USING KRIGING TO BOUND SATELLITE RANGING ERRORS DUE TO THE IONOSPHERE

A DISSERTATION

SUBMITTED TO THE DEPARTMENT OF AERONAUTICS AND ASTRONAUTICS

AND THE COMMITTEE ON GRADUATE STUDIES

OF STANFORD UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Juan Blanch

December 2003

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality as a dissertation for the degree of Doctor of Philosophy.

_____

Per Enge
(Principal Advisor)

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality as a dissertation for the degree of Doctor of Philosophy.

_____

Claire Tomlin

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality as a dissertation for the degree of Doctor of Philosophy.

_____

Sanjay Lall

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality as a dissertation for the degree of Doctor of Philosophy.

_____

Todd Walter

Approved for the University Committee on Graduate Studies:

_____

Abstract

The Global Positioning System (GPS) has the potential to become the primary navigational aid for civilian aircraft. GPS, however, lacks a fundamental feature for a safety critical system: it does not provide hard bounds on the position inaccuracy. Although most of the time the accuracy is excellent, the position error can become very large without any warning to the user. Consequently, satellite based augmentation systems (SBAS) for GPS have been developed to provide corrections and hard bounds on the user errors. Among the sources of error to GPS positioning, the ionosphere is the largest and least predictable. The only ionospheric information available to the United State's SBAS, termed the Wide Area Augmentation System (WAAS), at any given time is a set of slant range delay measurements taken at reference stations distributed across the continent. From this limited and randomly scattered data, the master station must compute in real time an estimate of the ionospheric delay and a hard bound valid for any user within range of the reference stations. The variability of the ionospheric behavior and the stringent integrity requirements have caused the confidence bounds corresponding to the ionosphere to be very large in WAAS, on the order of 10 meters of delay. In the position domain, these conservative ranging bounds translate into conservative bounds on vertical position error. These position error bounds are called protection levels and take values of 30 to 50 meters. Since these values fluctuate near the maximum tolerable, WAAS is not always available. In order to increase the availability of WAAS, we need to decrease the confidence bounds corresponding to ionospheric uncertainty while maintaining integrity.

In this work, I present an ionospheric estimation algorithm based on kriging. I first introduce a simple model of the Vertical Ionospheric Delay that captures both the deterministic behavior and the random behavior of the ionosphere. Under this model, the kriging method, a type of minimum mean square estimator adapted to spatial data, is optimal. More importantly kriging provides an estimation variance at each location that can be easily translated to a confidence bound. However, this method must be modified for two reasons; first, the state of the ionosphere is unknown and can only be estimated through real time measurements; second, because of bandwidth constraints, the user cannot receive all the measurements which are needed to apply kriging. I will show how these two obstacles can be overcome.

The algorithm presented here provides up to 40% reduction in the confidence bound corresponding to the ionospheric delay. This reduction in the ionospheric error results in a 20% reduction in Vertical Protection Levels across the United States. This is achieved without any changes in the hardware and without significantly increasing the complexity of the algorithm.

Acknowledgments

First, I would like to thank my advisor Professor Per Enge, for giving me the opportunity to pursue this research, and for the guidance he has given me in the course of my doctoral program. His insight and contagious enthusiasm has been a source of inspiration through these years. I would also like to thank Dr. Todd Walter, director of the WAAS Laboratory. He has helped me gain not only a deep and wide knowledge but a way of conducting research, and this with endless patience. Working with him has been a pleasure. I hope to continue learning from him in the future.

During the course of my research, I have received the generous advice from several professors at Stanford. Professor Paul Switzer introduced me to the field of spatial statistics, Professor André Journel gave me valuable advice on the use of geostatistics and Professor Ingram Olkin in the use of multivariate analysis. I am very grateful to them, as they have made my work much easier. Also, I would like to thank Professor Claire Tomlin and Professor Sanjay Lall for taking the time to read this thesis.

I would like to thank research associates and students at the Stanford WAAS laboratory for their help, always offered generously: Sriram Rajagopal, Seebany Datta-Barua, Juyong Do, Shau-Shiun Jan, Frank Bauregger, Eric Phelts, Sherman Lo, Denis Akos, Demoz Gebre-Egziabher, Frédéric Bastide, and Matt Deland. My thanks are extended to Doug Archdeacon, Sherann Ellsworth and Dana Parga. I also want to thank Fiona Walter for her careful and thoughtful proofreading of this thesis. Also, I thank the Federal Aviation

Administration for its financial support of the GPS Laboratory and my research in particular.

Finally, I would like to thank my parents for the support and understanding they have given me during these years, and my wife, Vanessa, without whom this thesis would probably not have been possible.

Table of Contents

List of Figures

Glossary of Acronyms

Wide-Area DGPS	
CNMP	Code Noise and Multipath
CONUS	Conterminous United States
GIVE	Grid Ionospheric Vertical Error
HAL	Horizontal Alert Limit
HMI	Hazardously Misleading Information
HPL	Horizontal Protection Level
IFB	Interfrequency bias
IGP	Ionospheric Grid Point
IPP	Ionospheric Pierce Point
MOPS	Minimum Operational Performance Standard. Refers to RTCA159 specification for the Wide-Area Augmentation System (WAAS).
PHMI	Probability of Hazardously Misleading Information
Protection Level	Broadcast indication of the bound on the accuracy of the state. This value is compared to the Alert Limit to determine if a flight operation can continue.
SBAS	Space Based Augmentation System
TEC	Total Electron Content
UDRE	User Differential Range Error

VAL	Vertical Alert Limit
VPL	Vertical Protection Level
WAAS	Wide-Area Augmentation System
WMS	WAAS Master Station
WRS	WAAS Reference Station

Chapter 1

Introduction

The largest obstacles for the Global Positioning System (GPS) to become the primary navigational aid in civil aircraft are the ionosphere [Chao] and radio frequency interference. Ionospheric disturbances can cause positioning errors larger than 50 meters. In this thesis, we present an algorithm that, from a set of measurements of the ionosphere, corrects most of the error induced by the ionosphere and lets the user know the bound on the inaccuracy of the correction. The method used is based on *kriging* [Cressie], [Journel], a type of minimum mean square estimator adapted to spatial data. However, many changes were necessary to adapt it to this particular problem.

In this chapter, we will first introduce some background on GPS, aircraft navigation, and the need for GPS augmentation systems. This will be followed by a description of satellite based augmentation systems and how they enable GPS to become a certified navigational aid. We will then focus on the ionospheric corrections broadcast by the augmentation systems and present previous work relevant to this thesis. Finally, an outline of the thesis will be given.

1.1 GPS AND CIVIL AIRCRAFT NAVIGATION

GPS is a positioning system based on ranging. Each of the 24 GPS satellites in medium earth orbit broadcasts a message that includes the position of the satellite and the time it was sent. The message is broadcast at the frequencies L1 (1575.42 MHz) and L2 (1227.60 MHz) and it is modulated by a pseudorandom (PRN) code, such that all satellites can broadcast at the same frequencies without time sharing [Enge]. The frequency L1 is public whereas L2 is encrypted so that only authorized users can take full advantage of it. A user measures the apparent distance to the satellite by comparing the time each message was sent with the current time. This quantity is called pseudorange. By combining the pseudoranges appropriately, users can find their position. Using stand alone GPS (L1 only), which is available to anyone, the horizontal position error is rarely above 10 meters [Enge01]. For this reason, the potential of GPS for civil aircraft navigation was recognized very early: never before was there a navigational aid so ubiquitous and precise at the same time.

However, for aircraft navigation, it is not enough to be precise most of the time. What is needed is the maximum possible position error: pilots want to know that the error is not going to be larger than a certain bound (called the Vertical Alarm Limit (VAL) in the vertical direction and the Horizontal Alarm Limit (HAL) in the horizontal). Figure 1.1 illustrates this point. If there is no guarantee that the position lies within the red box, the pilot will not be able to go between the two mountains. As a consequence, a useful navigation aid must notify within a certain time (time to alarm) that the positioning error might be larger than the VAL and HAL. This property is called integrity: the system must always be believable.

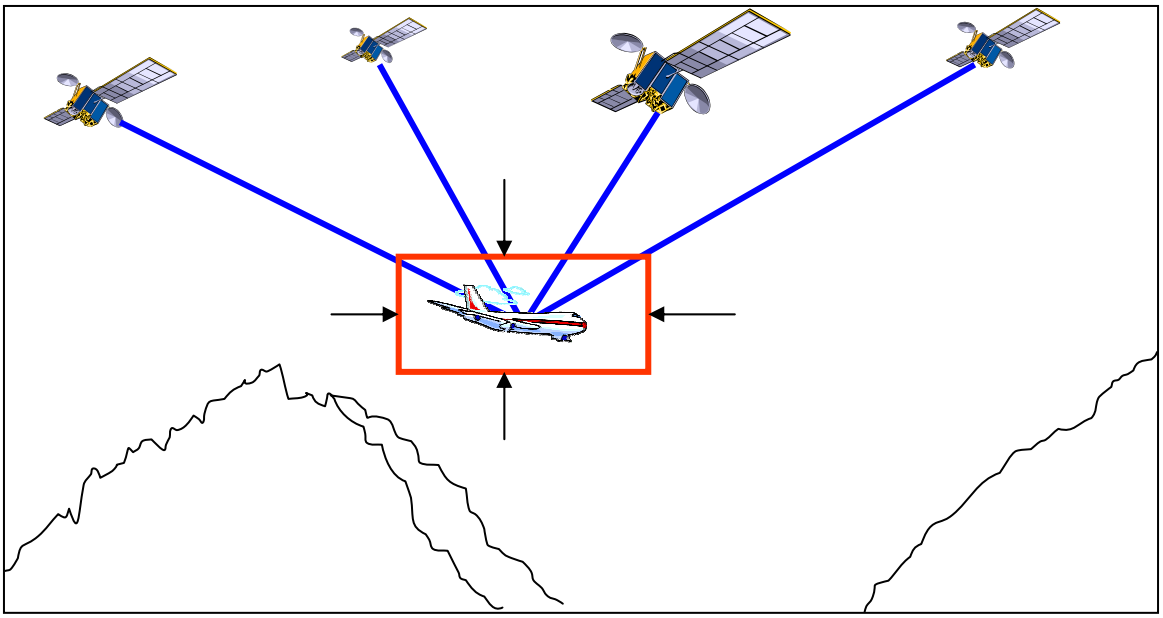

Figure 1.1: Vertical Alarm Limit and Horizontal Alarm Limit

GPS by itself does not have sufficient integrity. There is no mechanism to compute a hard bound on the position error. Although the errors are typically very small, they can become very large without any warning to the user. As we will see later in this chapter, several sources of error can corrupt the pseudorange measurement. These errors in the pseudorange have a direct effect on the position error. GPS by itself cannot therefore be used as a navigational aid for common civil aviation. To fill this gap, the concept of the augmentation system was introduced [Braff]. (Of course, all existing aids to air navigation have monitors, so it is only a continuation of this practice with the newest aid to navigation.) The idea is to monitor GPS signals through reference stations and send corrections and integrity flags to the user so that: first, the position error is reduced; second, the user can determine a hard bound on the position error. There are two concepts for augmentation systems: ground based and satellite based. Ground based augmentation systems are local and only need one reference station [Chou]. (Actually, there are several reference stations that check on each other and are collocated.) Since most of the errors have a strong spatial correlation and the position of the reference receiver is known, it is possible to correct most of the user position error. Satellite based augmentation systems (SBAS) operate in a different manner: the sources of error in the pseudoranges are isolated

and corrected one by one [Kee]. This concept requires several well equipped reference stations and a satellite that can broadcast the correction message. To this date the Wide Area Augmentation System (WAAS) is the only SBAS in operation. WAAS was commissioned by the FAA and started operations over a large part of CONUS on July 10, 2003. There are several other SBAS systems under development: the European GPS Overlay System in Europe [Benedicto], MSAS (MTSAT Satellite-Based Augmentation System) in Japan [Shimamura], GAGAN in India, and several other projects at an earlier stage.

1.2 SATELLITE BASED AUGMENTATION SYSTEMS

As said earlier, the purpose of satellite based augmentation systems is to allow users to correct their position error as much as possible and, more importantly, to generate a hard bound on the error over a large geographic area. This hard bound is called the *protection level*. In the vertical, the hard bound is labeled Vertical Protection Level (VPL) and in the horizontal, Horizontal Protection Level (HPL). More precisely, a protection limit is defined such that at any moment:

$$\text{Prob}\left(error < PL\right) \leq 10^{-7}$$

For an augmentation system to be useful, we need the protection levels to be smaller than the VAL and HAL corresponding to the maneuver that is underway.

In Subsection 1.2.1 we will list the disturbances to which the pseudoranges are subject and distinguish which ones can be corrected by an augmentation system. In Subsection 1.2.2, the overall architecture of WAAS is described (the architecture of all the other satellite based augmentation systems is identical). Finally, we explain how the user calculates his VPL and HPL based on the information contained in the WAAS message.

1.2.1 PSEUDORANGE ERRORS

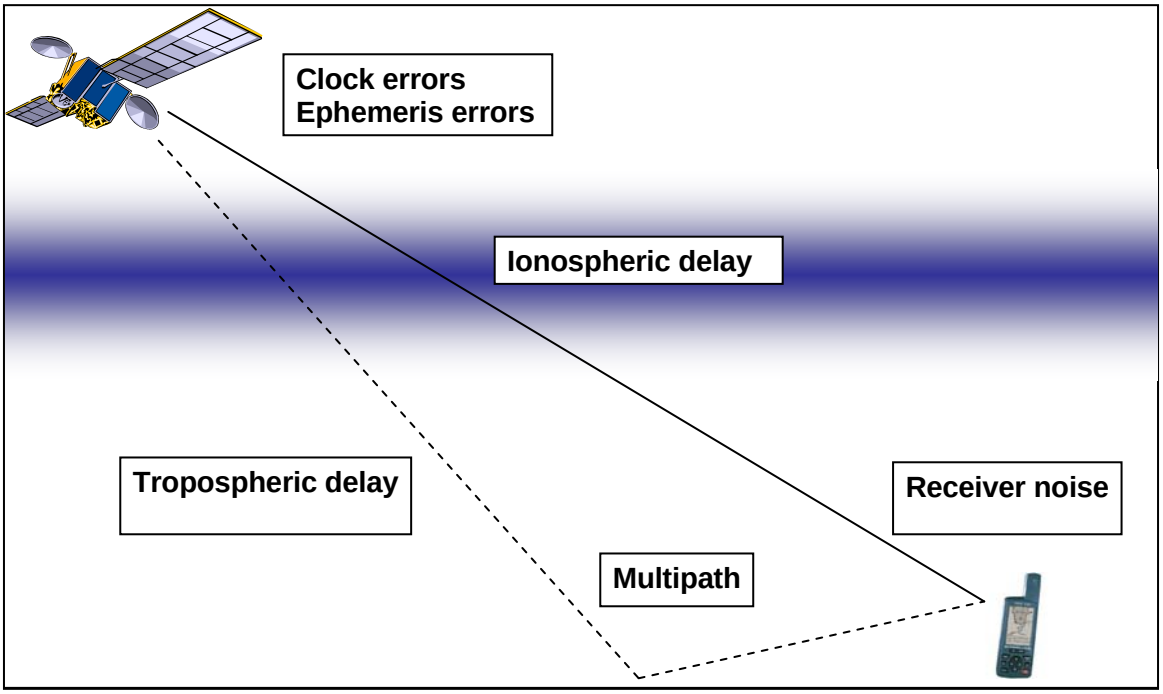

Figure 1.2: Sources of error in GPS

Figure 1.2 illustrates the different sources of errors affecting the pseudoranges for a single frequency user. As we said earlier, the GPS message includes the satellite position and a time tag indicating the time the message was sent. Going from the satellite to the user, the first possible set of errors is due to the ephemeris error (difference between true position of the satellite and broadcast position) and the clock error (difference between the true clock and the satellite clock) [Enge01]. The second error is introduced by the ionosphere. As we will detail later, the ionosphere introduces an additional delay in the message that depends on the free electron content of the upper atmosphere. Because this error is frequency dependent, dual frequency users can correct this error without help. However, single frequency users have no means to measure the additional delay introduced by the ionosphere. The third possible error is introduced by the troposphere, which also introduces an additional delay. This delay is not dispersive, so it cannot be extracted using dual frequency. Fortunately, it can be corrected using relatively simple models [Spilker]. The remaining errors are multipath and receiver noise. Multipath is due to the possible

transmission of the message through paths others than line of sight. These additional paths (which usually have lower power) can distort the correlation peak [Braasch] such that the measured pseudorange is either larger or smaller. The receiver noise is due to the limitations inherent to the receiver. In this thesis, we will explain the origin of the ionospheric delay. Also, because it is relevant to ionospheric estimation, we will see in Chapter 2 how multipath and receiver noise are characterized.

1.2.2 WAAS ARCHITECTURE

The architecture of WAAS is summarized in Figure 1.3. The GPS signals are received by a network of WAAS reference stations (WRS) located in the Conterminous United States (CONUS) and in neighboring regions. The GPS measurements are sent to be processed by the master station. Because the location of the WRS is known and the receivers have dual frequency, it is possible to isolate some of the errors in the pseudorange.

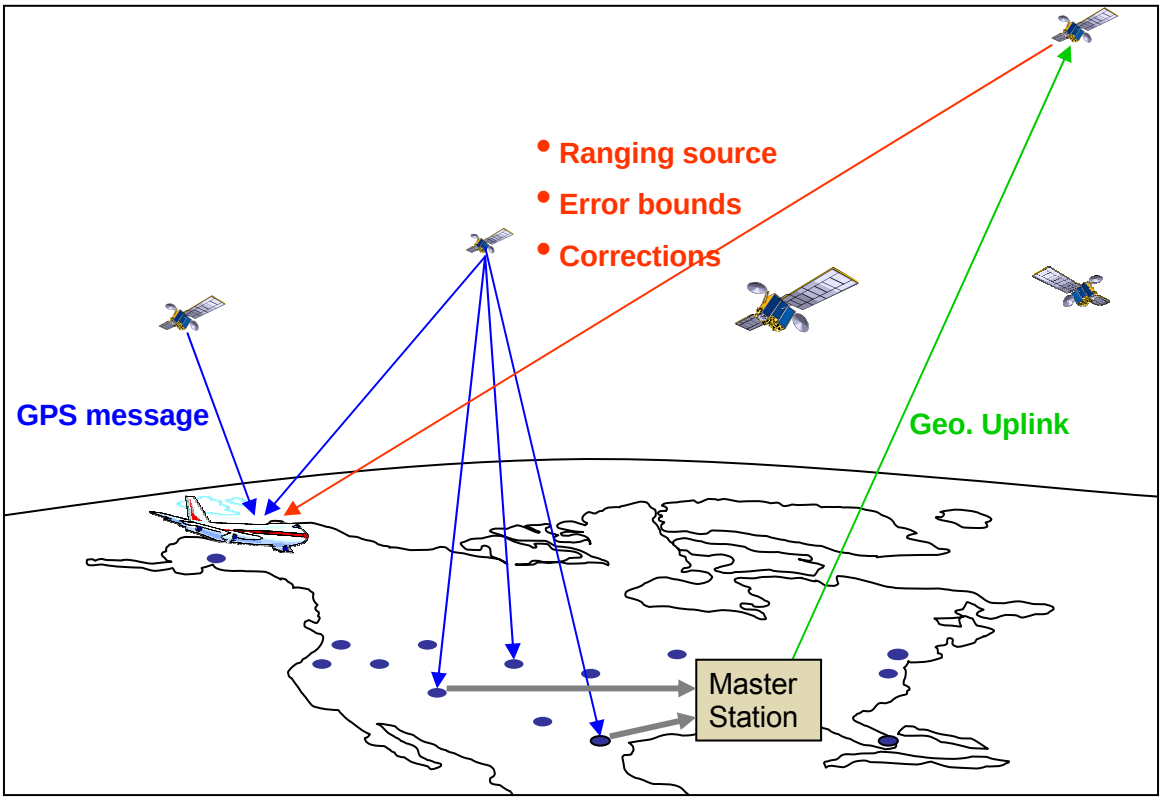

Figure 1.3: WAAS architecture

There are three types of errors that can be corrected – to a certain extent, as we will see -; the ephemeris error and the clock error for each of the satellites, and the ionospheric delay.

The master station produces a message – the WAAS message – that contains integrity information and real time corrections for the ephemeris and clock error for each satellite along with the ionospheric delay. The message is continually uploaded to a geostationary satellite. The geostationary satellite broadcasts the message to the users at 250 bits per second in a PRN code on L1 (1575.42 MHz). In addition to the WAAS message, the geostationary satellite is a valuable source of ranging. For a more detailed introduction to the WAAS architecture we refer the reader to [Enge96].

A WAAS user first measures the pseudoranges to the GPS satellites in view and to the geostationary satellite. Then, using the information contained in the WAAS message it determines which satellites can be used safely, and corrects their pseudoranges (ephemeris and clock error and ionospheric delay). Also, the user computes a VPL and an HPL. These are then compared to the VAL and HAL to determine if the system can be used for the desired procedure.

