Hall, Englewood Cliffs, N.J.

Jaynes, E.T., 1982: 'On the Rationale of Maximum-Entropy Methods', *Proc. IEEE*, Vol. 70, pp. 939-952.

Jazwinski, A.H., 1970: *Stochatic Processes and Filtering Theory*, Academic Press, New York.

Julian, P.R., and A. Cline, 1974: 'The Direct Estimation of Spatial Wavenumber Spectra of Atmospheric Variables,' *Jour. Atmos. Sci.*, Vol. 31, pp. 1526-1539.

Julian, P.R., and H.J. Thiebaux, 1975: 'On some properties of Correlation Functions used in Optimum Interpolation Schemes,' *Mon. Wea. Rev.*, Vol. 103, pp. 605-616.

Kailath, T., 1976: *Lectures on Linear Least Squares Estimation*, Springer-Verlag.

Kalman, R.E., 1963: 'Mathematical Description of Linear Systems,' *SIAM J. Contr.*, Vol. 1, pp. 152-192.

Kalouptsidis, N., D. Manolakis, and G. Carayannis, 1983: 'A family of Computationally Efficient Algorithms for Multichannel Signal Processing - A Tutorial Review', *Signal Processing* , Vol. 5, pp. 5-19.

Kantor, A.J. and A.E. Cole, 1977: 'Monthly 90 ^{o} Atmospheres and High Latitudes Warm and Cold Winter Stratosphere Mesosphere,' Air Force Geophysics Laboratory, Hanscom AFB, MA.

Kao, S.K., 1970: 'The wavenumber-frequency Spectra of Temperatures in the Free Atmosphere,' *Jour. Atmos. Sci.*, Vol. 27, pp. 1000-1007.

Kasahara, A., 1977: *Computational Aspects of Numerical Models for Weather Prediction and Climate Simulation*, in *Methods in Computational Physics, Advances in Research and Applications*, Vol. 17: General Circulation Model of the Atmosphere, J. Chang (ed). Academic Press, New York.

Kassam, S.A., T.L. Lim, and L.J. Cimini, 1980: 'Two-Dimensional Filters for Signal Processing Under Modeling Uncertainties,' *IEEE Trans. Geosci. Remote Sensing*, Vol. 18, pp. 331-336.

Kitagawa, G. and W. Gersch, 1985: 'A Smoothness Priors Time-Varying AR Coefficient Modeling of Nonstationary Covariance Time Series,' *IEEE Trans. Automat. Contr.*, Vol. 30, pp. 48-56.

Kitagawa, G. and W. Gersch, 1985: 'A Smoothness Priors Long AR Model Method for Spectral Estimation,' *IEEE Trans. Automat. Contr.*, Vol. 30, pp. 57-65.

Klein, L.A and C.T. Swift, 1977: 'An Improved Model for the Dielectric Constant of Sea Water at Microwave Frequencies,' *IEEE Trans. Antennas Propagat.*, Vol. 25, pp. 104-111.

Koehler, T.L., J.C. Derber, B.D. Schmidt, and L.H. Horn, 1983: 'An Evaluation of Soundings, Analyses and Model Forecasts Derived from TIROS-N and NOAA-6 Satellite Data,' *Monthly. Wea. Rev.*, Vol. 111, pp. 562-571.

Kolmogorow, A.N., 1941: 'The Local Structure of Turbulence in Incompressible Viscous Fluids for very Large Reynolds Numbers,' *Dokl. Acad. Nauk. SSSR*, Vol. 30, pp. 299-303.

Kuks, J. and W. Olman, 1971: 'Minimax Linear Estimation of Regression Coefficients,' *Iswestija Akademija Nauk Estonskoj SSR*, Vol. 20, pp. 480-482. (In Russian)

Kuks, J. and W. Olman, 1972: 'Minimax Linear Estimation of Regression Coefficients II,' *Iswestija Akademija Nauk Estonskoj SSR*, Vol. 21, pp. 66-72. (In Russian)

Kunaresan, R. and D.W. Tufts, 1982: 'Estimating the Parameters of Exponentially Damped Sinousiodes and Pole-Zero Modelling in Noise,' *IEEE Trans. Acoust., Speech, Signal Proc.*, Vol. 30, pp. 833-840, December.

Landweber, L. 1951: 'An Iteration Formula for the Fredholm Integral Equation of the First Kind,' *Am. J. Math.*, Vol. 73, pp. 615-624.