1.2.3 PROTECTION LIMIT CALCULATION

In this subsection we give a brief description of the Protection Level calculations. Let us first review the equations used in WAAS [Walter97]. The equations to solve are:

$$\rho_c^{(k)} = \left\| x^{(k)} - x \right\| + b + \tilde{\varepsilon}_\rho^{(k)}$$

In this equation, $\rho_c^{(k)}$ is the pseudorange after the WAAS correction for the kth satellite, $\tilde{\varepsilon}_\rho$ are the errors that remain after the correction and b is the user clock bias. After linearization, and assuming that we have initial estimates of the position and the clock bias x_0 and b_0, we have:

$$\rho_c - \rho_0 \doteq \delta\rho = G\begin{bmatrix} x - x_0 \\ b - b_0 \end{bmatrix} + \tilde{\varepsilon}_\rho = G\begin{bmatrix} \delta x \\ \delta b \end{bmatrix} + \tilde{\varepsilon}_\rho$$

In these equations, ρ_c is the vector of pseudoranges after the WAAS correction; ρ_0 is the pseudoranges that would be measured if the user were located in x_0 and its clock bias were b_0; G is a K x 4 matrix characterizing the user-satellite geometry (K is the number of valid satellites in view):

$$G = \begin{bmatrix} \left(-1^{(1)}\right)^T & 1 \\ \left(-1^{(2)}\right)^T & 1 \\ \vdots & \vdots \\ \left(-1^{(K)}\right)^T & 1 \end{bmatrix}$$

The first three columns are the components of the line of sight vector to each of the satellites. $\tilde{\varepsilon}_\rho$ needs to be characterized in order to compute the Protection Limits. Let us suppose now that the errors are Gaussian and independent. The covariance matrix of the errors can be written:

$$W^{-1} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \sigma_K^2 \end{bmatrix}$$

In these conditions, the optimal solution is given by weighted least squares:

$$\begin{bmatrix} \delta\hat{x} \\ \delta\hat{b} \end{bmatrix} = \left(G^T W G\right)^{-1} G^T W \delta\rho$$

(Even if the errors are non Gaussian, the mean squared error is minimized.) The vertical error is then characterized by the standard deviation:

$$\sigma^2_{VerticalError} = \left(G^T W G\right)^{-1}_{3,3}$$

if the third component in x refers to the height. Assuming Gaussian errors, the VPL can be computed as:

$$VPL = K_{V,PA}\sigma_{VerticalError}$$

Here, $K_{V,PA}$ is the quantile of a unit Gaussian distribution corresponding to 10^{-7} and is equal to 5.33. There is a difficulty here: the errors are not Gaussian and not necessarily uncorrelated. The approach taken in SBAS is, for each pseudorange error, to allow the user to compute a Gaussian overbound of the true error distribution assuming that the different errors are uncorrelated. The definition of Gaussian overbound is given in Appendix A, but for now it is enough to see it as a Gaussian distribution that conservatively describes the true distribution. Ignoring the correlation tends to be a conservative step, in the sense that it leads to larger protection limits [Walter97]. When it is not a conservative step, WAAS inflates the individual error bounds to account for the effect of the correlation [Schempp]. The validity of this approach has been confirmed by several years of experience first in the WAAS test bed (the NSTB network) [Dehel] and then in the operational WAAS [McHugh].

As said earlier, the WAAS message allows the user to form the matrix W. For each satellite we have [MOPS]:

$$\sigma_i^2 = \sigma^2_{i,flt} + \sigma^2_{i,UIRE} + \sigma^2_{i,receiver} + \sigma^2_{i,tropo}$$

The two last terms are provided by the user. The third one, labeled *i, receiver*, is the receiver noise and multipath overbound. The fourth term, labeled *i,tropo*, is the delay introduced by the troposphere. The first term, labeled *flt* (fast and long term) includes the User Differential Range Error, which is the error due the ephemeris and clock error. The information needed to compute this term is included in the WAAS message.

The second term, $\sigma^2_{i,UIRE}$, the User Ionospheric Range Error, is also computed using the WAAS message. This term is an overbound of the error remaining in the range delay once the ionospheric correction has been applied. The goal of this thesis is to make this term as small as possible while maintaining the integrity of the system.

1.3 REAL TIME IONOSPHERIC CORRECTIONS

As said earlier, the ionospheric delay corrections and the protection limits are computed in real time from the measurements collected at the reference stations. This is possible because of the strong spatial correlation of the ionospheric delay. (In Chapter 3, this correlation will be stated explicitly.) However, the degree of correlation is quite variable and must be monitored in real time. This variability forces the error bound on the ionospheric delay to be large. The problem is made harder by the fact that the available ionospheric measurements are sparse and subject to measurement noise. All of this has made the ionospheric correction term the dominant term in the VPL.

In this section, previous research on ionospheric estimation is presented. In the first subsection we cite ionospheric modeling and general estimation (not necessarily for navigation use.) Since there is a large amount of previous work in this field, we will only cite the works most relevant to this thesis. The second subsection presents work specifically addressing ionospheric estimation with an emphasis on integrity.

1.3.1 PREVIOUS WORK

The first class of previous work deals with ionospheric mapping. The goal of these studies is to find consistent representations of the ionosphere given the available measurements in order to draw conclusions about what is happening in the ionosphere.

To our knowledge, the first attempts to characterize the random properties of the ionospheric delay as measured by GPS were made by Delikaraoglou [Delika] and Wild [Wild]. In those early works, there are several interesting points relevant to this thesis.

First, temporal and spatial autocorrelation plots are computed from GPS ionospheric measurements. Second, the difference between the stochastic behavior of the ionosphere and the measurement noise is clearly drawn. Third, it is recognized that, within the thin shell model (which will be explained in Chapter 3) the ionospheric delay can be interpreted as having a deterministic component and a stochastic component. Finally, an estimation method close to kriging, called co-location, is presented (the difference with kriging, is that the determination of the trend is done visually, which prevents it from being made automatic). The application of kriging to ionospheric mapping was introduced by P.A. Bradley in [Bradley]. These techniques have been applied to the mapping of different ionospheric characteristics, in particular the Total Electron Content (TEC) by Stanislawska *et al* [Stanis]. For our purpose, what is lacking in these studies is the computation of a reliable error bound on the estimate.

Another class of ionospheric estimation algorithms is based on tomography. Because the ionosphere is three-dimensional and the GPS measurements are integrals along the ray path, it is natural to try to reconstruct the Total Electron Content using tomography. The most relevant effort to this research has been realized by A. Hansen [Hansen02]. In this same reference, there is an overview of all ionospheric estimation algorithms based on tomography. Compared to previous applications of tomography to ionosphere estimation, there is an attempt to produce a tight error bound on the estimate. In addition, this ionospheric estimation algorithm was adapted to WAAS.

Previous work on ionospheric estimation specific to satellite based augmentation systems includes all the early work done for WAAS [Chao]. At that point, the ionosphere was believed to be sufficiently well behaved so that a single model for ionospheric behavior would be enough. Error bounds were computed that depended on the geometry of the measurements, but they assumed smooth ionospheric models corresponding to low ionospheric activity.

1.3.2 CURRENT WAAS IONOSPHERIC ESTIMATION ALGORITHM

The previous work that is the most relevant to this thesis is all the effort that has led to the current ionospheric estimation algorithm for WAAS. In 1999, at the beginning of the solar maximum season (one of the periods in the 11 year solar cycle [Hargreaves]) it was found that assuming smooth ionospheric behavior during ionospheric storms could lead to integrity faults in WAAS.

The difference between quiet and stormy ionospheric behavior was exposed in [Hansen00] through decorrelation plots. The large difference between quiet and stormy ionospheric behavior resulted in the development of the storm detector [Walter00], which is the basis of the current WAAS ionospheric algorithm. It is interesting to summarize this paper here, as several ideas from it are used in this thesis. The idea of the algorithm is to test whether the real time measurements (at a given time) are compatible with a quiet ionosphere via a chi-square test on the measurements. If the chi-square test statistic is larger than a certain threshold, then the ionosphere is assumed to be stormy, no correction is attempted, and a very large error bound is set, which will virtually cover any possible error, but will make the system unavailable for precision approach. If the chi-square statistic is below the threshold, then the ionosphere is probably in a quiet state. However, the error bound is inflated because even if the statistic is low, there is still a possibility that the ionosphere is mildly disturbed. Improvements to this algorithm have been suggested, including but not limited to the use of a dynamic inflation factor [Angus], [Cormier].

The previous algorithm was designed assuming that the ionosphere was well sampled by the WAAS reference station measurements. Unfortunately, while this is true in most of the CONUS region, it does not hold at the edge of coverage. There, the measurements become sparse as there cannot be reference stations in the ocean. The concern is that the ionosphere might appear quiet in the WAAS measurements when, in reality, there are disturbances very close to the measurements that are not detected. This concern led to the development of the *threat model* [Sparks01]. The threat model is a methodology to put a hard bound on the possible threats present in archived real ionospheric data. A threat, here, is defined as a class of ionospheric behavior that could lead to an integrity failure. As it is

an essential part of the algorithm that is reused in this thesis, it will be presented with some details in Chapter 4. Further improvements of the ionospheric threat model have been proposed in [Sparks03].

1.4 CONTRIBUTIONS

Since many of the contributions need a long introduction which will be given in the thesis, we will only give a brief summary of them here. We found a simple local characterization of the vertical ionospheric delay as a deterministic component and random field correlated with distance; this is a refinement of previous work. To our knowledge, we have introduced the first application of kriging in a safety critical system and in real time (both are new). To do that, we evaluated the effect of measurement noise on the determination of the underlying random model. This is one of the most valuable contributions of this thesis as it can be adapted to other critical systems where there is both process noise and measurement noise but where the characteristics of the process noise are not well known. The last important contribution resides in the fact that the ionospheric estimation algorithm designed in this thesis could provide a 20% reduction in VPL compared to the current algorithm, with the same safety standard and without changing the current SBAS message standards.

1.4.1 OUTLINE

The second chapter gives a more detailed picture of the ionosphere and of the available sensors for estimation, an explanation of the algorithm requirements and the problem statement. The third chapter explains how to obtain a simple and useful model of the vertical ionospheric delay and how to reduce the degree of disturbance to a single parameter; this chapter also shows some of the limits of the model. The fourth chapter uses the model derived in the third chapter to derive an estimation algorithm. There are two different sections each one dealing with a different side of the problem: irregularly scattered measurements and uncertain underlying random model. At the end of Chapter 4

we include a description of the threat model. It is not a contribution in this thesis, but it needs to be explained since it is a required part of the algorithm. Chapter 5 shows how to modify the algorithm designed in Chapter 4 to the current SBAS standard. Finally, in Chapter 6 we evaluate the algorithm using a Service Volume Analysis tool.

Chapter 2

Problem Statement

The purpose of this chapter is to define the problems. What is the effect of the ionosphere on the GPS signal? How are the GPS measurements processed to extract ionosphere information? What are the requirements on the ionosphere estimation algorithm? First, we will give a brief description of the ionosphere and its effects on radio wave propagation. We will then give the formula that links the electron content with the range delay experienced by a user. From there, we will see how the dispersive nature of the ionosphere allows a dual frequency receiver to compute the range delay introduced by the ionosphere. We will review how the GPS measurements are processed, and how their noise is characterized. Then we will turn our attention to the specific requirements of the ionospheric estimation algorithm for satellite navigation. As this chapter is an overview of previous work, it will be brief.

2.1 IONOSPHERIC PHYSICS

We start by giving a brief description of the ionosphere and its effect on GPS measurements, and refer the reader to several references that cover the subject extensively. For the description of the ionosphere, the focus is on the Total Electron Content (which will be defined.)

2.1.1 IONOSPHERE STRUCTURE

The ionosphere is a region of the upper atmosphere constituted of partially ionized plasma. This region extends from around 100 km height to above 1000 km. The ionization results from the ultraviolet radiation from the Sun. The exact electron density at any location is caused by a complex interplay between the solar flux and the terrestrial magnetic field. Because the UV radiation is not constant and the earth is rotating, the electron density is variable both spatially and temporally. There is a diurnal pattern, where the maximum density is located near the equator and lags the sun by two to three hours. From that point, the electron density tapers off toward the poles. Despite the gross regularity, the details are variable and difficult to predict. We will not discuss the several models that have been developed to integrate all physical phenomena affecting the ionosphere and predict the electron density. What is important for us here is that none of them has been very successful in predicting this day-to-day variability [Klobuchar].

Because the electron density is a function of the solar flux, it is linked to the 11 year solar cycle. During this cycle the UV flux goes through a maximum. During this period, the ionosphere is more likely to be disturbed, in the sense that the electron density can become much higher and much more unpredictable than on a quiet day. Also, the daily pattern is disturbed during these stormy events. Although more likely during the high solar season, these events can happen during “non-peak” portions of the solar cycle.

2.1.2 IONOSPHERE PROPAGATION

The major effects of the ionosphere on the GPS signal are: Doppler shift, Faraday rotation of linearly polarized signals, bending of the radio wave, scintillation and delay. Here we will only discuss the delay introduced by the ionosphere on the GPS signal, as it is by far the most severe effect in mid-latitudes (in Equatorial and high latitudes, scintillation is also a problem). The ionosphere is a dispersive medium: the refractive index is dependent on the frequency. More precisely, we have the approximation (obtained from the Appleton and Hartree general equation for the refractive index of the ionosphere [Hargreaves]):

$$n(r,f) = 1 - \frac{N_e(r)e^2}{2\varepsilon_0 m f^2} = 1 - \frac{40.3}{2f^2} N_e(r) \tag{2.1}$$

In this equation, $n(r,f)$ is the refractive index at location r, f is the frequency, and $N_e(r)$ is the electron density.

This equation allows us to evaluate the effect of the ionosphere on the GPS signal. The group delay (the delay suffered by the message) of a ray path crossing the ionosphere is given by:

$$\Delta\rho = \int (1-n)\,dl$$

By plugging Equation (2.1) into this last equation we get:

$$\Delta\rho = \frac{40.3}{f^2} \int N_e dl \text{ (in number of electrons per m}^2\text{)} \tag{2.2}$$

This last equation shows that the group delay is proportional to the inverse of the frequency squared and to the quantity $\int N_e dl$ which is called the Total Electron Content (TEC) and it is expressed in number of electrons per square meter.

In addition to the group delay, the refractive index provokes an advance of the carrier phase [Klobuchar]. The carrier phase $\Delta\Phi$ advance is such that:

$$\lambda\Delta\Phi = -\Delta\rho \tag{2.3}$$

Equations (2.2) and (2.3) express the two effects that we are going to try to mitigate in this thesis. A user tracking the GPS signal at only one frequency cannot observe and correct the delay introduced by the ionosphere. This is the reason an ionospheric estimation algorithm is needed. (It is possible to detect relative changes in the ionosphere by comparing the carrier phase and the code delay [Klobuchar]. This possibility is not exploited in satellite based augmentation systems because it would put more responsibility

on the user side than is desirable, and would not be robust enough, because the bias is poorly observed [Enge96]).)

2.2 GPS IONOSPHERIC DELAY MEASUREMENTS

For a single frequency user, the delay (2.2) is not observable. However, using two different frequencies, the delay (2.2) can be isolated. GPS transmits its message in two frequencies: L_1 (1575.42 MHz) and L_2 (1227.6 MHz). Excluding all the other errors, two pseudorange measurements from the same receiver to the same satellite differ by:

$$\rho_{L_1} - \rho_{L_2} = 40.3 f_{L_1}^2 \left(1-\gamma\right) TEC$$

where $\gamma = \frac{f_{L_1}^2}{f_{L_2}^2}$. The delay in meters at L_1 is then (we can keep the notation TEC for the delay since it is proportional to the TEC in electrons per square meter):

$$TEC = \frac{\rho_{L_1} - \rho_{L_2}}{\left(\gamma - 1\right)}$$

Similarly, we have for the carrier phase measurements:

$$TEC = -\frac{\Delta\Phi_{L_1} - \Delta\Phi_{L_2}}{\left(\gamma - 1\right)}$$

While the code measurement is absolute and noisy, the carrier phase measurement has a much lower noise, but is ambiguous (in the sense that the measurement is accurate modulo the length of a wave. In the next section, this ambiguity is expressed through the unknown N_i) These two estimates can be combined to form an improved TEC estimate. We now go into more detail.

2.2.1 ERRORS AND BIASES

The situation is not as simple as the previous equations suggest. There are many more terms in the pseudorange measurement that impair our observability of the ionospheric delay. The four available measurements are:

$$pr_{L_1} = R + TEC + M_{L_1} + \xi_{L_1}$$

$$pr_{L_2} = R + \gamma\left(TEC + IFB + \tau_{gd}\right) + M_{L_2} + \xi_{L_2}$$

$$\phi_{L_1} = R + M_{\phi} + \xi_{\phi} - TEC + N_1\lambda_1$$

$$\phi_{L_2} = R + M_{\phi} + \xi_{\phi} - \gamma\left(TEC - IFB - \tau_{gd}\right) + N_2\lambda_2$$

In these equations, R designates the range and the clock biases (satellite and receiver). This term is common to all four measurements. M is the multipath (error due to the fact that not all the received signals are from the line of sight [Braasch]). The multipath term is negligible in the carrier phase measurements. ξ is the receiver noise; it is different for each of the terms and also negligible in the carrier phase measurements. IFB is the interfrequency bias in the receiver, and τ_{gd} is the interfrequency bias in the satellite. N_i is the ambiguity in the carrier phase measurement and is unknown. Neglecting multipath and receiver noise in the carrier phase measurement we get the two observables:

$$T\hat{E}C_{code} = TEC + \frac{\gamma}{\gamma - 1}\left(IFB + \tau_{gd}\right) + \frac{1}{\gamma - 1}\left(M_{L_2} - M_{L_1} + \xi_{L_1} - \xi_{L_2}\right)$$

$$T\hat{E}C_{phase} = TEC + \frac{\gamma}{\gamma - 1}\left(IFB + \tau_{gd}\right) - \frac{1}{\gamma - 1}\left(N_2\lambda_2 - N_1\lambda_1\right)$$

These two measurements are combined to provide a low noise estimate of the ionospheric delay. Second, we want to characterize our measurement uncertainty in a conservative way. To do that we need to calibrate the biases – which have a time constant of days - and get a bound on the receiver noise and multipath.

2.2.2 HARDWARE BIAS CALIBRATION

The hardware bias calibration (IFB and group delay) is done using software based on the Global Ionospheric Model [Wilson]. The idea is to assume that the ionosphere is stationary in a frame moving with the sun along the magnetic equator. The ionosphere is then described by a set of coefficients representing the magnitude of the electron density function. The problem is then to estimate the hardware bias and the coefficients at the same time from the observed measurements. There are two versions of the hardware calibration: a batch solution (used for post-processing) and a sequential solution. Because the biases evolve very slowly (on the order of days) the bias estimation algorithm can use long time constants (for the sequential filter) or several days worth of ionospheric data (for the batch solution). The equations describing the relationship between the coefficients, biases and measurements are linear, so the problem can be solved efficiently using least squares in the batch version of the estimator and a Kalman filter in the case of the real time estimator.

It has been shown that the bias determination is not very sensitive to the ionospheric model that is assumed [Hansen02], and that departures from the ionospheric model do not harm the accuracy of the bias estimates. The biases can be calibrated to better than 10 cm and errors in the biases do not exceed 1 meter.

2.2.3 RECEIVER NOISE AND MULTIPATH

There are several techniques to mitigate and characterize the receiver noise and multipath, also labeled code noise and multipath (CNMP). The technique that was selected for WAAS estimates and corrects for the multipath and then provides a noise estimate for the residual error in the corrected measurements. This approach is detailed in [Graas]. A high level description is provided in [Shallberg]. In this section we summarize the account given in [Shallberg].

The CNMP algorithm has three major steps: a mean filter, a mean error function, and cycle slip detection. Multipath error is calculated using the standard CNMP relationship for dual

frequency code pseudorange and the carrier phase (which can be found in [Shallberg].) Due to the ambiguities inherent to the carrier phase measurement, the estimate of the multipath error is biased. The bias is removed using the mean filter estimator. The final multipath correction is obtained by differencing the mean filter estimate from the current multipath estimate. Finally, the multipath corrected pseudoranges are carrier smoothed with a period equal to the mean filter time constant. The mean error function is the residual error once the estimate of the multipath has been removed. Figure 2.1 shows the overbound of the mean error function as a function of time.

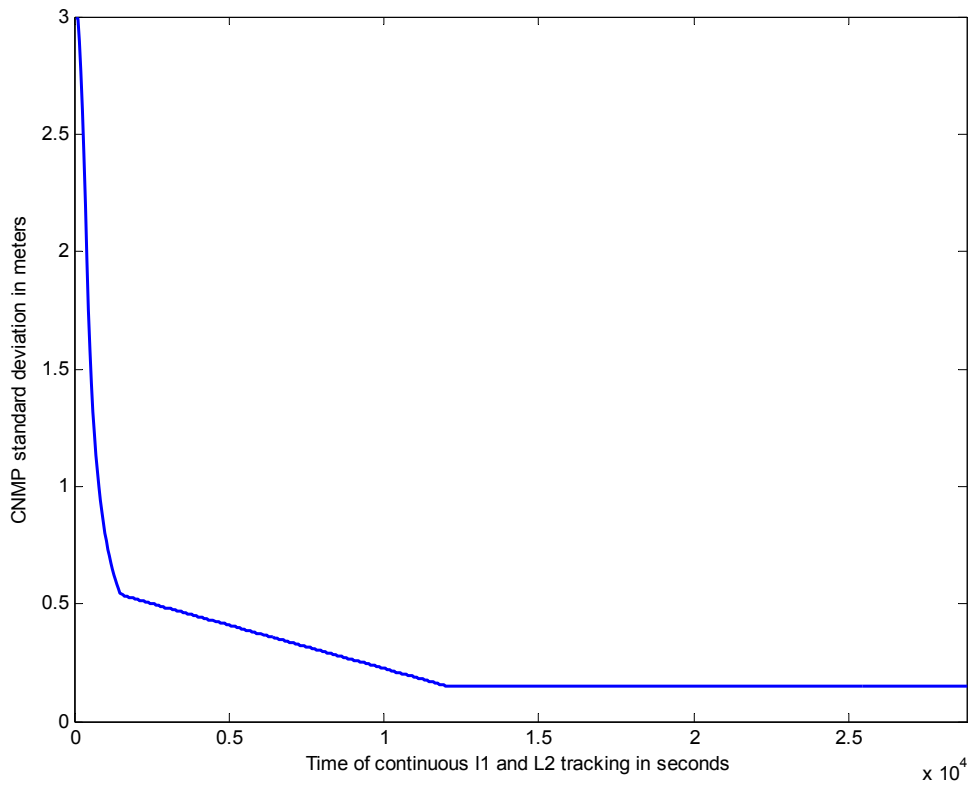

Figure 2.1: CNMP standard deviation

Because of the mean filter, the multipath correction estimate is not accurate at the beginning of the satellite track. However, as time passes, the estimate becomes increasingly accurate due to the more accurate determination of the carrier ambiguities. The mean error function was determined empirically making use of WAAS measurements.

It represents the standard deviation of a Gaussian overbound (see Appendix A) of the residual error. The mean error function starts out large and decreases towards a floor value that is reached after 12000 s. The last piece of the CNMP algorithm is the cycle slip detector. It uses both a dual frequency carrier phase detector and a single frequency carrier phase detector. The single frequency detector is there to rule out any simultaneous cycle slips in L1 and L2 that would be unobservable for the dual frequency detector. Cycle slips force data gaps to restart the algorithm.

2.2.4 POST-PROCESSED IONOSPHERIC DATA

In the previous sections we have dealt with the real time measurements that will be fed into the ionospheric estimation algorithm. We now turn our attention to post-processed data. Post-processing raw data allows us to remove noise that we would not be able to remove in real time, and thus obtain very accurate ionospheric delay measurements. A description of the processing can be found in [Hansen00]. As we will see in Chapter 3, it is critical to have low measurement noise and an extensive data set of ionospheric delay measurements to be able to correctly model the ionosphere. This is all the more important for WAAS, as it has been decided early on that the design of the algorithm and the safety analysis would rely on measurements. Here, we outline each step to produce a 'supertruth' data set from raw measurements collected at the WAAS reference stations. For more details, please refer to [Hansen00].

2.2.5 MEASUREMENT NOISE COVARIANCE

None of the calibration and data processing described above is unique to this thesis. What is important here, is that after this process (be it real time or post-processed) we end up with a set of measurement data at each time frame. The accuracy of the measurements is characterized by a covariance matrix M, which characterizes both the code noise and multipath error (after correction) and the uncertainty introduced by the bias (after correction also). This matrix is almost diagonal. There are off-diagonal terms because the error in the bias estimation is common to all measurements coming from the same satellite or the same reference station. For every measurement, the standard deviation (the diagonal

element corresponding to the measurement) is a Gaussian overbound of the true error distribution. The covariance M naturally varies with time and the magnitude of the standard deviation of the errors is on the order of 50 cm.

2.3 WAAS REFERENCE STATIONS

The measurements described earlier are collected through a network of reference stations. The location of these reference stations can be seen in Figure 1.3. Each reference station has several dual frequency capable receivers, a precise clock and an antenna located in a low multipath environment. Several dual frequency receivers provide diversity in the measurements. This diversity reduces the impact of noise and aids the detection and isolation of faults in every single measurement.

2.4 REQUIREMENTS OF THE ALGORITHM

As discussed in Chapter 1, the purpose of WAAS (and all SBAS systems) is to allow the user to compute a hard bound on the position error (VPL or HPL) or to let the user know that it is not safe to use a certain measurement (because the uncertainty on the error is too large). We will first present the exact requirement on the ionospheric estimation algorithm and then describe why it is difficult to achieve.

2.4.1 PROBABILITY OF HAZARDOUSLY MISLEADING INFORMATION

As we said in the Introduction, the user must receive a correction and Gaussian overbound for each of the errors, or a flag indicating that the pseudorange cannot be used safely [MOPS]. For the ionospheric delay, when a valid correction is sent (along with an error bound), the Gaussian overbound condition can be written:

$$P\left(\left|I_{est} - I_{true}\right| \geq K\sigma_{est}\right) \leq Q\left(K\right) \text{ for all } K \geq 0 \tag{2.4}$$

In this equation, I_{est} is the ionospheric delay computed by the user, I_{true} is the real ionospheric delay, and Q is the residual of the normal cumulative distribution function. σ_{est} is the standard deviation of the Gaussian overbound. We now describe how Condition (2.4) can be simplified. First of all, the Gaussian overbound theorem (See Appendix A and [DeCleene]) can be relaxed in our case, because we only need a VPL such that:

$$P\left(\text{vertical error} < K_{V,PA} d_{3,3}\right) < 10^{-7}$$

As a result, we only need to meet Condition (2.4) for K in a finite range. In this thesis we will assume that it is enough to satisfy (2.4) when $K<5.592$. We can further simplify the requirement by noticing that it is easier to fulfill condition (2.4) using small K than large K. In other words, the solution of the equation:

$$P\left(\left|I_{est} - I_{true}\right| \geq K\sigma_{est}\right) = Q\left(K\right)$$

where σ_{est} is the unknown, increases as a function of K. (We are not proving this now, but it will be revisited in Chapter 4.) Consequently, it is enough to have:

$$P\left(\left|I_{est} - I_{true}\right| \geq K\sigma_{est}\right) \leq Q\left(K\right) \text{ for } K = 5.592 \tag{2.5}$$

A failure to fulfill Condition (2.5) constitutes what is called Hazardously Misleading Information (HMI). The left side of Inequality (2.5) is called Probability of Hazardously Misleading Information (PHMI).

The problem now is defined: a WAAS user needs to be able to compute an estimate of each ionospheric delay corresponding to each satellite in view and an error bound such that (2.5) is fulfilled at any moment. It would be easy to send a very large error bound, we would then be sure that (2.5) is met. However, for WAAS to be useful, the VPL must be below a certain threshold. It is for this reason that the purpose of this work is to make the error bound as small as possible subject to Equation (2.5).

2.5 CONSTRAINTS

In this section we briefly discuss the difficulties of solving the problem stated in the previous section. We can divide these difficulties into two categories: constraints coming from the coupling of the unpredictable physical behavior of the ionosphere with the scattered and noisy measurements, and the operational constraints.

2.5.1 IONOSPHERIC CONSTRAINTS

The main difficulty in estimating each ionospheric delay and the error bound is the unpredictable nature of the ionosphere coupled with the limited information that we have about it. Because the ionospheric delays are correlated with distance – in a sense that will be explained in Chapter 3 - it is possible to estimate the ionospheric delay experienced by a user within the WAAS reference stations network. However, the characteristics of the ionosphere change significantly over time. In particular, the stochastic properties of the ionospheric delay – which are essential to form the error bound - are highly variable. We therefore need to either check that an assumed model holds or estimate in real time the magnitude of the decorrelation. This is all the more difficult due to the limited amount of measurements, and their noisy nature, which makes the true ionosphere decorrelation less observable. To make things even more difficult, there are sudden features in the ionosphere that might be unobservable to the network of reference stations but that could affect potential users. The ionospheric estimation algorithm needs to take all these threats into account such that Equation (2.5) holds.

2.5.2 SYSTEM CONSTRAINTS

The constraints introduced in the previous subsection are unavoidable, because we do not have any control over the ionospheric behavior. In this subsection, we introduce the constraints inherent to the system. It is possible to design an algorithm disregarding the limitations imposed by the current system, but such an algorithm could not be used in the short term, and could only be used if its performance was worth the change. In this work,

one of the main purposes of the algorithm is to fit within the frame that has been set up and put to use in the current system.

The most important constraint is the messaging structure for the ionospheric information, which is determined in the WAAS Minimum Operational Standard (WAAS MOPS). The details of the standard can be found in [MOPS]. What is important here is that the ionospheric information is sent as a grid of values on a thin shell, which will be described in Chapter 3 -. Each value is nominally updated every 5 minutes. However, the user can utilize a correction for up to 10 minutes, unless a warning is sent saying that the correction is no longer valid. From this set of grid points, the user can compute an estimate of the ionospheric delay for each of the satellites in view. In Chapter 5, we will see how the ionospheric estimation algorithm was adapted to the WAAS MOPS.

Chapter 3

Ionospheric Delay Structure

In the previous chapter, we saw how GPS measurements are processed in real time and post-processed to obtain clean GPS ionospheric measurements. This chapter makes use of the large amount of post-processed data to determine a simple model that encompasses both the deterministic and stochastic properties of the ionospheric delay. The methodology applied here is primarily based on proven methods used in geostatistics. However, there are several differences, the most important of which is the amount of data available. While the geostatistician can look at all the data available (typically less than a hundred points), we have to rely on mechanical analysis, due to the size of the data set. Also, the concern is different: where the geostatistician looks for accuracy, we look for integrity. That is, we want a conservative model of the ionosphere, in the sense that it leads to strong error bounds in the estimation process.

First, we need to explain a fundamental approximation that has simplified ionospheric estimation: the thin shell model. This model transforms a three dimensional problem into a two dimensional one, by transforming the slant ionospheric delay as measured by a GPS receiver into a vertical ionospheric delay. The shortcomings associated with the thin shell model will be briefly addressed as well. Second, we will show a series of maps of the vertical ionospheric delay over CONUS. As with any exploratory analysis of spatial data, we start with a close look at pictures of the phenomenon we are trying to characterize. This

analysis will allow us to determine qualitatively the driving features that can be characterized deterministically. After that, we will turn our attention towards the statistical properties of the vertical ionospheric delay. We will see that there is a simple and conservative model of the ionospheric delay that captures most of the effects. This study will continue by indicating the limits of the model, which frequently impacts any statistical characterization. Finally, we will examine the temporal decorrelation of the vertical ionospheric delay, and argue that, given our time scale, it is a second order effect.

3.1 THIN SHELL APPROXIMATION

The thin shell approximation states that all the free electrons lie within a thin shell at a given altitude. Figure 3.1 illustrates this approximation: the slab represents the thickness of the ionosphere and the blue traces of the ray paths show where they are affected by the free electrons. The path integral corresponding to a GPS ionospheric measurement can be approximated by [Klobuchar]:

$$\text{Iono. delay} = \frac{\alpha \int_{\text{Satellite}}^{\text{User}} \mathrm{n_e}(l)\,dl}{f^2} = \text{Vertical Iono delay*obliquity factor} \qquad (3.1)$$

Within this model, the path integral becomes a punctual measurement on the thin shell labeled 'vertical ionospheric delay,' scaled by a multiplying factor labeled 'obliquity factor.' The location of this point, which corresponds to the location where the ray path crosses the thin shell, is known as Ionospheric Pierce Point (IPP). The obliquity factor depends on the assumed altitude, H, and the elevation angle, α. It can be expressed as:

$$\text{Obliquity factor} = \sec\left(\arcsin\left(\frac{R_e}{R_e + H}\cos(\alpha)\right)\right) \qquad (3.2)$$

R_e is the radius of Earth. The altitude of the thin shell is set arbitrarily and is intended to represent the mean height of the ionosphere. In the current WAAS broadcast message, the altitude is fixed at 350 km [MOPS].

The drawbacks of this approximation are obvious: we ignore any vertical distribution of the electron density. For example, two slant delays with the same IPP but different elevation angles will have a different magnitude, because the region of the ionosphere crossed by each ray path is different (except at the IPP). Yet, we retain the thin shell approximation. First, it simplifies the estimation problem. Second, as we will see in Chapter 5, the current ionospheric correction message relies on the thin shell model. A method relying on another model would require changes in the broadcast standard - a standard already adopted by the SBAS-capable receiver manufacturers [MOPS], [SARPS]. Such a change would only be possible if an outstanding benefit could be expected. Finally, methods that relax the thin shell approximation to take into account the vertical density profile of the ionosphere have so far shown a low gain in performance compared to the thin shell model approximation under the circumstances investigated [Lejeune].

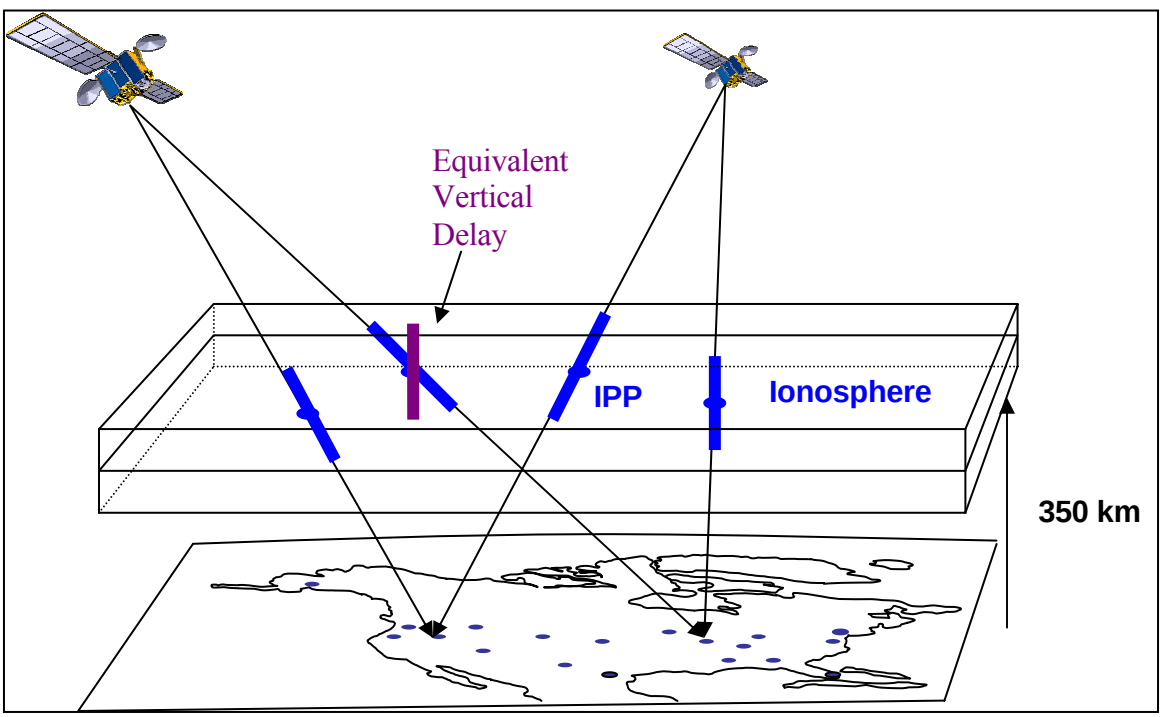

Figure 3.1: Thin shell approximation over CONUS

3.2 IONOSPHERE SNAPSHOTS

One of the useful properties of the thin shell model is to reduce each measurement to a data point (the IPP) on a two-dimensional map. In this section, we show some snapshots of the vertical ionospheric delay for different ionospheric conditions - quiet or stormy - and at different times of the day. The distinction between quiet and stormy conditions is loose here and there is a range of behaviors between them. For now it is sufficient to say that quiet conditions generally correspond to low vertical ionospheric delays (below 10 meters) and high correlation between neighboring measurements. For stormy conditions, it is the opposite.

3.2.1 QUIET CONDITIONS

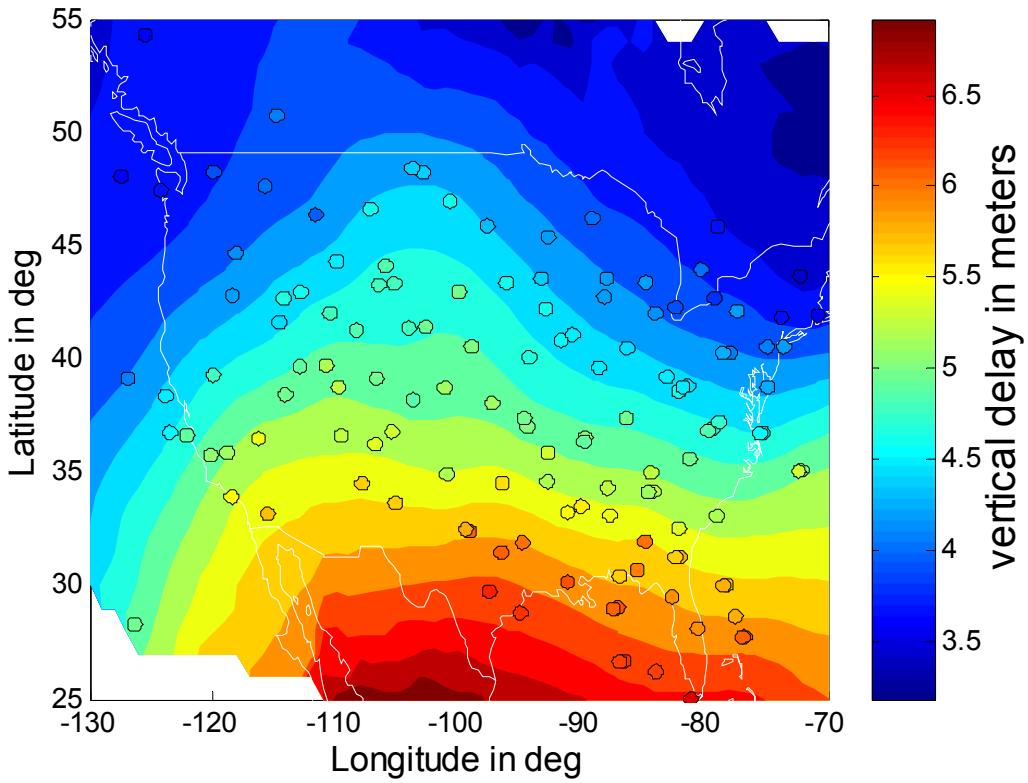

Figure 3.2: Vertical ionospheric delay on July 2, 2000 at 5:00 p.m. E.T

Quiet conditions occur most of the time (more than 95%, even in the high solar season). In Figures 3.2 and 3.3 we show snapshots of the ionosphere at two times on July 2, 2000. The circles represent the post-processed IPP measurements at a given time, that is, all the slant measurements from each of the reference stations to each satellite in view transformed to an equivalent vertical ionospheric delay. The color code indicates the magnitude of the vertical ionospheric delay measured at each location. The colored map is created by interpolating the measured values. At this point, the only purpose of the underlying map is to help us visualize the ionospheric delay.

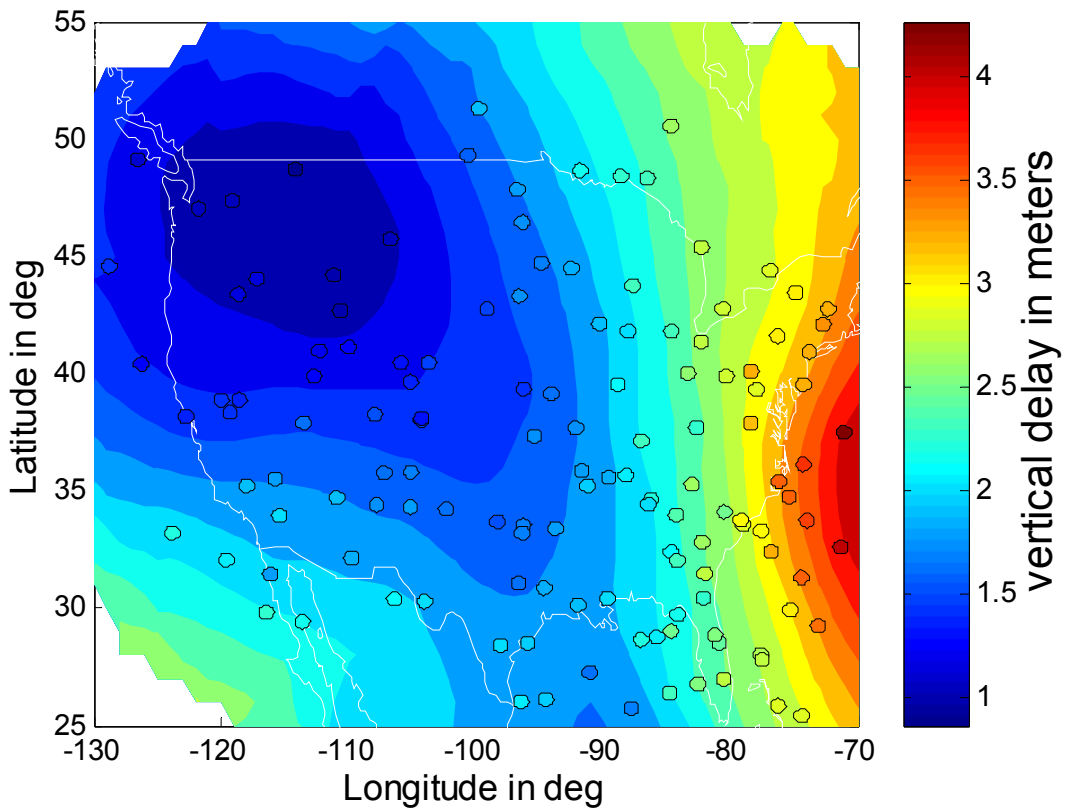

Figure 3.3: Vertical ionospheric delay on July 2, 2000 at 12:00 p.m. E.T

The most important feature of these maps is the presence of a clear trend. This trend is only a part of the overall diurnal trend that we saw in Chapter 2, Section 2.1.1. Added to this main planar trend, whose typical size is on the order of 4000 km, one can detect

smaller features which have a typical size of 500 to 1000 km. The shape of these features is more difficult to describe than the planar trend, and could be considered random, but correlated over distance. One can see that there is also a random noise above these superposed trends. We can interpret the smaller features and the decorrelation as a random spatial process correlated with distance.

3.2.2 DISTURBED CONDITIONS

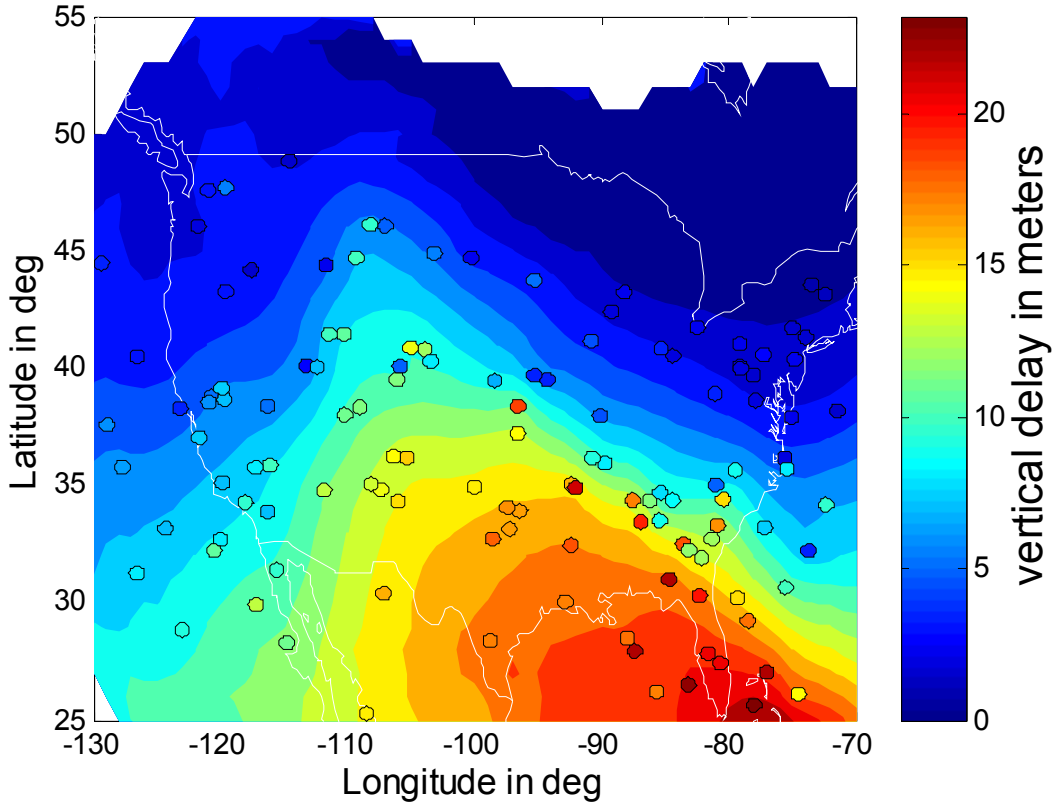

Figure 3.4: Vertical ionospheric delay on July 15, 2000 at 3:20 p.m. E.T

We now turn our attention to snapshots of the ionosphere during disturbed periods. Figures 3.4 and 3.5 show two typical vertical ionospheric maps during some of the worst ionospheric behavior ever observed over the CONUS region. The vertical ionospheric delay is now much higher than in quiet days and the decorrelation of the measurements is larger. Still, we can characterize the vertical ionospheric delay by a combination of a

planar trend, and a stochastic (spatial) process that describes both the decorrelation above the mean map and the smaller features, as we did for the quiet conditions. At this point, the only difference is the magnitude of the decorrelation, and also of the secondary features.