Lauter, H, 1975: ,'A Minimax Linear Estimator for Linear Parameters under Restrictions in Form of Inequalities,' *Math. Operationsforschung Statistik*, Vol. 6, pp. 689-696.

Ledsham, W.H. and D.H. Staelin, 1981: 'An extended Kalman-Bucy Filter for Atmospheric Temperature Profile Retrieval with a Pasive Microwave Sounder,' *J. App.l. Met.*, 17, pp. 1023-1033.

Ledsham, W.H., 1978: Optimum Retrieval Techniques in Remote Sensing of Atmospheric Temperature, Liquid Water and Water Vapor, Ph.D. Thesis, Massachusetts Institute of Technology, E.E.C.S. Department.

Leith, C.E. 1973: 'The Standard Error of Time-average Estimates of climatic Means', *Jour. Appl. Meteor.*, Vol. 12, pp. 1066-1069.

Leith, C.E. 1982: 'Statistical Methods for the Verification of Long and Short Range Forecasts,' *Seminar 181 - Problems and Prospects in the Long and Medium Range Weather Forecasting*, European Center for Medium Range Weather Forecasts, pp. 313-335.

Lenhard, R.W., 1970: 'Accuracy in Radio-sondes Temperature and Pressure Height Determination,' *Bull. Am. Met. Soc.*, Vol. 51, pp. 842-846.

Lenhard, R.W., 1973: 'A Revised Estimate of Radio-sondes Accuracy,' *Bull. Am. Met. Soc.*, Vol. 54, pp. 691-693.

Levinson, N., 1947: 'The Wiener Rms (Root Mean Square) Error Criterion in Filter Design and Prediction,' *Jour. Math. Phys.*, Vol. 25, pp. 261-278.

Levy, B.C. and J.N. Tsitsiklis, 1982: 'Linear Estimation of Stationary Stochastic Processes, Vibrating Strings, and Inverse Scattering,' Laboratory for Information and Decision Systems and the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology.

Levy, B.C. and J.N. Tsitsiklis, 1985: 'A Fast Algorithm for Linear Estimation of Two-dimensional Isotropic Random Fields,' *IEEE Trans. Inform. Theory*, Vol. 31, pp. 635-644, September.

Lilly, D.K., 1983a: 'Some Facets of the Predictability Problem for Atmospheric Mesoscales,' *Proceedings of the American institute of Physics*, number 106, 'Predictability of Fluid Motion'. G. Holloway and B.J. West (Ed.), American Institute of Physics.

Lilly, D.K., 1983b: "Stratified Turbulence and the Mesoscale Variability of the Atmosphere, *Jour. Atmos. Sci.*, Vol.40, pp. 749-761.

Lin, J.T., and Y.H. Pao, 1979: 'Wakes in Stratified Fluids,' *Annual Review of Fluids Mechanics*, Vol. 11, Annual Reviews, pp 317-338.

Liou, K.N. 1980: *An Introduction to Atmospheric Radiation* Academic Press, Orlando, FL.

Lorenc, A.C., 1981: 'A Global Three-dimensional Multivariate statistical interpolation scheme,' *Mon. Wea. Rev.*, Vol. 109, pp. 701-721.

Lorenz, E.N., 1969: 'The Predictability of a Flow which Possesses many Scales of Motion,' *Tellus*, Vol. 29, pp. 289-307.

Luenberg, D.G., 1976: *Optimization By Vector Space Method*, John Wiley & Sons, New York.

Lumley, J.L., 1964: 'The Spectrum of Nearly Inertial Turbulence in a Stably Stratified Fluid,' *J. Atmos. Sci.*, Vol. 21, pp. 99-102, January.

Makhoul, J., 1981: 'On the Eigenvectors of Symmetric Toeplitz Matrices,' *IEEE Trans. Acoust., Speech and Signal Processing*, Vol. 29, pp. 368-372.

Matusita, K, 1965: 'A Distance and Related Statistics in Multivariate Analysis,' in *Multivariate Analysis*, Proceedings of an International Symposium held in Dayton, Ohio, June 14-19, 1965, Paruchuri R. Krishnaiah, (Ed.), Academic Press, New York.

Maymin, Z.G., 1981: 'Minimax Estimation on Subsets of Parameters', Ph.D. Thesis, Massachusetts Institute of Technology, Department of Mathematics, May.