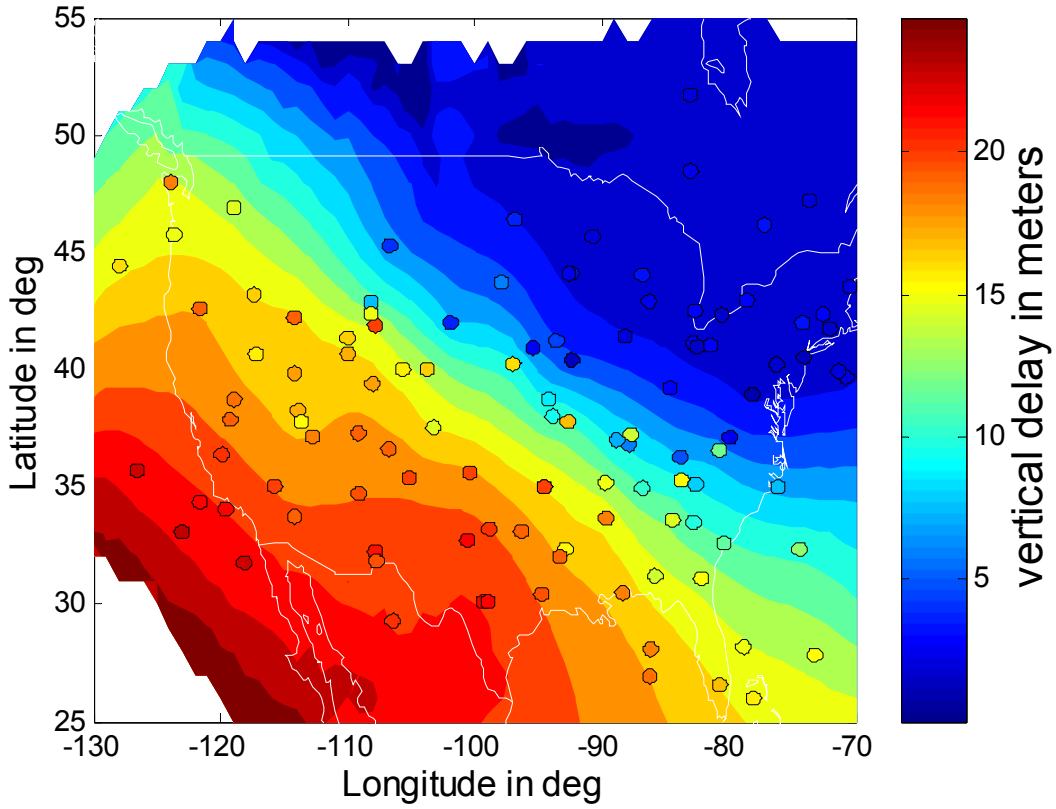

Figure 3.5: Vertical ionospheric delay on April 6, 2000 at 4:50 p.m. E.T

3.3 VERTICAL IONOSPHERIC DELAY MODEL

From the previous section, a possible model for the vertical ionospheric delay is:

$$I(x) = a_0 + a_1 x_{east} + a_2 x_{north} + r(x) \tag{3.3}$$

$I(x)$ is the vertical ionospheric delay at the IPP located in $x = (x_{east}, x_{north})$, here these coordinates can be considered to be respectively the longitude and the latitude. The coefficients a_0, a_1 and a_2 describe the planar trend. The scalar field, $r(x)$, includes the small features superposed on the planar trend and the decorrelation between neighboring measurements. It is the behavior of this field, $r(x)$, that is going to determine the estimation algorithm, so we need to study its properties. The questions we would like to answer about this field are: Is it correlated with distance? Is it reasonable to assume that it is Gaussian (meaning that if we take n measurements and consider them random, they can be seen as coming from a multivariate Gaussian distribution)? We can begin to answer these questions by studying the correlation between pairs of residuals $r(x)$.

3.3.1 SCATTER PLOTS

One of the tools used in geostatistics for exploratory data analysis is the scatter plot. To get the scatter plot, we first compute at a given time frame (i.e. a set of measurements similar to the ones plotted in Figures 4.2 through 4.5) all the pair-wise differences of residuals, and the distances between the IPP measurements. We then plot the pairs formed by distance and difference of residual on the x-axis and the y-axis, respectively. The cloud of points on this diagram, called a scatter plot, will give us an idea of how the difference of residuals increases as a function of distance. Notice that we want to compare the residuals, $r(x)$, of the IPP measurements and not the measurements themselves. If we compared the measurements without removing the planar trend, we would observe that the residuals increase with distance at a linear rate. This linear rate would be due to the deterministic planar trend and would not say anything about the stochastic properties of the field of residuals.

The difficulty now is to break any measurement into planar trend and residual field. De-trending the data is a classical problem in spatial statistics [Cressie]. It is problematic to just fit a plane to the data and compute the residuals. Some measurements might have high leverage on the fit, either due to their large value or their geometry [Hastie], contaminating the original data. For example, a measurement that represents an outlier to the plane will be smoothed if we include it in the planar fit. In our description of the ionosphere, we want to

be aware of such outliers, and want to understand their spatial dependency. We therefore need a method that does not contaminate the original data or that contaminates it in a conservative way. The common practice in spatial statistics is to compare only the pairs of measurements that lie in a direction unaffected by the trend [Journel], [Cressie]. This way we ensure that the original data is not affected by the fit. For our purposes, there are two problems with this approach. First, we end up with very few measurements, thus getting lower statistical significance. Second, we assume that the variability in the direction of the trend is the same as in the direction orthogonal to the trend. These two problems go against our worst case approach.

Instead, we proceed as follows. For each pair of measurements we fit a plane to the remaining measurements (up to a certain radius) and assume this planar fit describes the trend. To our knowledge, our approach is new. The advantage of this approach is that the data used to generate the pairs will not influence the trend. Also, we do not limit the number of pairs. The inconvenience is that we do not take into account the uncertainty that the plane might have due to the geometry of the measured locations, thus not defining the trend correctly. However, this is likely to bias the differences of residuals in a conservative way, because it will tend to give larger residuals. We mitigate this effect by considering only the densest regions in terms of IPP measurements. Figure 3.6 illustrates the procedure outlined above in a one-dimensional example.

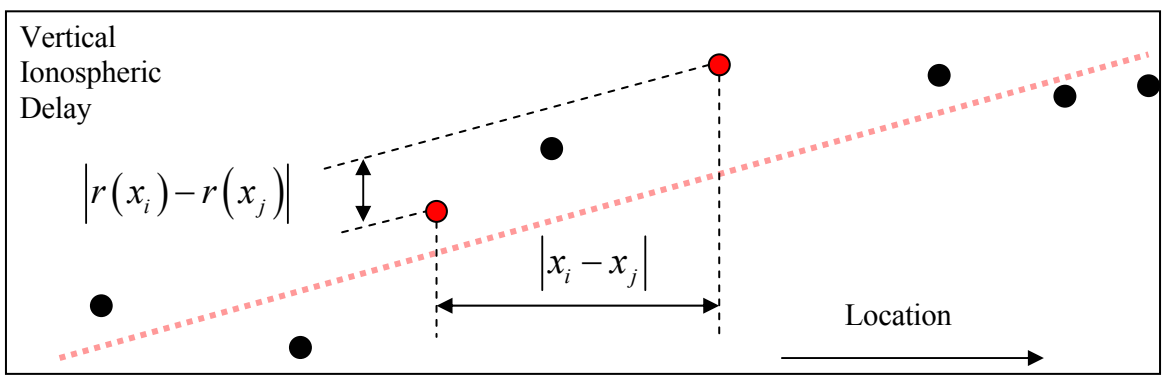

Figure 3.6: Estimating the differences of residuals

As we pointed out in the introduction to this chapter, we cannot visualize the scatter plots corresponding to each time frame: there are too many. Instead, we group them by periods of 24 hours. For quiet days, we can group all the scatter plots in a single plot. The large amount of data points makes it necessary for us to replace the "clouds of points" by bins whose color indicates the number of points lying in that region of the diagram. Figure 3.7 shows a scatter plot corresponding to July 2, 2000. The fit radius for the planar trend was taken to be 2000 km (in section 3.2 we noticed that the planar trend has a characteristic size of 4000 km). We made sure that the planar trend would be well defined by requiring a minimum of 20 IPP measurements.

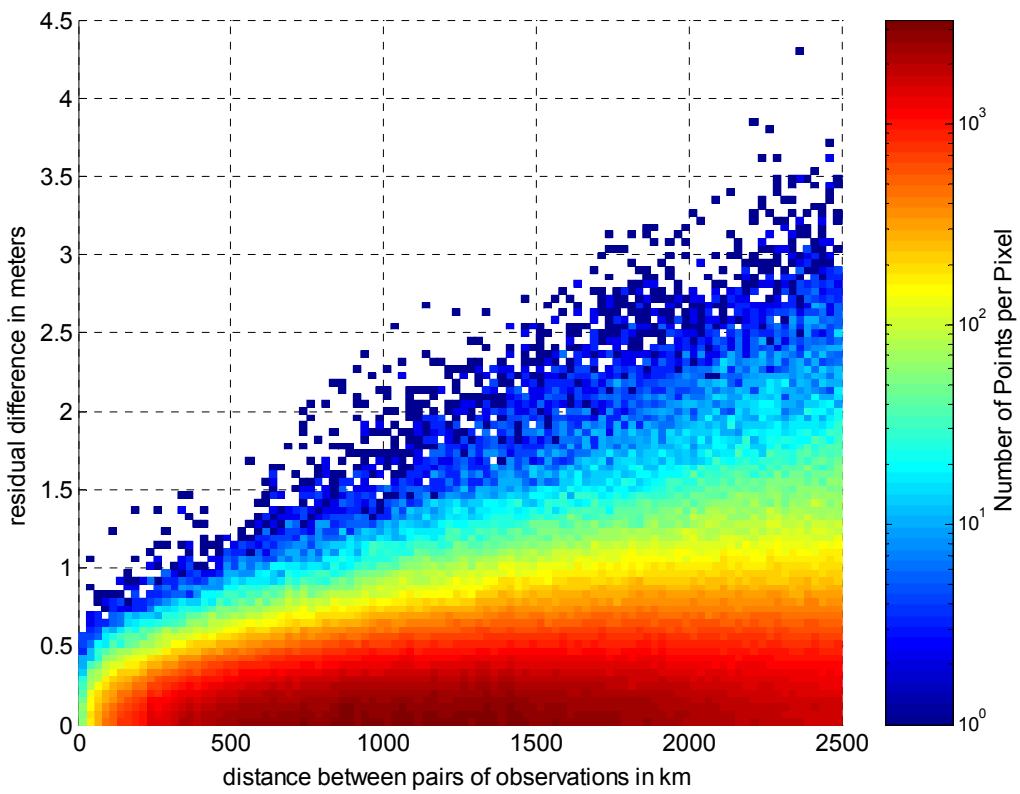

Figure 3.7: Scatter plot of residuals corresponding to July 2, 2000

We can see that the difference between residuals tends to increase as the distance increases between them. We remind the reader here that this decorrelation is not due to the underlying planar trend, since we have removed it. Instead, it is the decorrelation of the

residual field, *r(x)* (a planar trend would have produced linearly increasing residuals). There are two features that are worth mentioning here that will play an important role in the estimation. First the concave shape of the residuals suggests that what we are seeing is a random field [Webster]. Second, we notice that the decorrelation does not tend to zero as the distance goes to zero. There are at least three causes to this effect. First, there might be remaining biases in the measurements (see Chapter 2, Section 2.3). Second, there is the 'slant to vertical error' [Klobuchar]. This error is introduced by the thin shell model: measurements having IPPs close together might have different look angles, in such a way that when we consider the vertical profile of the electron density, the integrated TEC appears discontinuous. Third, the decorrelation of the ionosphere might be so fast at the origin that what is in reality continuous appears discontinuous. A scatter plot similar to Figure 3.7 (not shown here) which only keeps the pairs of measurements coming from the same station or same satellite, thus obviating the slant to vertical error, is still compatible with a discontinuity at the origin. Further evidence of this phenomenon is confirmed by the very fast temporal decorrelation of a single measurement [Datta]. The discontinuity or continuity cannot be observed because there are very few samples at small distances. The general characteristics of the scatter plots for other days remain the same, while the magnitude of the decorrelation changes.

The scatter plots give a picture of the vertical ionospheric delay over a long period. Even though most of the time the differences between residuals are small (red region), there are cases in which they can become large. These outliers are critical for any estimation algorithm with integrity. For each distance we can see that the differences between residuals follow a certain distribution. In the next subsections we examine the nature of this distribution and introduce the variogram which will help us reduce the information contained in the scatter plots into a single function of the distance.

3.3.2 GAUSSIAN ASSUMPTION

We can measure the difference between the actual distribution of differential residuals at a given distance and a Gaussian distribution. For this purpose, we will use the approach taken by [Hansen00]. The idea is to choose a set of probabilities and compute the quantile

corresponding to each of them. For a given distance, if the distribution were Gaussian there would be a specific ratio between the different quantiles. This ratio can be computed from the Gaussian distribution. We can visualize how far the actual distribution is from the Gaussian by normalizing each empirical quantile to the equivalent normal standard deviation and plotting the result. If the normalized quantiles line up, we can say that the distribution is close to Gaussian. The resulting plot is called *sigma containment*. Figure 3.8 shows the sigma containment plot corresponding to the data from July 2, 2000 for different time periods.

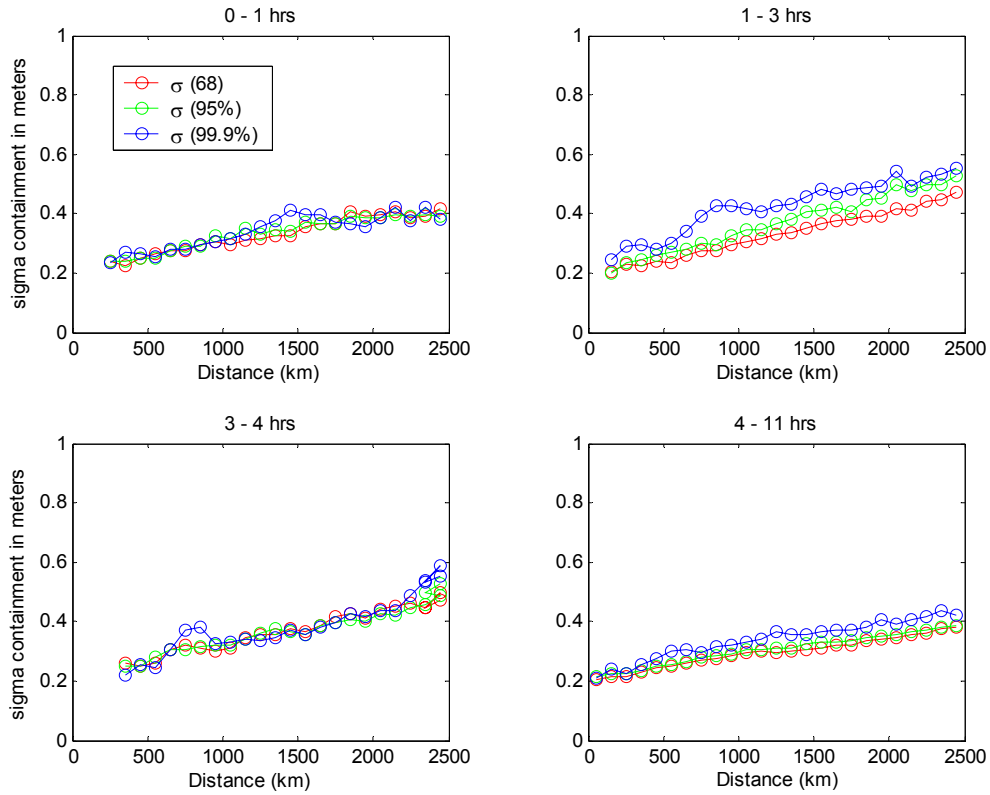

Figure 3.8: Sigma containment plot for July 2, 2000

We observe that the Gaussian assumption is appropriate here. There are cases where the tails of the experimental distribution exceed what would be expected. However, most of the time, the Gaussian hypothesis is a conservative one. The Gaussian distribution allows for infinite tails, when we know that the tails of the true distribution must be clipped.

3.3.3 THEORETICAL AND EXPERIMENTAL VARIOGRAM

The variogram [Webster] is defined as the expectation of the squared difference between two values of a random field at different locations as a function of the distance. This expectation is divided by two:

$$\gamma(x_1, x_2) = \gamma(|x_1 - x_2|) = \frac{1}{2} E\left(\left(r(x_1) - r(x_2)\right)^2\right)$$

The variogram measured from data is called an experimental variogram. In our case, we are interested in the de-trended values (the field of residuals, *r(x)*). The classical formula defining the experimental variogram is:

$$\gamma_{\exp}(d) = \frac{1}{2} \sum_{x_i, x_j \left| |x_i - x_j| \in \left[d - \frac{\delta}{2}, d + \frac{\delta}{2}\right]\right.} \left(r(x_1) - r(x_2)\right)^2$$

Because we do not have pairs of measurements separated by exactly *d*, we need to bin the pairs with a certain tolerance δ. δ was chosen to be 50 km. This is justified by the small change in the residuals over such a distance (see Figure 3.7). This expression supposes that the field has a uniform mean and that there is no dependency on the orientation of the line joining the two measurements, i.e. that the field is isotropic. This is not strictly true at a given time as the variability tends to be larger along a certain axis. However, this axis changes in an unpredictable way, such that it is easier and simpler to assume that the field is isotropic and average over axis orientation. The link between the scatter plot introduced in the previous subsection and the variogram is the following: to obtain the variogram we take the variance of the distribution of differences of residuals for each distance and we divide the result by two. Figure 3.9 shows variograms corresponding to the scatter plot shown in Figure 3.7 at different times. One of the problems of this definition of the variogram is that the variance of the distribution is not sensitive enough to the outliers. However, we have seen in the previous subsection that for each distance, the Gaussian assumption was an acceptable one, in which case the variance is enough to characterize the distribution.

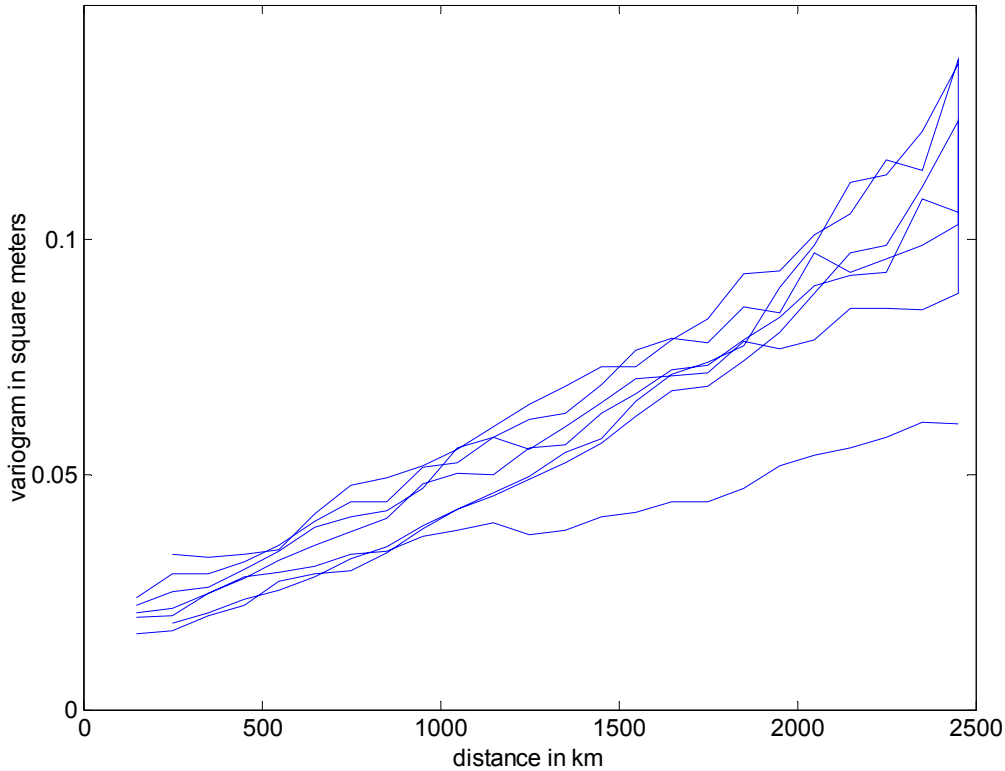

Figure 3.9: Variograms of the residual vertical ionospheric delay
on July 2, 2000

As the scatter plot suggested, we see that the intercept at zero distance is non-zero. This is called the 'nugget effect' in geostatistics [Webster]. By extension, the magnitude of the intercept is called the nugget (this name comes from gold mining, where the gold grade of the soil is discontinuous due to the existence of gold nuggets). We also see that the variogram increases with distance. This means that the decorrelation of the ionosphere increases as the distance between two points increases, even after we remove the planar trend. It is also reassuring to see that the variogram does not increase quadratically. Such a variogram would indicate that there is a deterministic trend unaccounted for, and that the field can no longer be considered random [Webster].

The variogram gives us a quantitative measure of the spatial decorrelation. Because the field is nearly Gaussian or can be overbounded by a Gaussian, the knowledge of the variogram conservatively characterizes the random field.

3.3.4 MODEL VARIOGRAM AND COVARIANCE

Thus far, we have shown how to compute the experimental variogram of the residual field *r(x)*. If we just wanted to have an idea of the behavior of the residual field, we could stop here. However, as stated earlier, the purpose of this chapter is to derive a simple model of the vertical ionospheric delay that can be used for estimation. In particular, the model variogram needs to be 'admissible' [Webster]. An admissible variogram is one that cannot give rise to negative variances when used in variance calculations. In order to understand better why there are constraints on the variogram, we translate the variogram in terms of covariance. We have the following relationship between the two:

$$Cov\left(x_1, x_2\right) = \sigma_\infty^2 - \gamma\left(x_1, x_2\right)$$

Notice that when we write this formula we need to assume that the variogram has a sill, σ_∞^2. In other words it is assumed that the variogram tends toward a finite value σ_∞^2 as the distance approaches infinity. This is not a problem, since we know that the vertical ionospheric delay is bounded, even if the sill cannot be seen in the experimental variogram shown in Figure 3.8. As long as the sill appears at large distances (larger than 5000 km), this is not a critical choice when modeling the variogram. The value of the sill is the variance of two measurements that are completely decorrelated.

Now that we have the covariance between two residuals, we can compute the covariance matrix of a set of residuals, $r(x_1),\ldots, r(x_n)$:

$$Cov\left(r\left(x_i\right), r\left(x_j\right)\right) = E\left(r\left(x_i\right) r\left(x_j\right)\right) = C_0\left(\left|x_i - x_j\right|\right) \tag{3.4}$$

The only requirement for the covariance to be admissible is that any linear combination of the residuals must have a positive variance. This is equivalent to requiring that the

covariance matrix be positive definite. An admissible and widespread model for the variogram is the exponential model [Webster]. This model variogram has the three properties we need. There is a nugget effect at the origin, the variogram increases linearly near the origin, and there is a sill. The expression for the model variogram is:

$$\gamma(d) = c\left(1 - e^{-\frac{d}{a}}\right) + \nu \tag{3.5}$$

The value of the variogram between two points is only a function of the distance between the two points, d. We should point out again that the value of the variogram at the origin is zero, and that there is a discontinuity, the nugget. The different parameters of the variogram were chosen visually, so that the function given by (3.5) would approximately overbound different experimental variograms for a quiet day. The parameters retained for CONUS were: c = 2 m^2, a = 32000 km, and ν = .04 m^2. Figure 3.10 shows the experimental variogram (blue) and the model variogram chosen (black). In Chapter 5, we will primarily use the covariance, given by:

$$C(d) = c + \nu - \gamma(d) = ce^{-\frac{d}{a}} \tag{3.6}$$

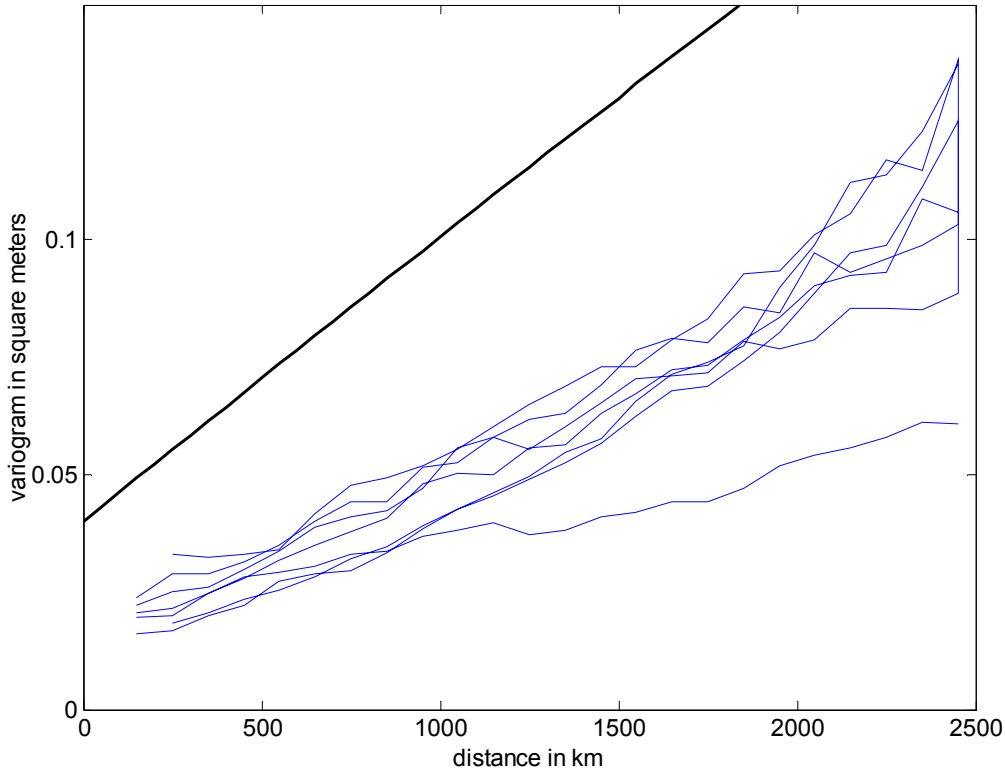

Figure 3.10: Experimental variogram and model variogram

3.3.5 DECORRELATION DURING A STORMY DAY

In the last two subsections we have seen the scatter plot and the variogram for a quiet ionospheric day (July 2, 2000). We now turn our attention briefly to more disturbed days: April 6, 2000 and September 8, 2002.

We show in Figure 3.11 sigma containment plots corresponding to the onset of the April 6, 2000 ionospheric storm, which is one of the worst ionospheric storms that the CONUS region has seen. We can see that the differences between residuals increase more rapidly with distance, and, what is more important, that the discontinuity at zero distance is much larger. Again this discontinuity might be due to the slant to vertical error due to the thin shell model, but it is also due to a very fast decorrelation of the electron density during

storms. We can also see that in the third plot, the Gaussian hypothesis is more problematic. This due to the mixing between variograms corresponding to different behaviors.

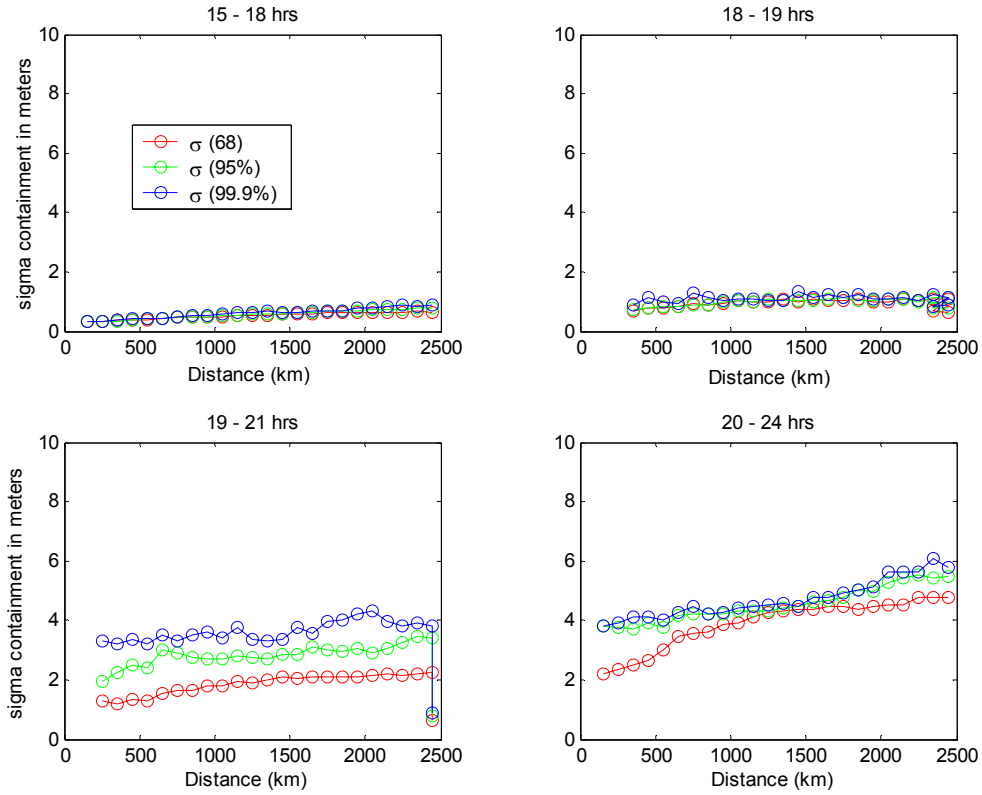

Figure 3.11: Sigma containment plots during April 6,2000

Overall, the sigma containment corresponding to the quiet day (Figure 3.8) and the sigma contatinment corresponding to the stormy day (Figure 3.12) are similar in shape. That suggests that we can describe the ionosphere with a similar model (planar trend plus residual field) in both cases. The only thing that changes is the value of the parameters describing the residual random field, i.e., the variogram.

Figure 3.12 shows the sigma containment for September 8, 2002 where the shape of the decorrelation is the same as for a quiet day, and the Gaussian hypothesis holds.

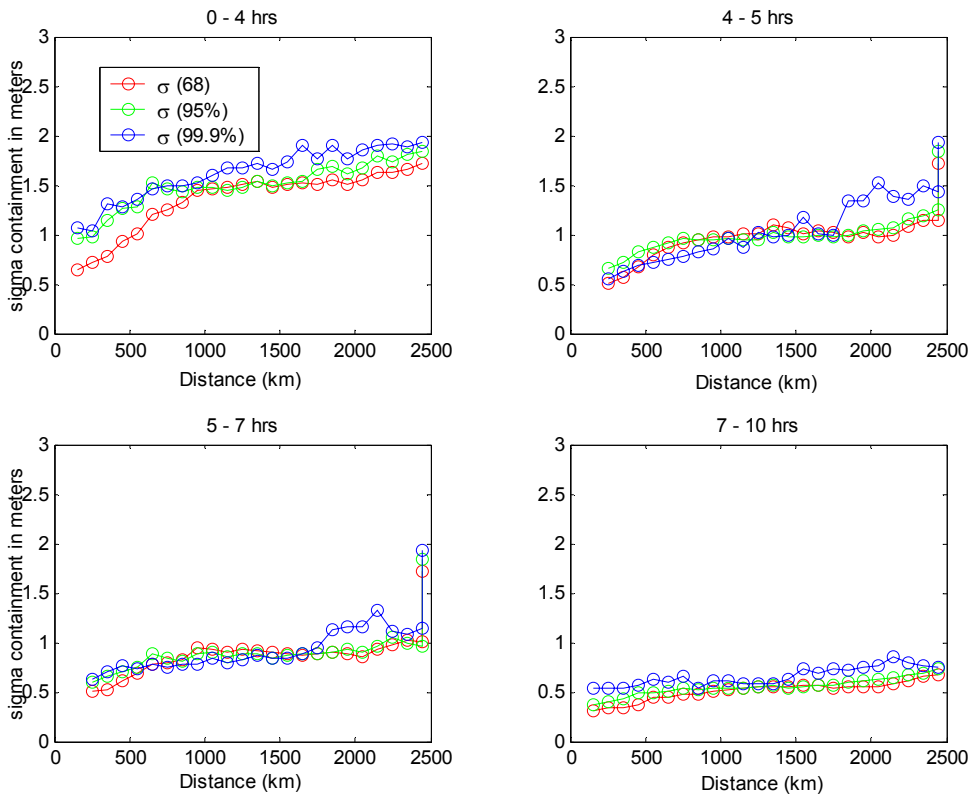

Figure 3.12: Sigma containment plots during September 8, 2002

These remarks lead us to postulate that there exists a parameter u such that the variogram can be written:

$$\gamma(d) = u^2 \gamma_0(d) \tag{3.7}$$

The function γ_0 is the variogram corresponding to quiet conditions, which was introduced in the previous section. The parameter u describes the degree of disturbance in the ionosphere: a larger u corresponds to a disturbed ionosphere, and u=1 corresponds to a quiet ionosphere. This formula supposes that the slope of the variogram is proportional to the nugget effect, which is close to what we observe.

3.4 IONOSPHERIC IRREGULARITIES

We now have a simple model of the vertical ionospheric delay which has two components: a planar trend (the deterministic component) and a random Gaussian residual field – the stochastic component. This last component is characterized by a variogram (or, equivalently, by the covariance) indexed by the parameter u, which describes the roughness of the vertical ionospheric delay. It is important at this point to be aware of the limitations of this model. The main limitation is the occasional lack of stationarity in the random field. The vertical ionospheric delay might be well described by a different variogram in different regions of the ionosphere. For a region like CONUS, what we would like to know is the shape of such irregularities: Can we have an irregularity in the middle of an otherwise quiet region? Can we have an irregularity coming from the side of an otherwise quiet region? Notice that here irregularity is any behavior that is not accounted for in the assumed model. It is exceedingly difficult to tell simply by looking at a certain situation whether it is compatible with a certain random model or not. The ultimate answer to that question will be given in Chapter 6, when we will test the bounding algorithm based on the model given above. In this section we will mostly hint at the possible drawbacks of such a characterization, and cite previous research that has attempted to characterize the lack of stationarity in the vertical ionospheric delay maps.

3.4.1 ISOLATED IRREGULARITIES

First we examine the threat constituted by isolated irregularities. Isolated irregularities include regions of the ionosphere where the surrounding region is well observed and well described by our model, but contained inside is a region that is not. Let us imagine that we have a dense set of measurements over CONUS and that a certain variogram similar well describes the random field of residuals. Suppose now that in the middle of it, there is an unobserved region (for example a disk of a certain radius) without any measurements. Is it possible that the residuals corresponding to that region might not be compatible with the variogram valid elsewhere? Again this question is difficult to answer, mainly because a given variogram does not prohibit a given situation from happening, it only makes it less likely. An attempt to answer this question was made in [ITM], and we now summarize the

methodology and conclusions of that work. The approach taken in that work used data deprivation in order to simulate the lack of data. The following procedure was applied: for any given measurement, all the measurements surrounding it at the same time frame up to a certain radius were excluded. The purpose of this step was to simulate the lack of data in that region. Then, the remaining measurements were checked to determine compatibility with a planar trend with a certain decorrelation around it (this was done by computing the sum of squared residuals). If they were found to be compatible, the residuals of the excluded measurements were computed and their size was compared to the observed residuals. This analysis determined that isolated irregularities were not a serious threat.

3.4.2 GRADIENTS

The other threats we need to consider are the 'gradients.' In this context, gradients are a sharp change in vertical ionospheric delay. If this sharp change occurs in an already disturbed region, it might not be a problem because the measurements will reflect the general disturbance. If, on the other hand, the gradient divides the region into a disturbed side and a quiet side, or two distinct quiet regions, there might a problem. The same analysis that was done for the isolated irregularities was carried out for gradients in [ITM]. Now, instead of excluding disks around every single measurement, whole half planes were excluded; then the same procedure (computing the observed residuals and comparing them with the 'unobserved' ones) was applied. This research determined that gradients could constitute a serious threat to stationarity. In Figure 3.13 we show the vertical ionospheric delay map toward the end of the ionospheric storm of July 15-16, 2000. One can see that the region above 35 degrees latitude corresponds to a quiet day ionosphere. South of it we see that the vertical ionospheric delays increase to more than 20 meters, and so does the variability. If we were to estimate the vertical ionospheric delay with the IPPs with a delay below 8 meters, the estimation error could be over 15 meters in Florida, which could lead to slant errors above 50 meters. (This picture also illustrates the difficulty of such studies as we are always limited by the sampling that was done that day.) As we will see later, situations like the one pictured in Figure 3.13 are currently handled in WAAS by the *threat model* [Sparks01], which will be outlined in Chapter 4, in conjunction with the algorithm designed in this thesis.

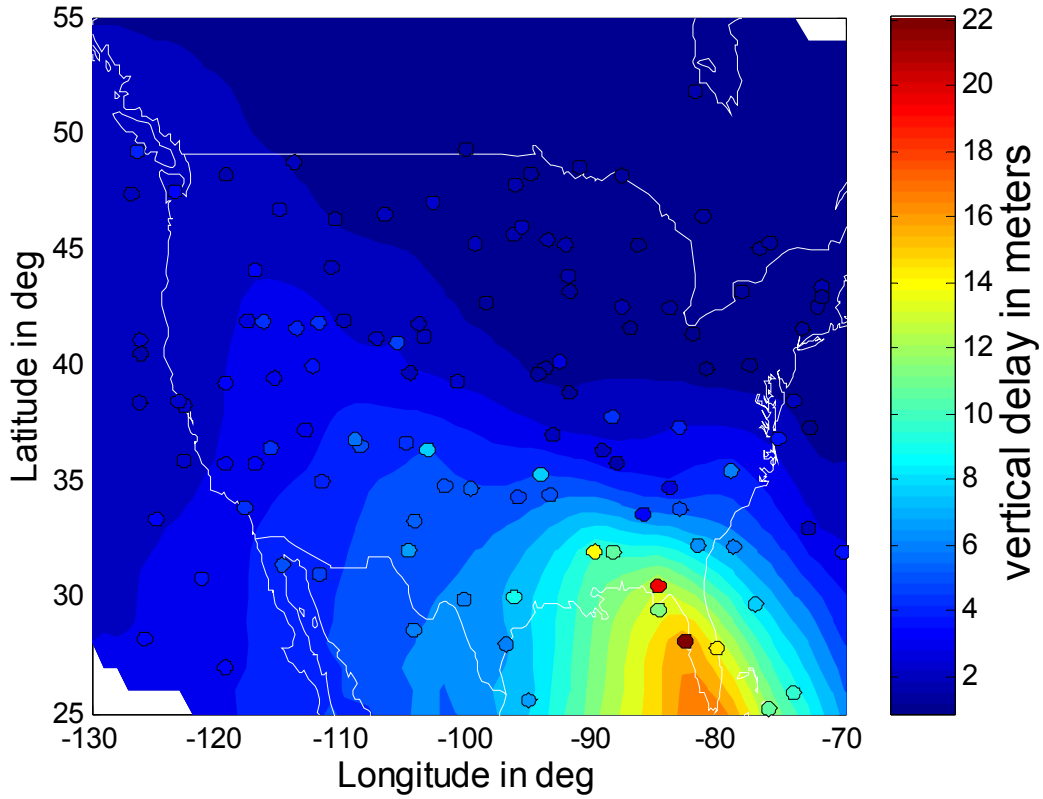

Figure 3.13: Very large gradient during the July 15-16, 2000 ionospheric storm

3.5 TEMPORAL DECORRELATION

So far, we have dealt with the problem of determining the structure of the vertical ionospheric delay as a static problem. We have ignored the fact that the vertical ionospheric delays evolve with time (at a fixed location in the thin shell). A model for the random behavior of the ionosphere that includes both the spatial decorrelation and the temporal decorrelation would be difficult to obtain and express for any time period. Among other problems, we would have to include the deterministic temporal behavior in our description. However, as we saw in Chapter 2, the ionospheric corrections and bounds need only be valid for ten minutes. Based on the speed of the reference in which the ionosphere appears stationary, 360 degrees of longitude per day, we see that the underlying trend shifts by 2.5 degrees longitude, or about 200 km at typical CONUS latitudes, over

ten minutes. This is a small distance compared to the characteristic distances of the problem (see Figures 3.7 and 3.8). The vertical ionospheric delay maps shown above suggest that as a first approximation we can consider that the underlying trend is fixed, and that the evolution is not deterministic but stochastic.

To visualize the short term temporal and spatial structure of the vertical ionospheric delay, we form a generalized scatter plot. That is, we add the time as a new dimension with the pairs of IPP measurements separated by distance and time. Instead of having two axes like in Figure 3.7, we will have three: difference of residuals, distance between IPP locations and time delay between the two IPP measurements. Practically, we examine 10 minute periods (600 seconds) where each time step is 200 seconds. Each period contains three time frames. For a given measurement we fit a plane (planar fit radius is taken to be 2000 km) at the time frame and remove this from all three periods. This step should approximately remove the trend. After that, we form the difference of residuals between our test measurement and each of the other ones. We apply this to all the measurements and plot each difference of residual in a three dimensional scatter plot. Since the resolution in the time domain is only four, we can plot it as a series of four scatter plots similar to Figure 3.7. In Figure 3.14, we show the sigma containment plots corresponding to these scatter plots. We can see that there is not much difference between the four plots. For time lags below 10 minutes, there is no noticeable change in the structure of the residual random field. These plots suggest that the model chosen to describe the vertical ionospheric delay is stationary over periods of 10 minutes, which simplifies greatly the design of the algorithm, since now the problem becomes purely spatial and not temporal.