Malvar, H.S., 1986: 'Optimal Pre- and Post-Filters in Noisy Sampled-Data Systems,' Ph.D. Thesis, Massachusetts Institute of Technology, E.E.C.S. Dept., August.

McLachlan, N.W., 1955: *Bessel Functions for Engineers*, Clarendon Press, Oxford.

McMillin, L.M. 1978: 'An improved Technique for Obtaining Clear-Column Radiances from Cloud-Contamined Radiances', *Mon. Wea. Rev.*, Vol. 106, pp 1590-1597.

McPherson, R.D., K.H. Bergman, R.E. Kistler, G.E. Rash, and D.S. Gordon, 1979: 'The NMC Operational Global Data Assimilation System,' *Mon. Wea. Rev.*, Vol. 207, pp. 1445-1461.

Mook, D.R., 1983: 'The Numerical Synthesis and Inversion of Acoustic Fields Using the Hankel Transform With Application To The Estimation of The Plane Wave Reflection Coefficient of The Ocean Bottom.', Sc.D Thesis Massachusetts Institute of Technology/ Woods Hole Oceanog. Inst.

Musicus, B.R., 1981: 'Levinson and Fast Choleski Algorithms for Toeplitz and Almost Toeplitz Matrices,' Research Laboratory of Electronics, Massachusetts Institute of Technology.

Nathan, K.S., 1983: Application of a Multidimensional Spatial Filter to Temperature Profiles Retrievals, S.M. Thesis, Massachusetts Institute of Technology, E.E.C.S. Department.

Neggi, J.G. and V.P. Dimri, 1984: 'On a generalization of Maximum Entropy Method for Processing of Nonstationary Multichannel Complex Valued Data,' *IEEE Trans. Geos. & Remo. Sens.*, Vol. 22, pp. 461-465.

Njoku, E.G., E.J. Christensen, and R.E. Cofield, 1980: 'The Seasat Scanning Microwave Radiometer (SMMR): Antenna Patterna Corrections- Development and Implemetation,' *IEEE J. Oceanic Eng.*, Vol. 5, pp. 125-137.

North, G.R, 1975: 'Theory of Energy-Balance Climate Models,' *J. Atmos. Sci.*, Vol. 32, pp. 2033-2043.

North, G.R and R.F. Cahalan, 1981: 'Predictability in a Solvable Stochastic Climate Model,' *J. Atmos. Sci.*, Vol. 38, pp. 504-513.

Nuttall, A.H., 1958: 'Theory and Application of the separable Class of Random Processes,' Technical report 343, Research Laboratory of Electronics, Massachusetts Institute of Technology., Cambridge.

Oort, A.H. and E.M. Rasmusson, 1971: *Atmospheric Circulation Statistics*, NOAA professional paper 5, National Oceanic and Atmospheric Administration, US department of Commerce, Rockville, MD.

Oppenheim A.V. and R. Shaffer, 1975: *Digital Signal Processing*, Prentice-Hall, Englewood Cliffs, New Jersey.

Ortega, J.M. and W.C. Rheinboldt, 1970: *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York.

Palmen, E. and C.W. Newton, 1969: *Atmospheric Circulation Systems. Their Structure and Physical Interpretation*, Academic Press, New York.

Panchev, S., 1971: *Random Functions and Turbulence*, Pergamon Press, Oxford.

Papoulis, A., 1975: 'A New Algorithm in Spectral Analysis and Bandlimited Signal Extrapolation,' *IEEE Trans. Cir. Sys.*, Vol. 22, pp. 735-742, Sept.

Papoulis, A., 1965: *Probability, Random Variables and Stochastic Processes*, McGraw-Hill, New York.

Parks, T.W. and R.G. Meier, 1971: 'Reconstruction of Signals of a Known Class From a Given Set of Linear Measurements,' *IEEE Trans. Info. The.*, Vol. 17, pp. 37-44, Jan.

Parsen, E., 1962: 'Spectral Analysis of Asymptotically Stationary Time Series,' *Bull. Inst. Internat. Statist.*, Vol. 39, livraison 2, pp 87-103.

Prata, A.J., 1984: 'An Appraisal of Retrieval Methods used for Obtaining Temperature Profiles from TOVS Radiances', The Technical Proceedings of the First International TOVS Study Conference, W.P. Menzel (Ed.), Cooperative Institute for Meteorological Satellite Studies, Space Science and Engineering Center, University of Wisconsin, WI.