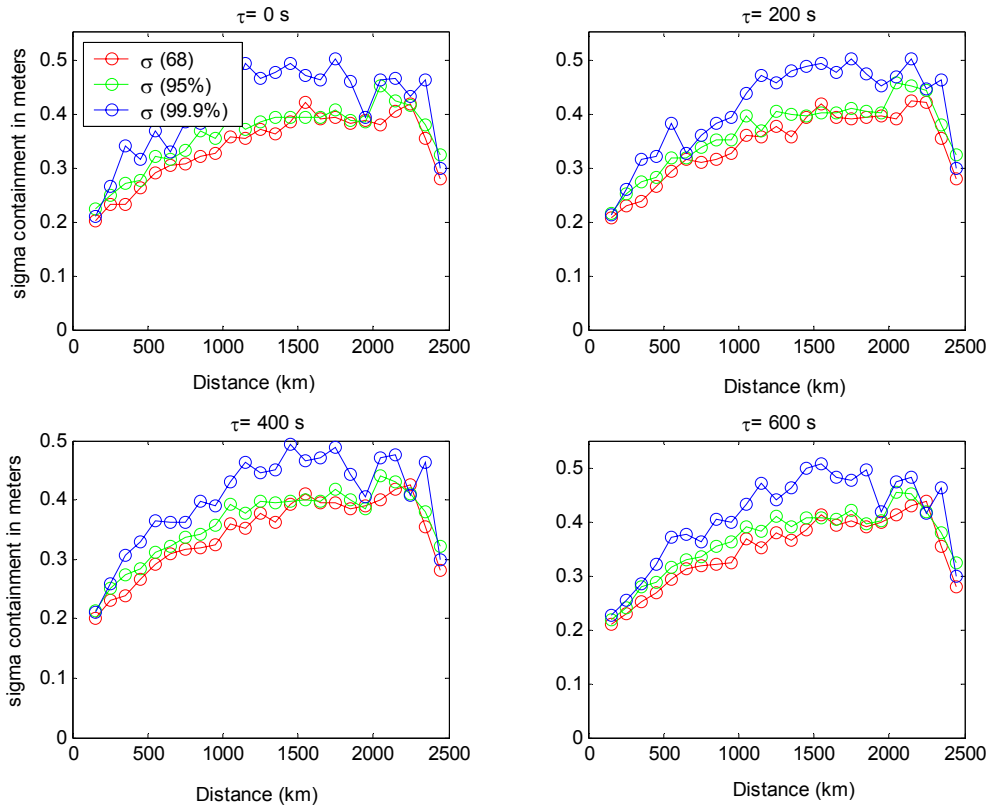

Figure 3.14: Spatial and temporal decorrelation during July 2, 2000

Unfortunately, a difficulty arises due to the lack of stationarity during the onset of storms in the time domain. During these times the vertical ionospheric delay can increase rapidly. Here again, as for the spatial irregularities, this difficulty is treated by a piece of the algorithm that deals specifically with threats that lie outside the assumed model. This part of the algorithm will be dealt with in Chapter 5 and Chapter 6.

3.6 CONCLUSION

In this chapter, we have extracted from the post-processed TEC data a simple characterization of the vertical ionospheric delay on a quiet day:

$$I(x) = a_0 + a_1 x^{(east)} + a_2 x^{(north)} + r(x) \tag{3.8}$$

In this equation, *r(x)* is a random residual field that has zero mean and a covariance between two measurements, *i* and *j*:

$$Cov\left(r(x_i), r(x_j)\right) = E\left(r(x_i) r(x_j)\right) = C_0\left(\left|x_i - x_j\right|\right) \tag{3.9}$$

From now on we will call this model the *nominal* covariance: it is a covariance that describes conservatively a quiet vertical ionospheric delay. We have also seen that this characterization is also sound for a stormy day provided we suppose that the covariance is:

$$Cov\left(r(x_i), r(x_j)\right) = u^2 C_0\left(\left|x_i - x_j\right|\right) \tag{3.10}$$

The parameter *u* describes the roughness of the vertical ionospheric delay: the larger it is the more disturbed the ionosphere is.

We have also seen that within a certain time lag, the vertical ionospheric delay field is stationary and that there is no need for a complex spatial and temporal stochastic model (the spatial one is enough).

The last part of this chapter has examined where this model could fail. Potential failures occur when there are severe stationarity violations, both spatial and temporal. However, it has been determined elsewhere [ITM], [Sparks01] that those seem to be limited to irregularities coming from the side of the well observed regions (an irregularity cannot appear in the middle of CONUS). The ultimate answer to the question to whether the assumed model is conservative enough will come when we check the final algorithm against real data in Chapter 6.

Chapter 4

Ionospheric Estimation Algorithm

In Chapter 3 we found a characterization of the vertical ionospheric delay that is comprised of a deterministic part (the planar trend) and a stochastic part (the residual field.) Here, we are going to use this model to design the estimation algorithm. The estimation algorithm must provide at each time epoch and at each location an estimate of the vertical ionospheric delay and, more importantly, a hard bound on the possible error of this correction. As seen in Chapter 2, this bound must provide a Probability of Hazardously Misleading Information below $2.25X10^{-8}$. This error bound must be as small as achievable, such that the availability is as high as possible given the current information. The estimation problem can be seen as a constrained optimization problem: for each location in the thin shell we minimize the error bound such that the PHMI is below $2.25X10^{-8}$. For a given location, we can write the following problem:

$$\text{Minimize } \Delta \text{ such that } \operatorname{Prob}\left(\left|I_{true} - I_{est}\right| > \Delta\right) \leq 2.25 * 10^{-8} \tag{4.1}$$

In this equation, I_{est} is the estimated vertical ionospheric delay, I_{true} is the true vertical ionospheric delay, and Δ is the estimated error bound. This equation must hold for any user of the estimate. There are several difficulties to overcome. First, the underlying ionospheric model is uncertain: the covariance model is indexed by the parameter u, which is unknown. This means that the roughness of the vertical ionospheric delay will have to

be safely bounded. This is all the more difficult as the vertical ionospheric delay measurements are randomly scattered over the region of interest. Second, the real time measurements used by the estimation algorithm have measurement noise (see Chapter 2.) This would not be a problem by itself if the underlying ionospheric behavior were known. We have a safe upper bound on the measurement noise, and it is easy to translate this to the error bound of the estimate. What is more problematic about the measurement noise is the fact that it blinds our observability of the underlying ionospheric behavior (the parameter u).

Due to the multiple layers of complexity of the estimation algorithm, in this chapter we will proceed as follows. In the first section we will assume that there is no model uncertainty (the parameter u is known). This situation is easily modeled and we can find the exact solution, which turns out to be kriging. In the second section we will drop the assumption that we know the underlying model and introduce a factor that accounts for this uncertainty, the inflation factor.

4.1 ALGORITHM ASSUMING KNOWLEDGE OF THE MODEL

We remind the reader here that when we say that we assume the model to be known we mean that the covariance of the residual field is known:

$$Cov\left(x_k, x_l\right) = E\left(r\left(x_k\right) r\left(x_l\right)\right) = C\left(x_k, x_l\right) \tag{4.2}$$

From now on we will call the residual field the process noise, as opposed to measurement noise. The problem is as follows: Given n IPP measurements, $I(x_1)\ldots I(x_n)$, what is the best estimate and the smallest error bound on the estimate that can be given at a certain location x? In this section we will see that we can solve Problem (4.1) exactly. We will start with the general form of the estimator. After that, we will express the estimation variance as a function of the variables of the problem. The last step will consist of a straightforward constrained minimization problem that will lead to the kriging equations. The beginning of

this section is not new. However, some modifications are needed that are not easily found in the literature. In addition, the final expression of the kriging equations is original. This section will also allow us to introduce several notations.

4.1.1 UNBIASED LINEAR ESTIMATOR

The estimate at location x is going to be a function of the surrounding measurements $I_{\text{meas}}(x_1)\ldots I_{\text{meas}}(x_n)$. We are going to limit the search of such a function among linear functions, that is, the estimate is going to be a linear combination of the measurements:

$$I_{est}\left(x\right)=\sum_{k=1}^{n}\lambda_k I_{meas}\left(x_k\right) \tag{4.3}$$

We only need to find the coefficients $\lambda_1,\ldots,\lambda_n$. Now we remind the reader of the different components of the measurement. We have:

$$I_{meas}\left(x_k\right)=a_0+a_1x_k^{(east)}+a_2x_k^{(north)}+r\left(x_k\right)+m\left(x_k\right) \tag{4.4}$$

The first four terms were examined in Chapter 3. The first three terms define the underlying planar trend (deterministic component) and the fourth term represents the residual field (the process noise.) The fifth term, $m(x_k)$, is the measurement noise specific to the receiver and the satellite pair (this term does not depend on ionospheric behavior). As stated in Chapter 2, this noise can be assumed to be zero mean and Gaussian, provided that a large enough standard deviation is taken to bound the tails of the true error distribution. The estimator should be unbiased, i.e.:

$$E\left(I_{est}\left(x\right)\right)=E\left(\sum_{k=1}^{n}\lambda_k I_{meas}\left(x_k\right)\right)=E\left(I_{true}\left(x\right)\right)=E\left(a_0+a_1x^{(east)}+a_2x^{(north)}+r\left(x\right)\right) \tag{4.5}$$

By noticing that $E(r(x))=E(r(x_k))=E(m(x_k))=0$ we can write, after some algebra, that:

$$G^T\lambda=\begin{bmatrix}1 & x^{(east)} & x^{(north)}\end{bmatrix}^T=X \tag{4.6}$$

The definitions of G and λ are the following:

$$G^T = \begin{bmatrix} 1 & \cdots & 1 \\ x_1^{(east)} & \cdots & x_n^{(east)} \\ x_1^{(north)} & \cdots & x_n^{(north)} \end{bmatrix} \quad , \qquad \lambda^T = \begin{bmatrix} \lambda_1 & \cdots & \lambda_n \end{bmatrix}$$

The constraints in (4.5) show that the λ coefficients must be compatible with the planar trend. Notice that one could generalize this result to any kind of trend by augmenting the matrix G appropriately.

4.1.2 ESTIMATION VARIANCE

The estimation variance is defined as the following expectation:

$$\sigma_{est}^2 = E\left(\left(I_{true}(x) - I_{est}(x)\right)^2\right) \tag{4.7}$$

The fact that the estimator is unbiased allows us to write:

$$\sigma_{est}^2 = E\left(\left\{r(x) - \left(\begin{bmatrix} r(x_1) & \cdots & r(x_n) \end{bmatrix} + \begin{bmatrix} m(x_1) & \cdots & m(x_n) \end{bmatrix}\right)\lambda\right\}^2\right)$$

This expression can be developed. Noticing that the measurement noise, m, and the ionospheric process noise, r, are uncorrelated, after some algebra we have:

$$\sigma_{est}^2 = \lambda^T\left(C(x_k, x_l) + M(x_k, x_l)\right)\lambda - 2\lambda^T C(x, x_k) + C(x, x) \tag{4.8}$$

Here, $C(x_k, x_l)$ is an n by n matrix whose elements are computed using the assumed covariance for the process noise. Similarly, $M(x_k, x_l)$ is the covariance matrix of the measurement noise, m. $C(x, x_k)$ is a vector whose elements are the covariance between $r(x)$ and each of the residuals at the measurement locations, $r(x_k)$. $C(x, x)$ is the value of the covariance at a distance of zero and does not depend on the location x.

4.1.3 ERROR BOUND AND ESTIMATION VARIANCE

In this subsection we explain the relationship between the estimation variance and the error bound. First, recall the assumption that the process noise is Gaussian. Under this assumption the random variable $I_{true}(x) - I_{est}(x)$ is zero mean Gaussian with a standard deviation σ_{est}. Considering all the possible realizations of the random field, we have:

$$\text{Prob}\left(\left|I_{true} - I_{est}\right| > \Delta\right) = 2Q\left(\frac{\Delta}{\sigma_{est}}\right) \tag{4.9}$$

Here Q is the cumulative distribution function of a Gaussian. If we want this probability to be $2.25\text{X}10^{-8}$, we need to have $\Delta = K\sigma_{est}$ where K=5.92. Minimizing Δ is then equivalent to minimizing σ_{est}^2. This is what is done in the next subsection.

4.1.4 FINDING THE OPTIMAL COEFFICIENTS: KRIGING EQUATIONS

The problem of determining the λ coefficients can be cast as a constrained optimization problem:

$$\text{Minimize } \sigma_{est}^2(\lambda) = \lambda^T\left(C(x_k, x_l) + M(x_k, x_l)\right)\lambda - 2\lambda^T C(x, x_k) + C(x, x)$$

$$\text{Subject to } G^T\lambda = X \tag{4.10}$$

This is a classical optimization problem and there are many methods to solve it. We show here the one using Lagrange multipliers. The Lagrangian is:

$$L(\lambda, \mu) = \sigma_{est}^2(\lambda) + \mu^T\left(G^T\lambda - X\right)$$

We find the solution of the minimization problem by setting the partial derivatives of L to zero. We skip the details of this calculation. At the end, we get the following matrix equation for the coefficients:

$$\begin{bmatrix} C(x_k,x_l)+M(x_k,x_l) & G \\ G^T & 0 \end{bmatrix}\begin{bmatrix} \lambda \\ \mu \end{bmatrix}=\begin{bmatrix} C(x,x_k) \\ X \end{bmatrix}$$

The solution is then:

$$\begin{bmatrix} \lambda \\ \mu \end{bmatrix}=\begin{bmatrix} C(x_k,x_l)+M(x_k,x_l) & G \\ G^T & 0 \end{bmatrix}^{-1}\begin{bmatrix} C(x,x_k) \\ X \end{bmatrix} \tag{4.11}$$

This equation defines the coefficients, but does not give an idea of the behavior of the coefficients. We would like to know how the λ coefficients depend on the covariance properties of the process noise and the measurement noise. For that, we invert the matrix by blocks:

$$\begin{bmatrix} C(x_k,x_l)+M(x_k,x_l) & G \\ G^T & 0 \end{bmatrix}^{-1}=\begin{bmatrix} W-WG\left(G^TWG\right)^{-1}G^TW & WG\left(G^TWG\right)^{-1} \\ \left(G^TWG\right)^{-1}G^TW & -\left(G^TWG\right)^{-1} \end{bmatrix} \tag{4.12}$$

The definition of W is:

$$W=\left(C(x_k,x_l)+M(x_k,x_l)\right)^{-1}$$

The λ coefficients are defined by:

$$\lambda=\left(W-WG\left(G^TWG\right)^{-1}G^TW\right)C(x,x_k)+WG\left(G^TWG\right)^{-1}X \tag{4.13}$$

This expression shows the different components of the estimate: the stochastic component in the first term and the deterministic component in the second one. Since the matrices involved in this formula are going to appear several times in the next pages, we include the notation:

$$P=\left(W-WG\left(G^TWG\right)^{-1}G^TW\right) \quad , \quad H=WG\left(G^TWG\right)^{-1}$$

By replacing these expressions in the formula for the estimation variance (4.8) we get:

$$\sigma_{est}^2 = C(x,x) + X^T \left(G^T W G\right)^{-1} X - C(x,x_k)^T PC(x,x_k) - 2C(x,x_k)^T HX \qquad (4.14)$$

We can give an approximate interpretation to this formula. The first two terms correspond to the estimation variance that we would get if we were to estimate the vertical ionospheric delay by the value of the estimated trend. The last two terms correspond to the reduction we can achieve by taking into account the spatial properties of the process noise –or residual field. For example, if the process noise was completely decorrelated from one location to another, then $C(x,x_k)$ would be zero, so there would not be any reduction in the estimation variance, which is what we expect.

The formulas presented in this subsection differ from the classical kriging equations [Webster], [Cressie], [Journel] in two ways. First, we have included the effect of measurement noise (it has not previously been included in kriging) which has a different role in the estimation formulas. Second, classical formulations of kriging stop at Equation (4.11), whereas here we have obtained a decomposition of the different contributions in the coefficients.

4.1.5 KRIGING MAPS

In this subsection, we give an example of the estimation variance map produced by the equations derived in the previous section over the CONUS region. We will also provide some details, among which are the choice of measurements and the coordinate and metric system used on the thin shell, which are the same that are used in the current WAAS ionospheric estimation algorithm [Walter00].

The distance used between IPP measurements on the thin shell is the true distance between the two locations, as opposed to the great circle distance. The coordinate system used is a local one. For each IPP location we use the East –North –Up coordinate system [Enge01]. The curved surface of the thin shell is flattened by just considering the East and North coordinates. The map gets 'deformed' slightly as we depart from the location of interest, but it has been shown to be insignificant. As the shape of the variogram suggests in Figure 3.8, a difference in distance of 100 km is not going to modify the covariance matrix

significantly. For each location we do not use the whole set of IPP measurements. Instead, the measurements lying within a certain radius (the search radius) are selected. The determination of the search radius is the same as in the current WAAS estimation algorithm. If there are more than 30 measurements lying within 2100 km, it is the smallest radius containing 30 measurements (but larger than 800 km). Otherwise, it is 2100. An estimate is only given if there are at least 10 measurements. At this state of the design, the search radius is not a key parameter: the coefficients assigned to measurements lying at more than 1000 km have virtually no weight on the estimate.

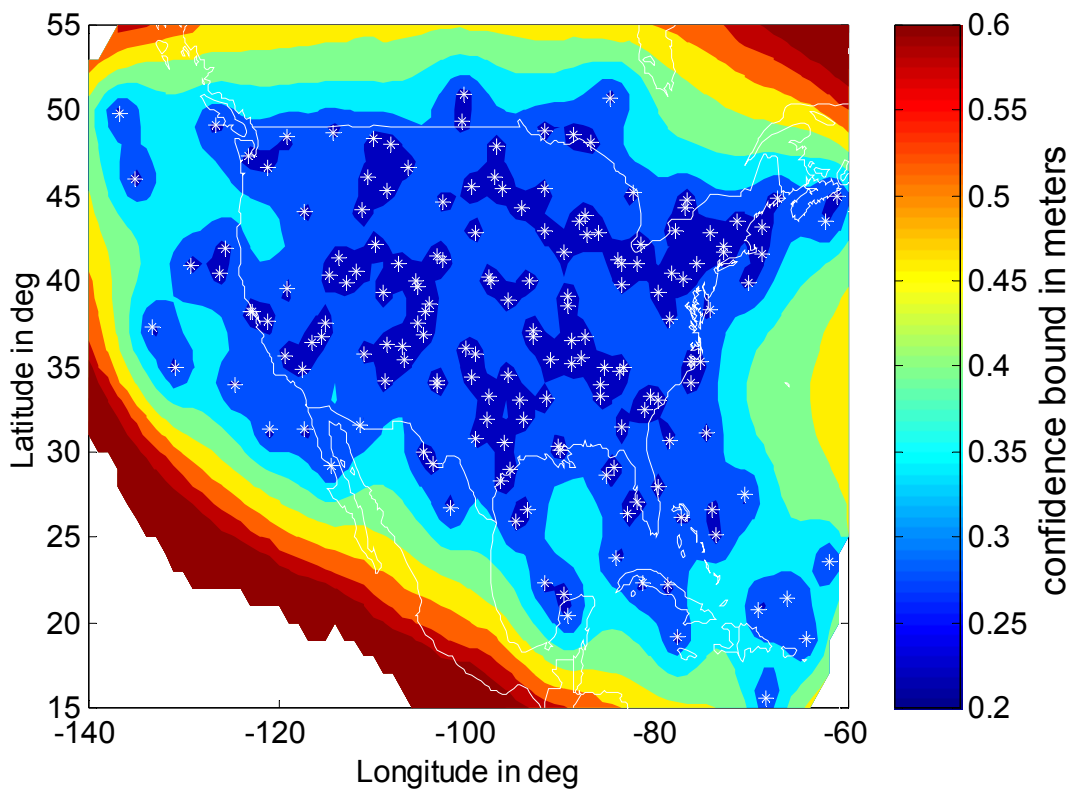

Figure 4.1: Map of the estimation variance on July 2, 2000 at 12:00 pm E.T.

To generate the map shown in Figure 4.1, we took all the measurements at a certain time during July 2, 2000 from the post-processed data. The covariance model chosen here is the one specified in Section 4.3.4, Equation (4.5). For this example, the measurement noise is taken to be zero. The covariance used is the nominal covariance derived in Chapter 3.

This is approximately true for the post-processed data and it allows us to see the effect of the spatial decorrelation on the error bounds.

In this map of σ_{est} we see how the regions rich in measurements (the white stars) have the lowest estimation error, and how the estimation error increases as we depart from them. The degradation in the error bound is a function of the variogram derived in Chapter 3. It is the behavior of the estimation variance and its link to the estimated stochastic properties of the ionosphere that makes kriging so attractive for an ionosphere estimation algorithm requiring tight error bounds. If the vertical ionospheric delay process noise were always bounded by the model covariance (which it is in all quiet days, that is, 99% of the time), we could stop the analysis here.

Unfortunately, as we saw earlier, the process noise varies greatly. To solve this problem, we could assume a nominal process noise that bounds all the ionospheric behavior ever seen, but given the characteristics of the worst ionospheric conditions, that would mean low availability [Walter00]. We therefore need to look for another solution, one that makes a better use of the information when we compute the correction.

4.1.6 ESTIMATION VARIANCE DEPENDENCE ON MEASUREMENT NOISE

In this subsection we study how the estimation variance computed using kriging depends on the measurement noise and on the process noise. The purpose here is to show that, when the process noise is well known, the measurement noise has very little effect on the estimation variance. The results from this section will support approximations made below. First, let us divide the estimation variance into two terms: the part coming from the process noise and the part coming from the measurement noise. We have:

$$\sigma_{est}^2 = \sigma_{process}^2(\lambda) + \sigma_{meas}^2(\lambda)$$

In this equation:

$$\sigma_{process}^2(\lambda) = \lambda^T C(x_k, x_l)\lambda - 2\lambda^T C(x, x_k) + C(x, x)$$

$$\sigma_{meas}^2(\lambda) = \lambda^T M(x_k, x_l)\lambda$$

There is an important qualitative difference in these two expressions. The first one cannot go to zero as the number of measurements increases, whereas the second one can. This comes from the fact that measurement noise is averaged out as the number of measurements increases. On the other hand, the process noise cannot go to zero because even with an infinite number of measurements there will always be random behavior in between the measurements. The first term has a lower bound in ν, the nugget effect in the covariance. The purpose here is to show that in the real system there is a large dissymmetry between the two terms. The first one is always much more important than the second one. Figure 4.2 shows the ratio $\sigma_{meas.}^2(\lambda)/\sigma_{process}^2(\lambda)$ computed using the nominal model on July 2, 2000. Notice that this ratio is only a function of the measurement covariance, the process noise covariance and the geometry of the measurements.

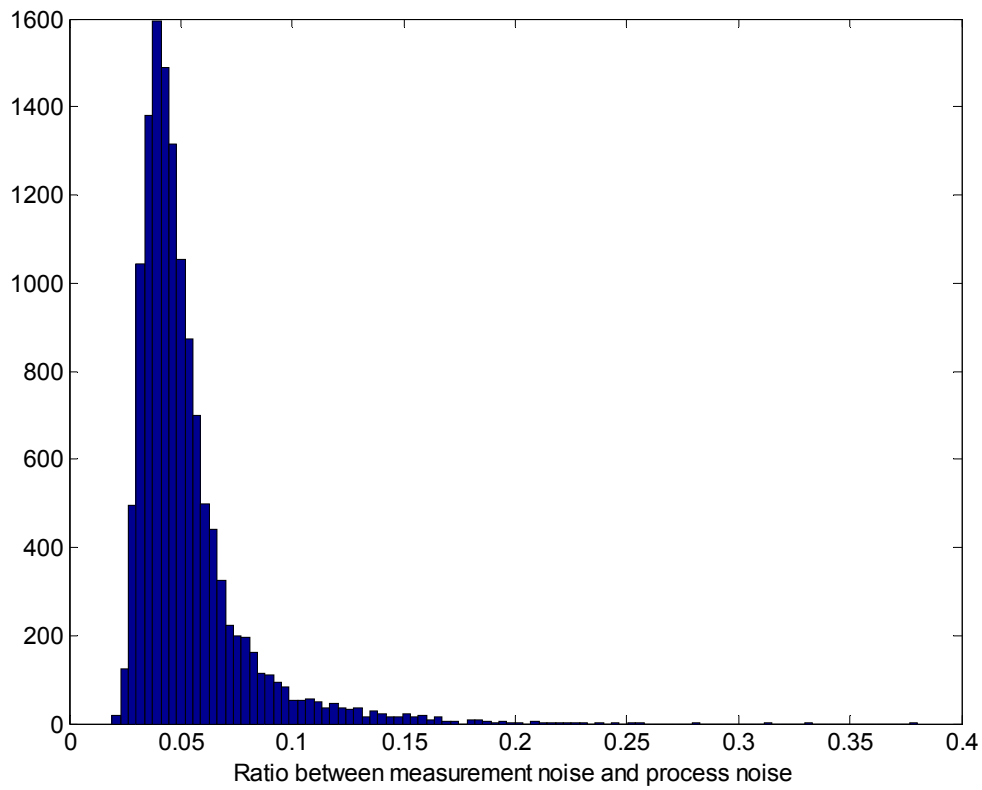

Figure 4.2: Histogram of the ratio $\sigma_{meas}^2/\sigma_{process}^2$ on July 2, 2000

4.2 IONOSPHERIC MODEL INFERENCE

In this section we drop the assumption that the process noise is known. Now we assume that the process noise is the one given by Equation (3.10). The parameter u is now unknown. At each time step it is possible to get an estimate of u; however, there is an uncertainty associated with u, which is dependent on the number of measurements and their quality (the measurement noise.) The idea of this section is largely based on the current WAAS ionospheric estimation algorithm [Walter00], and some proposed changes [Cormier]. New here is the method to evaluate the Probability of Hazardously Misdetection Information and the more realistic treatment of the effect of measurement noise.

In this section we will again take an incremental approach, by first solving an easier problem. The first subsection will show the underlying idea through a simple example. The second one describes how to 'reduce' the measurements to a more usable form. After that, we will solve the problem assuming that there is no measurement noise (M=0). The following subsections introduce the necessary changes when the measurement noise is no longer considered zero.

4.2.1 MOTIVATION

Let us assume that $M=0$, i.e., there is no uncertainty due to imperfect measurements: any deviation from the trend is due to the process noise. The idea is to estimate the value of u and then use an inflated model which is a function of this estimated value. Instead of using the nominal covariance matrix, C_0, we use an inflated model, $u_0^2 C_0$, where u_0 is based on the measurements. The motivation behind this is as follows: Let us suppose that we have a set of m independent zero mean Gaussian random variables, Y_k, with standard deviation $u\sigma_{decorr}$. We have:

$$\frac{1}{m} E\left(\sum_{k=1}^{m} y_k^2 \right) = u^2 \sigma_{decorr}^2 \tag{4.15}$$

As a result, a good estimate of u is therefore the experimental variance of the sample. It is therefore natural to use it in the inflation factor. However, there is an uncertainty associated with this estimate: the experimental variance is itself a random variable which, normalized by $u^2\sigma_{decorr}^2$, is chi-square distributed [Ross]. It is this fact that gives us a handle on the unknown parameter u.

4.2.2 DE-TRENDING AND DECORRELATING THE MEASUREMENTS

In our case, we need to reduce the problem to the previous one. First, we do not measure directly the field $r(x)$, but $r(x)+m(x)$ plus a trend, as shown in Equation (4.4). Second, the residuals are Gaussian but not uncorrelated; their covariance is $C_{\text{tot}}=C+M$. We therefore need to first filter the trend, and decorrelate the resultant residuals. The formation of a chi-square distributed statistic from correlated measurements with a trend is a classical one [Hastie02]. However, as later in the analysis we will need the reduced residuals (de-trended and decorrelated), we go over the derivation. In other words, we want to find Γ such that $y=\Gamma I_{meas}$ is a random Gaussian vector whose covariance is given by the identity matrix. We now show how to obtain the matrix Γ.

First of all we want to filter out the trend. This is equivalent to requiring $\Gamma Ga = 0$ for any 3 by 1 vector a. This implies:

$$\Gamma G = 0 \tag{4.16}$$

Since G is an n by 3 matrix, Γ is at most of rank n-3, which means that we will only be able to extract n-3 orthogonal residuals. So from now on we assume that y is a random Gaussian vector with n-3 components. Condition (4.16) means that if a matrix F has null space G, then there exists a matrix H such that $\Gamma=HF$. A suitable matrix F is:

$$F = I - G\left(G^T W G\right)^{-1} G^T W$$

F is a projection matrix of rank $n-3$ whose null space is the range of G. Now, to find H we set the covariance of $y=\Gamma I_{meas}$ to be the identity, i.e.:

$$E\left(\Gamma I_{meas}\left(\Gamma I_{meas}\right)^T\right) = \Gamma C \Gamma^T = I_{n-3}$$

The equation for H is:

$$H\left(W^{-1} - G\left(G^T W G\right)^{-1} G^T\right) H^T = I_{n-3}$$

We now do the change of variable $\tilde{H} = HW^{-\frac{1}{2}}$. The equation is now:

$$\tilde{H}\left(I - W^{\frac{1}{2}} G\left(G^T W G\right)^{-1} G^T W^{\frac{1}{2}}\right)\tilde{H}^T = I_{n-3} \tag{4.17}$$

The matrix $I - W^{\frac{1}{2}} G\left(G^T W G\right)^{-1} G^T W^{\frac{1}{2}}$ is an orthogonal projection. There exists U orthogonal such that:

$$U^T P U = \begin{bmatrix} I_{n-3} & 0 \\ 0 & 0 \end{bmatrix}$$

If we call $\tilde{U}$ the first n-3 columns of U, then $\tilde{H} = \tilde{U}^T$ fulfills Condition (4.17). We take Γ to be:

$$\Gamma = \tilde{U}^T\left(W^{\frac{1}{2}} - W^{\frac{1}{2}} G\left(G^T W G\right)^{-1} G^T W\right) \tag{4.18}$$

Now, to obtain a chi-squared distributed statistic, we form:

$$y^T y = I_{meas}{}^T \Gamma^T \Gamma I_{meas} = I_{meas}^T\left(W - WG\left(G^T W G\right)^{-1} G^T W\right) I_{meas}$$

This formula coincides with the classical formula [Hastie2]. This formula assumes that W, the inverse of the true covariance, is known, which it is not. There are still advantages to this transformation. Even if Γ is based on a different covariance, y is still zero mean. Also, in the case where there is no measurement noise, the covariance of y is still diagonal. In the

next subsections, the reduced measurements are computed using an arbitrary covariance. We will specify the covariance to use later on.

4.2.3 PHMI FORMULA

Here we develop an expression for the PHMI that will be used to design the algorithm. The PHMI is defined as:

$$\text{PHMI}=\text{Prob}\left(\left|I_{true}-I_{est}\right|>\Delta\right)$$

Because we do not know what the parameter u is at a given time, let us assign a density to it, $p(u)$. The formula for PHMI can then be written:

$$\text{Prob}\left(\text{HMI}\right)=\int_{u=0}^{u=+\infty}\text{Prob}\left(\text{HMI}|u\right)p\left(u\right)du$$

Ideally we would like to have the PHMI below the requirement for any distribution of u. So the new requirement is:

$$\text{Prob}\left(\text{HMI}|u\right)\leq 2.25\ 10^{-8}\ \text{for any } u$$

It is therefore a whole curve that must be under the critical value. Let us now develop the expression for $\text{Prob}\left(\text{HMI}|u\right)$, where we use an inflation factor to multiply the nominal covariance of the process noise. For this we need to introduce some additional notation. The estimation variance using a set of coefficients λ and assuming a process noise u can be written as we did above:

$$\sigma_{est}^{2}\left(\lambda,u\right)=u^{2}\sigma_{process,0}^{2}\left(\lambda\right)+\sigma_{meas}^{2}\left(\lambda\right)$$

where the first term corresponds to the error due to the process noise and the second one corresponds to the measurement noise. The expression for each term comes from Equation (4.8). We have:

$$\sigma^2_{process,0}(\lambda) = \lambda^T C_0(x_k, x_l)\lambda - 2\lambda^T C_0(x, x_k) + C_0(x, x)$$

$$\sigma^2_{meas}(\lambda) = \lambda^T M(x_k, x_l)\lambda$$

Notice that the only assumption on the coefficients here is that the estimate is unbiased. Let us suppose that we assume an inflation factor, u_0, which is a function of the (reduced) measurements y. Then, given u and y, the PHMI is:

$$\mathrm{P}(\mathrm{HMI}|u, y) = 2Q\left(K\frac{\sigma_{est}(\lambda, u_0(y))}{\sigma_{est}(\lambda, u)}\right)$$

To obtain the expression for $\mathrm{P}(\mathrm{HMI}|u)$, we need to integrate over all possible y (after the submission of this work, it was found that this expression is slightly incorrect. However, a correct treatment of the problem has very little impact on the final inflation factor as derived by this equation (see Appendix A.4)):

$$\mathrm{P}(\mathrm{HMI}|u) = \int_y \mathrm{P}(\mathrm{HMI}|u, y)\, p(y \mid u)\, dy$$

We can write the final expression for $\mathrm{P}(\mathrm{HMI}|u)$ by replacing the different terms by their analytical expressions:

$$\mathrm{P}(\mathrm{HMI}|u) = \int_y 2Q\left(K\frac{\sigma_{est}(\lambda, u_0(y))}{\sigma_{est}(\lambda, u)}\right)\frac{1}{(2\pi)^{\frac{n-3}{2}}|S(u)|^{\frac{1}{2}}}e^{-\frac{y^T S(u)^{-1} y}{2}}dy \qquad (4.19)$$

Here *S(u)* is the covariance of the reduced measurements; it is equal to:

$$S(u) = \Gamma\left(u^2 C_0 + M\right)\Gamma^T = u^2\Gamma C_0\Gamma^T + \Gamma M\Gamma^T = u^2 C_0^{(red.)} + M^{(red.)}$$

We can at this point formulate the problem of finding the function $u_0(y)$ and the coefficients λ as follows:

$$\text{Minimize } E\left(\sigma_{est}^2\left(\lambda, u_0(y)\right)\right)$$

$$\text{Subject to } \mathrm{P}\left(\mathrm{HMI}|u\right) \le 2.25\ 10^{-8} \text{ for any } u \tag{4.20}$$

We now have an analytic formulation of the problem. However, it is still a difficult problem. Not only does the constraint hold for a continuum of parameters, but the evaluation of (4.20) involves an n-3-fold integral, where the typical size of n is 30. In addition, the problem of finding the coefficients is not independent of the search of u_0.

4.2.4 CASE WITH NO MEASUREMENT NOISE

We now consider the case where there is no measurement noise: M=0. This case is simpler so it will give us further insight. In addition, the calculations introduced here will be used in the general case. We have $C=u^2C_0$ and $W = u^{-2}C_0^{-1}$. Let us suppose that we compute the matrix Γ based on C_0. Then the reduced residuals y are zero mean and their covariance is $u^2 I_{n-3}$. Expression (4.20) for $\mathrm{P}\left(\mathrm{HMI}|u\right)$ is simpler now:

$$\mathrm{P}\left(\mathrm{HMI}|u\right) = \int_y 2Q\left(K\frac{u_0(y)}{u}\right)\frac{1}{(2\pi)^{\frac{n-3}{2}}u^{\frac{n-3}{2}}}e^{-\frac{y^Ty}{2u^2}}dy \tag{4.21}$$

This expression does not depend on the coefficients λ: the determination of the optimal coefficients λ is independent from the search for u_0. Now, based on the remarks made in the first subsection, we are going to limit the search of u_0 to a positive definite quadratic form of y:

$$u_0^2(y) = y^T R y$$

This is a fundamental assumption for this whole section. We further simplify (4.21) by performing the change of variables $uz = y$:

$$\mathrm{P}\left(\mathrm{HMI}|u\right)=\int_{z}2Q\left(K\sqrt{z^{T}Rz}\right)\frac{1}{\left(2\pi\right)^{\frac{n-3}{2}}}e^{-\frac{z^{T}z}{2}}dz \tag{4.22}$$

This expression does not depend on u. We can now apply Craig's formula for the Q-function [Craig], which is:

$$Q\left(t\right)=\int_{t}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{s^{2}}{2}}ds=\frac{1}{\pi}\int_{\phi=0}^{\frac{\pi}{2}}e^{-\frac{t^{2}}{2\sin\left(\phi\right)^{2}}}d\phi$$

The advantage of this formula, which is extensively used in communication theory, is that the integration limits do not depend on the argument of Q. Replacing the expression in (4.22) we get:

$$\mathrm{P}\left(\mathrm{HMI}|u\right)=2\int_{z}\frac{1}{\pi}\left(\int_{\phi=0}^{\frac{\pi}{2}}e^{-K^{2}\frac{z^{T}Rz}{2\sin\left(\phi\right)^{2}}}d\phi\right)\frac{1}{\left(2\pi\right)^{\frac{n-3}{2}}}e^{-\frac{z^{T}z}{2}}dz$$

If we now reverse the order of integration we get:

$$\mathrm{P}\left(\mathrm{HMI}|u\right)=\frac{2}{\pi}\int_{\phi=0}^{\frac{\pi}{2}}\int_{z}e^{-K^{2}\frac{z^{T}Rz}{2\sin\left(\phi\right)^{2}}}\frac{1}{\left(2\pi\right)^{\frac{n-3}{2}}}e^{-\frac{z^{T}z}{2}}dzd\phi$$

The argument of the one-dimensional integral is:

$$\int_{z}e^{-K^{2}\frac{z^{T}Rz}{2\sin\left(\phi\right)^{2}}}\frac{1}{\left(2\pi\right)^{\frac{n-3}{2}}}e^{-\frac{z^{T}z}{2}}dz=\int_{z}\frac{1}{\left(2\pi\right)^{\frac{n-3}{2}}}e^{-\frac{1}{2}z^{T}\left(K^{2}\frac{R}{\sin^{2}\left(\phi\right)}+I_{n-3}\right)z}dz$$

We can find a closed form for this integral by noticing that:

$$\left|K^{2}\frac{R}{\sin^{2}\left(\phi\right)}+I_{n-3}\right|^{\frac{1}{2}}\int_{z}\frac{1}{\left(2\pi\right)^{\frac{n-3}{2}}}e^{-\frac{1}{2}z^{T}\left(K^{2}\frac{R}{\sin^{2}\left(\phi\right)}+I_{n-3}\right)z}dz=1$$

This is true because we are integrating a Gaussian distribution density over the whole space. The final result is:

$$\mathrm{P}\left(\mathrm{HMI}|u\right) = \frac{2}{\pi}\int_{\phi=0}^{\frac{\pi}{2}} \left| K^2 \frac{R}{\sin^2\left(\phi\right)} + I_{n-3} \right|^{-\frac{1}{2}} d\phi \tag{4.23}$$

This is now a one-dimensional integral, so it is possible to evaluate it. Recall that we want to:

$$\text{Minimize } E\left(z^T R z\right) \text{ subject to (4.23)}$$

Although we will not do it here, it is possible to show that the best R is a multiple of the identity (the proof uses the fact that to minimize a sum of positive real numbers such that the product is constant we need to have all numbers equal). As a result, the best inflation factor in the case of no measurement noise is a multiple of the chi-square statistic:

$$u_0^2\left(y\right) = \alpha y^T y$$

The equation to determine α is given by (4.23) and can be written:

$$2.25*10^{-8} = \frac{2}{\pi}\int_{\phi=0}^{\frac{\pi}{2}} \left(K^2 \frac{\alpha}{\sin^2\left(\phi\right)} + 1 \right)^{-\frac{n-3}{2}} d\phi \tag{4.24}$$

The result only depends on the number of measurements n, so it is possible to compute a table off-line. Let us label each solution α_n. Then, in real time, we only need to compute the chi-square statistic based on the quiet model, C_0. Once we have the inflated model, we can apply the kriging equations. Because the inflated model is a multiple of the nominal model, it is easy to see from Equation (4.13) that the coefficients are invariant under that transformation. For the case without measurement noise the problem is solved.

4.2.5 GENERAL CASE (WITH MEASUREMENT NOISE)

The case with measurement noise is more difficult. We need to consider the PHMI Equation (4.19) and transform the problem so it is treatable. Again, we look for a function $u_0^2(y)$. As we did in the previous case, we are only going to consider quadratic forms. If we do so, we can apply Craig's formula for the Q-function again and reverse the order of integration.

We first want to get rid of the coefficients λ. We do this by noticing that:

$$Q\left(K\frac{\sigma_{est}(\lambda,u_0(y))}{\sigma_{est}(\lambda,u)}\right)\sim Q\left(K\frac{u_0(y)}{u}\right)$$

More precisely, we have:

$$Q\left(K\frac{\sigma_{est}(\lambda,u_0(y))}{\sigma_{est}(\lambda,u)}\right)\leq Q\left(K\frac{u_0(y)}{u}\right)\text{ for } u_0(y)\leq u$$

and for $u_0(y)\geq u$ we have:

$$Q\left(K\frac{\sigma_{est}(\lambda,u_0(y))}{\sigma_{est}(\lambda,u)}\right)\geq Q\left(K\frac{u_0(y)}{u}\right)$$

which is a problem. However, in that case we also have:

$$Q\left(K\frac{\sigma_{est}(\lambda,u_0(y))}{\sigma_{est}(\lambda,u)}\right)\leq 2.5\ 10^{-8}$$

This indicates that for $u_0(y)\geq u$ the contribution of the integral (4.19) will not be a large one. We could take into account this approximation by simply inflating *K* from 5.592 to 5.69 (2%). The proof is straightforward but tedious. However, this approximation has

consistently proven to be a conservative one, due to the fact that $\sigma^2_{meas}(\lambda) \ll \sigma^2_{process,0}(\lambda)$ - as we saw in Subsection 4.1.5. After this approximation, we have:

$$\mathrm{P}(\mathrm{HMI}|u) = \int_y 2Q\left(K\frac{u_0(y)}{u}\right)\frac{1}{(2\pi)^{\frac{n-3}{2}}\left|S(u)\right|^{\frac{1}{2}}}e^{-\frac{y^T S(u)^{-1} y}{2}}dy \tag{4.25}$$

Expression (4.25) is a big step because we have decorrelated the search for the inflation factor from the search for the coefficients.

By writing that $u_0^2(y) = y^T R y$ and applying Craig's formula in the same way we did in the previous subsection, we obtain:

$$\mathrm{P}(\mathrm{HMI}|u) = \frac{2}{\pi}\int_{\phi=0}^{\frac{\pi}{2}}\left|K^2\frac{RS(u)}{u^2\sin^2(\phi)} + I_{n-3}\right|^{-\frac{1}{2}}d\phi \tag{4.26}$$

where, as we saw earlier, $S(u) = u^2\Gamma C_0\Gamma^T + \Gamma M\Gamma^T = u^2 C_0^{(red.)} + M^{(red.)}$. As we did for the noiseless case, we have reduced an *n*-3-fold integral to a one-dimensional integral. Now we can evaluate this function of *u* for a given *R*, and make sure that it is always below the critical PHMI value. To gain some insight on the behavior of $\mathrm{P}(\mathrm{HMI}|u)$ as expressed in (4.26), we now study an example.

4.2.6 EXAMPLE

Let us suppose that $C_0^{(red.)} = \beta I_{n-3}$ and $M^{(red.)} = \eta I_{n-3}$. The factor β reflects the relative magnitude of the process noise versus the measurement noise in the nominal covariance model. This example can be treated easily (there is no need to compute determinants) and it is important since we will be able to reduce the general case to it. Because of the definition of these matrices, we have $\beta+\eta=1$. Because it is optimal in the case without measurement noise, it is reasonable to take $R = \alpha I_{n-3}$. This means that the inflation factor

is a multiple of the chi-square statistic computed assuming the nominal covariance model. Equation (4.26) is now:

$$\mathrm{P}\left(\mathrm{HMI}|u\right)=\frac{2}{\pi}\int_{\phi=0}^{\frac{\pi}{2}}\left|K^{2}\frac{\alpha\left(u^{2}\beta I_{n-3}+\left(1-\beta\right)I_{n-3}\right)}{u^{2}\sin^{2}\left(\phi\right)}+I_{n-3}\right|^{-\frac{1}{2}}d\phi$$

Since the matrix in the integral is diagonal, this equation is simplified:

$$\mathrm{P}\left(\mathrm{HMI}|u\right)=\frac{2}{\pi}\int_{\phi=0}^{\frac{\pi}{2}}\left(K^{2}\frac{\alpha\left(u^{2}\beta+\left(1-\beta\right)\right)}{u^{2}\sin^{2}\left(\phi\right)}+1\right)^{-\frac{n-3}{2}}d\phi \tag{4.27}$$

The difference now is that there is no determinant in the term to be integrated, only a product. We can now study the behavior of $\mathrm{P}\left(\mathrm{HMI}|u\right)$ as a function of u, parameterized on the relative contribution of the process noise, β. Figure 4.3 shows several of these curves for different values of β in the case n=20. The parameter α was computed using Equation (4.24).