Penk, L., 1965: 'A Modeling Study of the Meridional Temperature Profile and Energy Transformations of the Lower Stratosphere,' Report 13, Departement of Meteorology, Planetary circulation project, Massachusetts Institute of Technology, April 1965.

Petersen, D.P., 1973: 'Static and Dynamic Constraints on the Estimation of Space-Time Covariance and Wavenumber Frequency Spectral Fields,' *J. Atmos. Sci.*, Vol. 30, pp. 1252-1266.

Petersen, D.P., and D. Middeleton, 1964: 'Reconstruction of Multidimensional Stochastic Fields from Discrete Measurements of Amplitude and Gradient,' *Information and Control*, Vol. 7, pp. 445-476.

Petiau, G., 1955: La Theorie des Fonctions de Bessel Exposee en Vue de ses Applications a la Physique Mathematique, Centre National de la Recherche Scientifique, Paris, (In French).

Phillips, N.A., 1982: 'On the Completeness of Multivariate Optimum Interpolation for Large-scale Meteorological Analysis,' *Mon. Wea. Rev.*, Vol. 110, pp. 1329-1334.

Phillips, N., L. McMillin, A. Gruber, and D. Wark, 1979: 'An Evaluation of Early Operational Temperature Soundings from TIROS-N,' *Bull. Amer. Meteor. Soc.*, Vol. 60, pp. 1195-1201.

Pokrosky, O.M., 1969: 'Optimal Conditions of Indirect Sensing of the Atmosphere,' *Izv. Acad. Sci. CCCP, ser. fiz. atm. i okeana*, Vol. 5, pp. 1324-1326. (In Russian)

Pond, S., S.D. Smith, P.F. Hamblin, and R.W. Burling, 1966: 'Spectrum of Velocity and Temperature Fluctuations in the Atmospheric Boundary Layer Over the Sea,' *J. Atmos. Sci.*, Vol. 23, pp. 376.

Poon, R.K.L., 1974: Atmospheric Opacity near Half Centimeter Wavelength, Sc.D. Thesis, Massachusetts Institute of Technology, EECS department.

Pratt, W.K., 1978: *Digital Image Processing*, John Wileys & Sons, New York.

Priestley, C.H.B, 1959: *Turbulent Transfer in the Lower Atmosphere*, The University of Chicago Press, Chicago.

Rao, C.R., 1973: *Linear Statistical Inference and Its Applications*. 2nd Ed. John Wiley & Sons, London.

Ramanathan, Y., P. Kulkarni, and D.R. Sikka, 1972: 'A Comparative Study of Fourier Analysis Procedure and Cressman's Method in Objective Analysis of the Wind field,' *Jour. Appl. Meteor.*, Vol. 11, pp. 1318-1321.

Rife, D.C., and R.R. Boorstyn, 1974: 'Single Tone Parameter Estimation from Discrete-Time Observations,' *IEEE Trans. Inform. Theory*, Vol. 20, September.

Rinne, J., and S. Jarvenoja, 1985: 'Autocorrelation Functions Computed from Daily 500 mb Geopotential Height Analyses,' *Mon. Wea. Rev.*, Vol 113, pp. 1681-1691, October.

Robinson, E.A. and S. Treitel, 1978: Digital Signal Processing in Geophysics, in it Applications of Digital Signal Processing, A.V. Oppenheim, (Ed.), Prentice-Hall, Englewood Cliffs, NJ.

Rodgers, C.D., 1976: 'The Vertical Resolution of Remotely Sounded Temperature Profiles with a priori Statistics,' *Jour. Atmos. Sci.*, Vol 33., pp.707-709.

Rosenkranz, P.W., 1975: 'Shape of the 5mm Oxygen Band in the Atmosphere,' *IEEE Trans. Anten. and Prop.*, Vol. 23, pp.498-506.

Rosenkranz, P.W., 1978:'Inversion of Data from diffraction-Limited Multiwavelength Remote Sensors, 1. Linear Case, *Radio Sci.*, Vol. 13, pp. 1003-1010.

Rosenkranz, P.W., 1982: 'Inversion of Data from diffraction-Limited Multiwavelength Remote Sensors, 2. Nonlinear Dependence of Observables on the geophysical Parameters,' *Radio Sci.*, Vol. 17, pp. 245-256.

Rosenkranz, P.W. and W.T. Bauman, 1980: 'Inversion of Multiwavelength Radiometer Measurememt by Three-dimensional Filtering,' in *Remote Sensing of Atmosphere and Oceans, Deepak (Editor)*, Academic Press, 1980.

Rosenkranz, P.W., M.J. Komichak, and D.H. Staelin, 1982: 'A Method for Estimation of Atmospheric Water Vapor Profiles by Microwave Radiometry,' *Jour. Appl. Meteo.*, Vol. 21, pp. 1364-1370.

Rosenkranz, P.W., K.S. Nathan, and D.H. Staelin, 1984: 'Use of Two- and Three- Dimensional Spatial Filtering for Inversion of Radiometric Measurements,' *Interactive Workshop on Advances in Remote Sensing Retrieval Method*, Williamsburg, VA..

Rosenkranz, P.W. and D.H. Staelin, 1984: 'Experimentals Studies of the Evolution of Errors in a Numerical Weather Prediction Model,' *Submitted for Publication.*

Rutherford, I.D., 1972: 'Data Assimilation by Statistical Interpolation of Forecast Error Field,' *Jour. Atmos. Sci.*, Vol. 29, pp. 809-815.

Sabri, M.S. and W. Steenaart, 1978: 'An approach to Band-Limited Signal Extrapolation: The Extrapolation Matrix,' *IEEE Trans. Cir. Sys.*, Vol. 25, pp. 74-78, Feb.

Schafer, R.W., R.M. Mersereau, and M.A. Richards, 1981: 'Constrained Iterative Restoration Algorithms,' *Proceedings IEEE*, Vol. 69, pp. 432.-450, April.

Scharf, L.L. and J.C. Luby, 1979: 'Statistical Design of Autoregressive Moving-Average Digital Filters,' *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. 27, pp. 240-247, June.

Scheffe, H., 1959:*The Analysis of Variance*, Wiley, New York, 1959.

Schlatter, T.W., G.W. Branstator, and L.G. Thiel, 1976: 'Testing a Global Multivariate Statistical Objective Analysis Scheme with Observed Data,' *Mon. Wea. Rev.*, Vol. 104, pp. 765-783.

Schlatter, T.W., 1975: 'Some Experiments with a Multivariate Statistical Objective Analysis Scheme,', *Mon. Wea. Rev.*, Vol. 103, pp. 246-257.

Schlatter, T.W., 1981: 'An Assessment of Operational TIROS-N Temperature Retrievals over the United States,' *Mon. Wea. Rev.*, Vol. 109, pp. 110-119.

Schweppe, F.C., 1973: *Uncertain Dynamic Systems*, Prentice Hall, Englewood Cliffs, New Jersey.

Schur, I, 1911: 'Bermerkungen zur Theorie der beschrankten Bilienarformen mit unendlich vielen Veranderlichen,' *J. Reine Angew. Math.*, Vol. 140, pp. 1-29, (In German)

Shannon, C.E. and W. Weaver, 1949: *Mathematical Theory of Communication.* University of Illinois Press.

Shannon, C.E., R.G. Gallager, and E.R. Berlekamp, 1967: ' Lower Bounds to Error Probability for Coding on Discrete Memoryless Channels: I ,' *Information and Control*, Vol. 10., pp. 65-103.

Smith, W.L. and H.M. Woolf, 1976: 'The Use of Eigenvectors of Statistical Covariance Matrices for Interpreting Satellite Sounding Radiometer Observations,' *Jour. Atmos. Sci.*, Vol. 33 pp. 1127-1140.

Smith, W.L., H.M. Woolf, C.M. Hayden, D.Q. Wark, and L. McMillin, 1979: 'The TIROS-N Operational Vertical Sounder,' *Bull. Amer. Meteor. Soc.*, Vol. 60, pp. 1177-1187.

Smith, W.L., 1980 : IFAORS Course 401, 'Inversion Methods in Remote-Sensing. Practical Methods for Reducing Satellites Sounding Data for Operational Applications,' March.

Sommerfeld, A., 1949: *Partial Differential Equations in Physics*, Academic Press Inc., New York.

Staelin, D.H., 1969: 'Passive Remote Sensing at Microwave Wavelength', *Proceedings of the IEEE*, Vol. 57, pp. 427-439.

Staelin, D.H., 1981: 'Passive Microwave Techniques for Geophysical Sensing of the Earth from Satellites', *IEEE Trans. on Antennas and Propagation*, Vol. AP-29, No 4, pp. 683-687, July.

Staelin, D.H., 1977: 'Inversion of Passive Microwave Remote Sensing Data from Satellites', in *Inversion Methods in Atmospheric Remote Sensing, A. Deepak (Editor)*, Academic, New York, pp. 361-394.

Staelin, D.H., W.H. Ledsham, R.L. Pettyjohn, P.W. Rosenkranz, R.K.L. Poon, and J.W. Waters, 1975a: 'Microwave Sensing of Atmospheric Temperature and Humidity from Satellites,' Proceedings of the Open Meetings of Working Groups on Physical Sciences of the Eighteenth Plenary Meeting of COSPAR, Varna, Bulgaria, 29 May-June 7, pp. 143-147.

Staelin, D.H., P.W. Rosenkranz, B.G. Anderson, K.F. Kunzi, R.M. Paroskie, R.J. Parr, and J.W. Waters, 1975b: Final Report to NASA, Contract NASA-10693, July.

Stakgold, I., 1979: *Green's Functions and Boundary Value Problems*, John Wiley & Sons, New York.

Stanford, J., 1979: ' Latitudinal-Wavenumber Power Spectra of Stratospheric Temperature fluctuations,' *Jour. Atmos. Sc.*, Vol 36, pp. 921-931.

Stephenson, J.A., 1977: 'The vertical Temperature Structure of the Mid-Latitude Troposphere: A Simple Model,' S.M. Thesis, Department of Meteorology, Massachusetts Institute of Technology.

Strand, O.N. and E.R. Westwater, 1968a: 'Statistical Information Content of Radiation Measurements used in Indirect Sensing,' *J. Atmos. Sci.*, Vol. 25, pp. 750-758.

Strand, O.N. and E.R. Westwater, 1968b: 'Statistical Estimation of the Numerical Solution of a Fredholm Integral Equation of the First Kind,' *J. ACM*, Vol. 15, pp. 100-113.

Strand, G., 1980: *Linear Algebra and its Applications*, Academic Press, New York.

Susskind, J. and D. Reuter, 1985: 'Intercomparison of Physical and Statistical Retrievals from Simulated HIRS2 and AMTS data,' *Interactive Workshop on Advances in Remote Sensing Retrieval Method*, Williamsburg, VA.

Tatarskii, V.I., 1971: *The effects of the turbulent Atmosphere on Wave Propagation*, Washington, D.C., National Science Foundation.

Taylor, G.I, 1935: 'Statistical Theory of Turbulence,' *Proc. Royal Soc.*, Vol. 151, Series A, pp. 421-478.

Tennekes, H., 1978: 'Turbulent Flow in Two and Three Dimensions,' *Bull. Amer. Meteor. Soc.*, Vol. 59, pp. 22-28,.

Tennekes, H. and J.L. Tumley, 1972: *A First Course in Turbulence Theory*, The MIT Press, Cambridge, Mass. 1972.

Thiebaux, H.J., 1975: 'Experiments with Correlation Representations for Objective Analysis,' *Mon. Wea. Rev.*, Vol. 103, pp. 617-627.

Thiebaux, H.J., 1976: 'Anisotropic Correlation Functions for Objective Analysis,' *Mon. Wea. Rev.*, Vol. 104, pp. 994-1002.

Thiebaux, H.J.,1977: 'Extending Estimation Accuracy with Anisotropic Interpolation,' *Mon. Wea. Rev.*, Vol. 105, pp. 691-699.

Thompson, P.D., 1982: 'A Generalized Class of Exact Time-Dependent Solutions of the Vorticity Equation for Nondivergent barotropic Flow,' *Mon. Wea. Rev.*, Vol. 110, pp. 1312-1324.

Tikhonov, A. N. and V.Y. Arsenin, 1977: *Solutions of Ill-posed Problems,* Winston, Washington, DC.

Toldalagi, P.M., 1980: ' Adaptive Filtering Methods Applied to the Remote-sensing of the Atmosphere for Meteorological Purposes,' Ph.D. Thesis, Massachusetts Institute of Technology, E.E.C.S. Department, 1980.

Tom, A.S., 1986: 'Prediction of FIR Pre- and Post-Filter Performance Based upon a Visual Model,' S.M. Thesis,Massachusetts Institute of Technology, E.E.C.S. Department, 1986.

Toutenburg, H., 1982: *Prior Information in Linear Models* , John Wiley & Sons, New York.

Trussell, H.J., 1981: "Maximum Power Signal Restoration,' *IEEE Trans. Acoust., Speech and Signal Processing*, Vol. 29, pp. 1059-1061.

Trussell, H.J. and M.R. Civanlar, 1984: 'The feasible Solution in Signal Restoration,' *IEEE Trans. Acoust., Speech and Signal Processing*, Vol. 32, pp. 201-212.

Tsang, L., J.A. Kong, and R.T. Shin, 1985: *Theory of Microwave Remote Sensing*, John Wiley & Sons, New York.

Tsitsiklis, J.N. and B. C. Levy, 1981: 'Integral Equations and Resolvents of Toeplitz plus Hankel Kernels', Laboratory for Information and Decision Systems and the Department of Electrical Engineering and the Computer Science, M.I.T.

Tveskoi, P.N., 1962: *Physics of the Atmosphere: A Course in Meteorology* Gidrometeorologicheskoe Izdatel'stuc, Leningrad, Israel Program for Scientific Translation, Jerusalem, 1965.

Twomey, S., 1963: 'On the Numerical Solution to the Freedholm Integral Equation of the First Kind by the Inversion of a Linear System Produced by Quadrature,' *J. Assoc. Comput. Mach.*¡ Vol. 10, pp. 97-101.

Twomey, S., 1977: *Introduction to the Mathematics of Inversion in Remote Sensing and Indirect Measurements*, Elsevier Scientific Publishing Company, 1977.

Unser, M., 1984: 'On the Approximation of the Discrete Karhuven-Loeve Transform for Stationary Processes, '*Signal Processing*, Vol. 7, pp. 231-249.

U.S. Standard Atmosphere Supplements, 1966 presented under sponsorship of ENVIRONMENTAL SCIENCE SERVICES ADMINISTRATION, NATIONAL AERONAUTICS AND SPACE ADMINISTRATION, and UNITED STATES AIR FORCE. US Government Printing Office.

VanMarcke, E.H., 1983: *Random fields, Analysis and Synthesis*, Massachusetts. Institute of Technology Press, Cambridge, Ma.

Van Trees, H.L., 1968: *Detection, Estimation and Modulation Theory*, Wiley, New York.

Van Vleck, J.H., 1947: 'The Absorption of Microwaves by Oxygen,' *Phys. Rev.*, Vol. 71, pp. 413-424.

Van Vleck, J.H., and V.F. Weisskopf, 1945: 'On the Shape of Collision Broadened Lines, *Rev. Mod. Phys.*, Vol. 17, pp. 227-236.

Van Zandt, T.E., 1982: 'A Universal Spectrum of Buoyancy Waves in the Atmosphere,' *Geophys. Res. Lett.*, Vol. 9, pp. 575-578.

Vinnichenko, N.K., 1970: 'The Kinetic Energy Spectrum in the free atmosphere-1 second to 5 years,' *Tellus*, Vol. 22, pp. 158-166.

Vinnichenko, N.K, N.Z. Pinus, S.M. Shmeter, and G.N. Shur, 1973: *Turbulence in the Free Atmosphere*, Central Areological Observatory Dolgoprudny, USSR, Consultants Bureau, New York.

Vinod, H.S., 1978: 'A survey of ridge regression and related techniques for improvements over ordinary least squares,' Review of economics and Statistics, Vol. 60, pp. 121-131.

Wagner, A.J., 1965: 'Mean Temperature from 1000 mb to 500 mb as a Predictor of Precipitation Type,' Scientific reports 2, Department of Meteorology, Massachusetts Institute of Technology, May.

Wahba, G. amd J. Wendelberger, 1980: 'Some New Mathematical Methods for Variational Objective Analysis Using Splines and Cross Validation,' *Mon. Wea. Rev.*, Vol. 108, pp. 1122-1143.

Wahba, G., 1981: 'Some New techniques for Variational Objective Analysis on the Sphere using Splines, Hough Functions, and Sample Spectral Data,' *7 th Conference on Probability and Statistics in Atmospheric Sciences*, Monterey, CA, Nov.

Wahba, G., 1984:'Variational Methods for Multidimensional Inverse Problems,' *Interactive Workshop on Advances in Remote Sensing Retrieval Method*, Williamsburg, VA.

Wallace, J.M. and D.S. Gutzler, 1981: 'Teleconnections in the Geopotential Height Field during the Northern Hemisphere Winter,' *Mon. Wea. Rev.*, Vol. 109, pp. 784-812.

Wallace, J.M. and P.V. Hobbs, 1977: *Atmospheric Science: An Introductory Survey*, Academic Press, New York.

Wang, R.J., 1969: 'The Determination of Optimum Gate Length for Time Varying Wiener Filter,' *Geophys.*, Vol. 34, pp. 683-695.

Watanabe, S., 1965: 'Karhunen-loeve Expansion and Factor Analysis,' *Transactions of the Fourth Prague Conference on Information Theory, Stastistical Decision Functions, Random Process*, Academia.

Watson, G.N., 1944: *A Treatise on the Theory of Bessel Functions*, Cambridge University Press.

Westwater, E.R., W. Zhenhui, N.C. Grody, and L.M. McMillin, 1985: 'Remote Sensing of Temperature Profiles from a Combination of Observations from the Satellite Based Microwave Sounding Unit and the Ground Based Profiler,' *J. Atm. Ocea. Tech.*, Vol. 2., pp. 97-109.

Whittle, P., 1962: 'Topographic Correlation, Power-Law Covariance Functions, and Diffusion,' *biometrika*, Vol.49, pp. 305-314.

Wiggins, R.A. and E.A. Robinson, 1965: 'Recursive Solution to the Multichannel Filtering Problem', *Jour. Geophys. Res.*, Vol. 70, pp. 1885-1891.

Wilcox, R.W., 1974: 'An Evaluation of Atmospheric Temperature and Water Vapor Inferences Made from Measurements by a Satellite-Borne Microwave Sensing Device,' S.M. Thesis, Massachusetts Institute of Technology, Department of Meteorology.

Wilheit, T.T. and A.H. Barrett, 1970: 'Microwave Spectrum of Molecular Oxygen,' *Phys. Rev. A.*, Vol. 1, pp. 213-215.

Wilkinson, J.H., 1965: *The Algebraic Eigenvalue Problem*, Clarendon Press, Oxford.

Willett, H.C. and F. Sanders, 1959: *Descriptive Meteorology*, Academic Press, New York.

Woods, J.W. and M.P. Ekstrom, 1976: 'Two-Dimensional Spectral Factorization with Applications in Recursive Digital Filtering,' *IEEE Trans. Acoust., Speech and Signal Proc.*, Vol. 24, pp. 115-128.

Woods, J.W. and C.H. Radewan, 1977 'Kalman Filtering in Two dimensions,' *IEEE Trans. Info. The.*, Vol. 23, pp. 473-482.

Wornell, G.W., 1986: Personal Communication.

Wozencraft, J.M and I.M. Jacobs, 1965:*Principles of Communication Engineering*, John Wileys & Sons, New York.

Wu, C.F.J.,1983: 'On the Convergence Properties of the EM algorithm,' *Ann. Stat.*, Vol. 11, pp. 95-103.

Wyngaard, J.C. and S.F. Clifford, 1977: 'Taylor's Hypothesis and High-Frequency Turbulence Spectra,' *J. Atmos. Sci.*, Vol. 34, pp. 922-929.

Yaakov, B.S., 1971: 'On the Asymptotic Properties of the Maximum Likelihood Estimate Obtained from Dependent Observations,' *Jour. Royal. Stat. Soc.*, Ser. B, Vol. 33, pp. 72-77.

Yang, C.H. and R. Shapiro, 1973: 'The Effects of the Observational System and the Method of Interpolation on the Computation of Spectra,' *J. Atmos. Sci.*, Vol. 30, pp. 530-536.

Yen, T.L, 1956: 'On Non-Uniform Sampling of Bandwidth Limited Signals,' *IRE Trans. Ciruit Theory*, Vol. 3, pp. 251-257.

Youla, D., 1978: 'Generalized Image Restoration by the Method of Alternating Orthogonal Projections,' *IEEE Trans. Cir. Sys.*, Vol. 25, pp. 694-702, Sept.

Zacks, S., 1981: *Parametric Statistical Inference*, Pergamon Press.