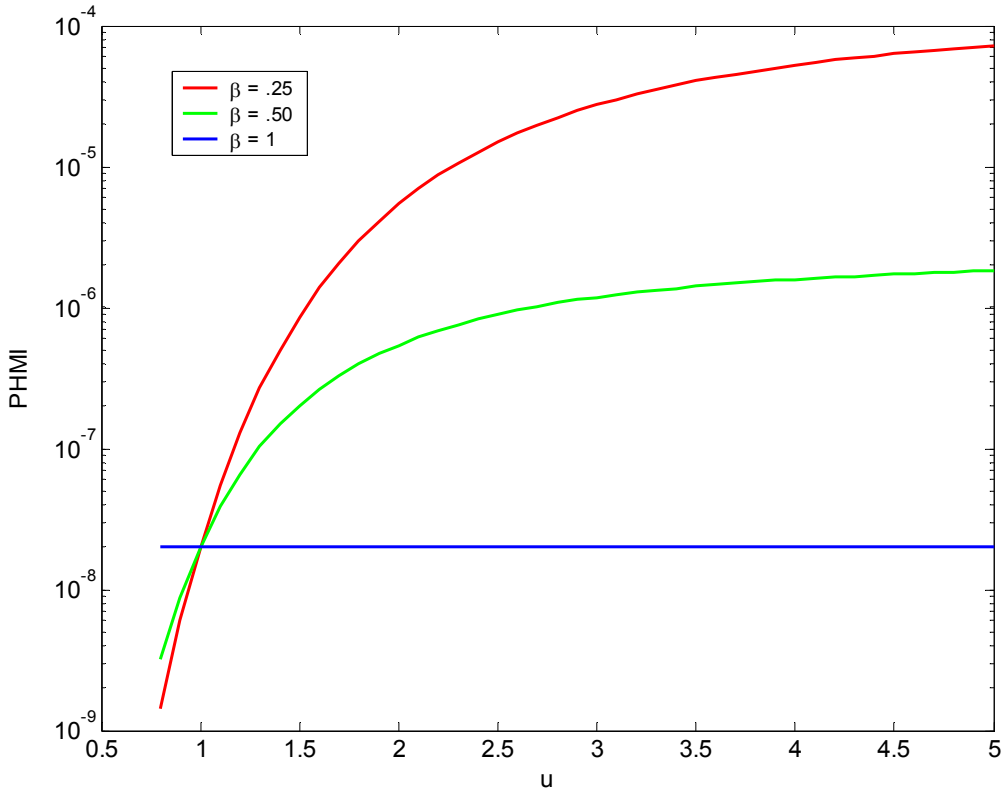

Figure 4.3: PHMI as a function of u parameterized by β

The blue curve corresponds to the case of pure process noise. As expected, it does not depend on the parameter u. As we said, P(HMI|u) increases with u. It increases faster as the measurement noise becomes larger compared to the nominal process noise.

This function has two important properties. It increases with u and has a limit as u goes to infinity. It is straightforward to compute this limit:

$$\mathrm{P}\left(\mathrm{HMI}|+\infty\right)=\frac{2}{\pi}\int_{\phi=0}^{\frac{\pi}{2}}\left(K^{2}\frac{\alpha\beta}{\sin^{2}\left(\phi\right)}+1\right)^{-\frac{n-3}{2}}d\phi \tag{4.28}$$

This formula is identical to (4.24) except for the factor β. It suggests that to get a safe inflation factor, we could compute the table as we did in the case with no measurement noise, and then scale the result by the inverse of β.

The behavior of P(HMI|u) is due to the fact that when we compute the chi-square statistic, we assume a certain ratio between process noise and measurement noise. However, if the ratio is in reality larger than assumed, the inflation factor is going to act as if some of the variability that is really due to the process noise is due to the measurement noise, thus underestimating the process noise. That would not be a problem if the ratio of the contribution of process noise to the measurement noise in the estimation variance was the same as in the ratio between the covariances. But this is not the case. As we saw in Section 4.1.5, the term coming from the measurement noise is much smaller, because the large number of measurements allows us to average it down. What this section shows is that the ratio between process noise and measurement noise is critical. We have also shown that when the reduced matrices are proportional to the identity, it is possible to compute an inflation factor safe for all u.

4.2.7 RATIO BETWEEN PROCESS NOISE AND MEASUREMENT NOISE

In this section we are going to show that by using the general case, we can find an upper bound of the PHMI where the matrices are proportional to the identity (like in the previous section). Also, we are going to evaluate the ratio between process noise and measurement noise in the real system.

Because it is optimal in the noiseless case, we take again $R = \alpha I_{n-3}$. The PHMI condition is:

$$\mathrm{P}\left(\mathrm{HMI}|u\right) = \frac{2}{\pi}\int_{\phi=0}^{\frac{\pi}{2}} \left| K^2\alpha \frac{u^2 C_0^{(red.)} + M^{(red.)}}{u^2 \sin^2\left(\phi\right)} + I_{n-3} \right|^{-\frac{1}{2}} d\phi$$

We remind the reader that we have $C_0^{(red.)} + M^{(red.)} = I_{n-3}$. There is no reason for each term in the sum to be a multiple of the identity. However, we can suppose without loss of generality that both matrices are diagonal (to see this, diagonalize one of them; the other one needs to be diagonal in the same basis). Let us write:

$$C_0^{(red.)} = diag\left(\begin{bmatrix} c_1 & \cdots & c_{n-3} \end{bmatrix}\right) \text{ and } M^{(red.)} = diag\left(\begin{bmatrix} m_1 & \cdots & m_{n-3} \end{bmatrix}\right)$$

By construction, $c_i + m_i = 1$ and $c_i, m_i \geq 0$. We have:

$$\mathrm{P}\left(\mathrm{HMI}|u\right) = \frac{2}{\pi} \int_{\phi=0}^{\frac{\pi}{2}} \prod_{i=1}^{n-3} \left(K^2 \alpha \frac{u^2 c_i + m_i}{u^2 \sin^2\left(\phi\right)} + 1 \right)^{-\frac{1}{2}} d\phi$$

It is possible to show that:

$$\mathrm{P}\left(\mathrm{HMI}|u\right) \leq \frac{2}{\pi} \int_{\phi=0}^{\frac{\pi}{2}} \left(K^2 \alpha \frac{u^2 \left(\prod_{i=1}^{n-3} c_i\right)^{\frac{1}{n-3}} + \left(\prod_{i=1}^{n-3} m_i\right)^{\frac{1}{n-3}}}{u^2 \sin^2\left(\phi\right)} + 1 \right)^{-\frac{n-3}{2}} d\phi \tag{4.29}$$

This inequality is based on the following general inequality (see Appendix A for proof):

$$\left(1 + \left(\prod_{k=1}^{p} b_k\right)^{\frac{1}{p}}\right)^p \leq \prod_{k=1}^{p} \left(1 + b_k\right) \text{ for } b_k \geq 0 \tag{4.30}$$

The upper bound in (4.29) of the PHMI is similar to (4.27) with $\beta = \left(\prod_{i=1}^{n-3} c_i\right)^{\frac{1}{n-3}} = \left|C_0^{(red)}\right|^{\frac{1}{n-3}}$.

When the matrices are a multiple of the identity matrix, there is equality. The closer these matrices are to a multiple of the identity, the tighter the bound is. Let us now examine the empirical distribution of the parameter β in the system. It will be a function of the covariance of the measurement noise and the geometry of the measurements (we assume the nominal model for the process noise covariance). Figure 4.4 shows the histogram of the computed values of β for July 2, 2000.

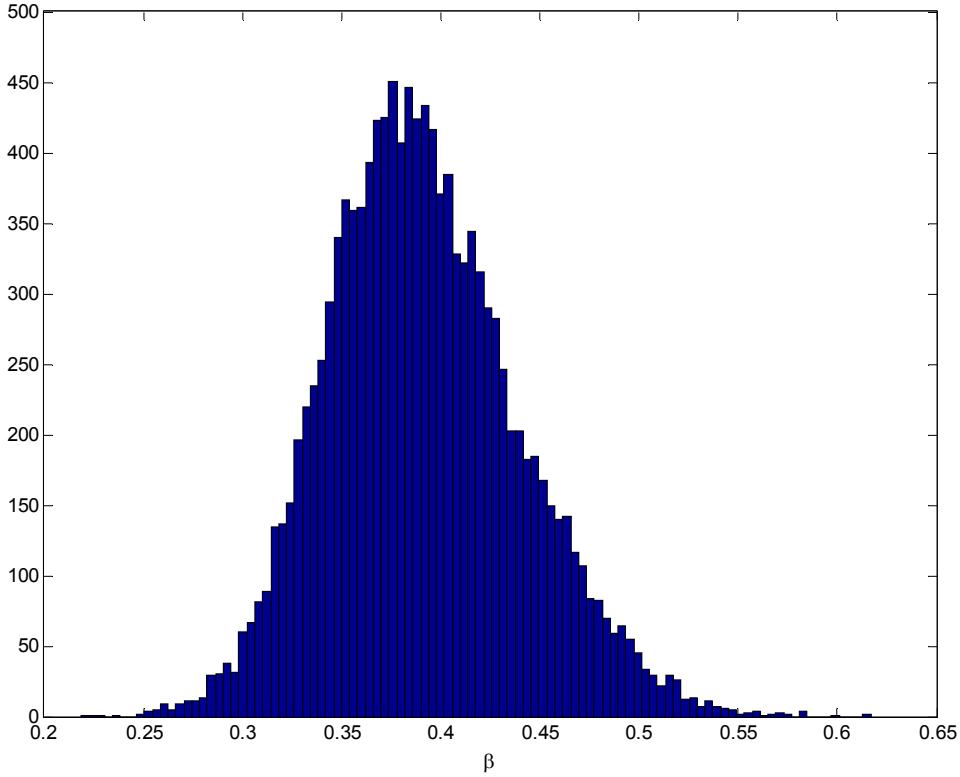

Figure 4.4: Process noise contribution

At this point, we could choose to pick a fixed value for β or to compute it in real time. A fixed value would need to be a lower bound of all possible values of β. In the final algorithm we will compute β in real time. This way we only multiply by what is needed in a particular situation. The computation of the parameter β is not as expensive as this proof might suggest. We have:

$$\beta^{n-3} = \left|C_0^{(red)}\right| = \left|\Gamma C \Gamma^T\right| = \prod_{\ell_i \neq 0} \ell_i \left(C\Gamma^T\Gamma\right) = \prod_{\ell_i \neq 0} \ell_i \left(C\left(W - WG^T\left(GWG^T\right)GW\right)\right)$$

In this expression, ℓ designates the eigenvalues of a matrix. The computation of the last expression only requires finding the eigenvalues of $C\left(W - WG^T\left(GWG^T\right)GW\right)$. In this expression, C is taken to be the nominal covariance, C_0, and W is computed based on C_0.

4.2.8 ADDING A THRESHOLD

In the previous sections we have taken the inflation factor to be a quadratic form of the measurement, and showed that we could compute an inflation factor that satisfies the PHMI. The quadratic form was chosen for two reasons. It is a good choice in the measurement noise free case and it allows us to obtain a relatively simple analytical form of the PHMI. However, the curves in Figure 4.3 suggest that we would like the inflation factor to increase faster than a quadratic as the residuals increase. It is an open problem to find a functional form for $u_0^2(y)$ that is close to quadratic but increases faster, and that makes the integral (4.22) easily computable. Another solution consists of adding a threshold to the chi-square statistic. If the chi-square statistic is larger than the threshold, the error bound is set to a maximum value (the maximum ever observed) [ADD], [Walter00]. The idea behind this is as follows. The PHMI becomes larger than desired for large values of u. These values of u are likely to result in a large chi-square statistic. If we only accept chi-square statistics below a certain threshold, we will be less likely to incur in HMI. In this section we study the effect of adding a threshold on the PHMI. The analysis done in this section relies on two unproven approximations. Despite the lack of proof, these two approximations are intuitively correct.

The new PHMI formula is similar to (4.26). Let us call T the threshold:

$$\mathrm{P}(\mathrm{HMI}|u) = \int_{y^T y \le T} 2Q\left(K\frac{u_0(y)}{u}\right)\frac{1}{(2\pi)^{\frac{n-3}{2}}\left|S(u)\right|^{\frac{1}{2}}}e^{-\frac{y^T S(u)^{-1} y}{2}}dy$$

The only difference lies in the integration domain. Now we only consider y inside the ellipsoid $y^T y = T$. This change again makes the computation of the PHMI difficult and unpractical (because there is no formula to integrate the density of a multivariate Gaussian distribution over an arbitrary ellipsoid). Instead, we are going to rely on the following approximation:

$$\int_{y^T y \le T} 2Q\left(K\frac{u_0(y)}{u}\right)p(y\,|\,u)\,dy = \left[\int_{y} 2Q\left(K\frac{u_0(y)}{u}\right)p(y\,|\,u)\,dy\right]\left[\int_{y^T y \le T} p(y\,|\,u)\,dy\right]$$

This approximation is obtained by assuming that:

$$E\left(Q\left(K\frac{u_0(y)}{u}\right)1\left(y^T y \le T\right)\right) = E\left(Q\left(K\frac{u_0(y)}{u}\right)\right)E\left(1\left(y^T y \le T\right)\right)$$

where $1\left(y^T y \le T\right)$ is a function equal to one over the ellipsoid and zero otherwise. Now we can use the results obtained in the previous section. We know how to find a computable upper bound of the first term (Subsection 4.2.7), so we turn our attention to the other term:

$$\text{Prob}\left(y^T y \le T\right) = \int_{y^T y \le T} \frac{1}{(2\pi)^{\frac{n-3}{2}}\left|S(u)\right|^{\frac{1}{2}}} e^{-\frac{y^T S(u)^{-1} y}{2}}\,dy$$

There is no known analytical solution for this integral. Even worse, the numerical integration for such dimensions is not trivial. We will first show the link from this integral to a weighted sum of chi-squares, which is a problem which has been studied [Gabler]. Let us do the change of variables $z = S(u)^{-\frac{1}{2}} y$. We have:

$$\text{Prob}\left(y^T y \le T\right) = \int_{z^T S(u) z \le T} \frac{1}{(2\pi)^{\frac{n-3}{2}}} e^{-\frac{z^T z}{2}}\,dz \tag{4.31}$$

As we said earlier, *S(u)* can be considered to be diagonal without loss of generality: $S(u) = u^2 diag\left(\left[c_1 \quad \cdots \quad c_{n-3}\right]\right) + diag\left(\left[m_1 \quad \cdots \quad m_{n-3}\right]\right)$, so we have:

$$z^T S(u) z = \sum_{i=1}^{n-3} z_i^2\left(u^2 c_i + m_i\right) \tag{4.32}$$

This shows that (4.31) is the cumulative distribution function of a sum of weighted chi-squares. In [Gabler], the following approximation is given:

$$\text{Prob}\left(\sum_{i=1}^{n-3} z_i^2\left(u^2 c_i + m_i\right) \le T\right) \approx \text{Prob}\left(\sum_{i=1}^{n-3} z_i^2 \le T\left(\prod_{i=1}^{n-3} u^2 c_i + m_i\right)^{-\frac{1}{n-3}}\right)$$

The right term is a chi-square cumulative distribution, so it is easily computed. Now take one step further by applying the inequality that was used to prove (4.31). We get:

$$\text{Prob}\left(\sum_{i=1}^{n-3} z_i^2 \le T\left(\prod_{i=1}^{n-3} u^2 c_i + m_i\right)^{-\frac{1}{n-3}}\right) \le \text{Prob}\left(\sum_{i=1}^{n-3} z_i^2 \le \frac{T}{u^2\left(\prod_{i=1}^{n-3} c_i\right)^{\frac{1}{n-3}} + \left(\prod_{i=1}^{n-3} m_i\right)^{\frac{1}{n-3}}}\right)$$

In order to get an idea of the benefit the chi-square yields, we apply these results to the example studied in subsection 4.2.6: $c_i = \beta$, $m_i = 1-\beta$. The threshold was chosen such that:

$$\text{Prob}\left(\sum_{i=1}^{n-3} z_i^2 \le T\right) \le 10^{-3}$$

This is the probability of false alarm if we are in the nominal state, but here it is not a specification of the algorithm. Figure 4.5 shows the new PHMI curves generated by adding the threshold to the algorithm. We can see that now the PHMI is no longer an increasing function of *u*, and that it has a maximum value. This maximum value needs to be below the design requirement ($2.25\text{X}10^{-8}$). We see that despite the threshold, the PHMI is unacceptably large and that there is little benefit compared to the situation without the threshold.

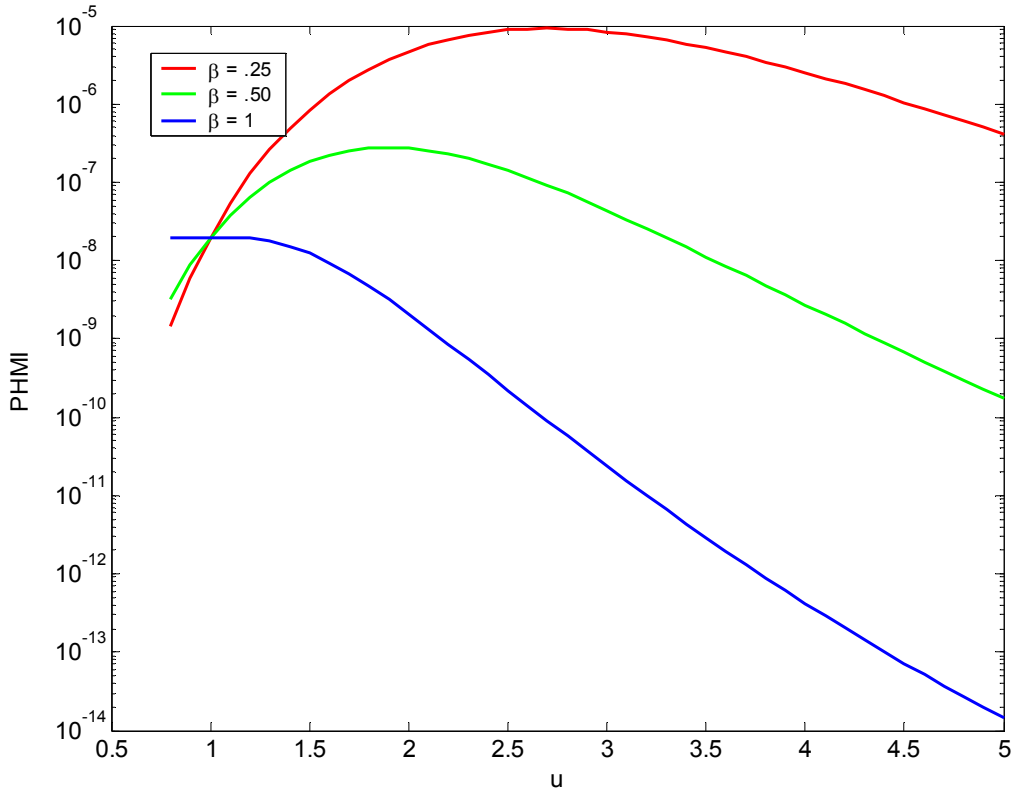

Figure 4.5: PHMI as a function of u when a threshold is added

There are two ways of bringing down the curves: by decreasing the threshold T or by increasing the inflation factor α. Decreasing the threshold is not an attractive option, because when the ionosphere is well behaved (i.e. when the nominal covariance is an upper bound of the true covariance) we want the probability of false alarm to be very small. As a result, we will have to rely only on the inflation factor. None of the analysis done in this subsection will be used to generate the inflation factor because it relies on some unproven approximations and the benefit seems to be small. The conclusion of this subsection is that, within our model, adding a threshold does not help much. However, the implementation of the threshold will be important when we deal with the irregularities that are not well described by the assumed model, as we will see later.

4.2.9 DETERMINATION OF THE INFLATION FACTOR

The purpose of this section was to determine an algorithm that computes an inflation factor u_0 such that the model $u_0^2 C_0$ can be used as if there was no uncertainty. This subsection summarizes the different steps in the computation, and distinguishes what can be done offline from what must be done in real time.

The inflation factor is given by $u_0^2(y) = \alpha y^T y$, where we have:

$$y^T y = I_{meas}^T \left(W_0 - W_0 G \left(G^T W_0 G \right)^{-1} G^T W_0 \right) I_{meas}$$

where $W_0 = \left(C_0 + M \right)^{-1}$. We still need to determine α.

Subsection 4.2.7 showed that:

$$\mathrm{P}\left(\mathrm{HMI}|u\right) \leq \frac{2}{\pi} \int_{\phi=0}^{\frac{\pi}{2}} \left(K^2 \frac{\alpha\beta}{\sin^2(\phi)} + 1 \right)^{-\frac{n-3}{2}} d\phi \text{ where } \beta = \left(\prod_{i=1}^{n-3} c_i \right)^{\frac{1}{n-3}} = \left| C_0^{(red)} \right|^{\frac{1}{n-3}} \quad (4.33)$$

The parameter β is given by:

$$\beta = \left[\prod_{\lambda_i \neq 0} \ell_i \left(C_0 \left(W_0 - W_0 G^T \left(G W_0 G^T \right)^{-1} G W_0 \right) \right) \right]^{\frac{1}{n-3}} \quad (4.34)$$

In Subsection 4.2.4, we determined for a given n a value α_n such that:

$$2.25 * 10^{-8} = \frac{2}{\pi} \int_{\phi=0}^{\frac{\pi}{2}} \left(K^2 \frac{\alpha_n}{\sin^2(\phi)} + 1 \right)^{-\frac{n-3}{2}} d\phi$$

According to Equation (4.33), if we choose α such that: $\alpha\beta = \alpha_n$, we will have:

$$\mathrm{P}\left(\mathrm{HMI}|u\right) \le \frac{2}{\pi}\int_{\phi=0}^{\frac{\pi}{2}}\left(K^2\frac{\alpha\beta}{\sin^2\left(\phi\right)}+1\right)^{-\frac{n-3}{2}}d\phi = \frac{2}{\pi}\int_{\phi=0}^{\frac{\pi}{2}}\left(K^2\frac{\alpha_n}{\sin^2\left(\phi\right)}+1\right)^{-\frac{n-3}{2}} = 2.25 * 10^{-8}$$

Once we have the inflation factor, we proceed as if the covariance model was the inflated model and not worry anymore about the model uncertainty. To summarize, for a given set of measurements we have:

$$u_0^2 = \frac{\alpha_n}{\beta} I_{meas}^T \left(W_0 - W_0 G\left(G^T W_0 G\right)^{-1} G^T W_0\right) I_{meas}$$

where β is given by Equation (4.33) and α_n is read in a table. Figure 4.6 shows the flow chart of the real time estimation of the inflation factor. The different variables were defined at the beginning of the chapter.

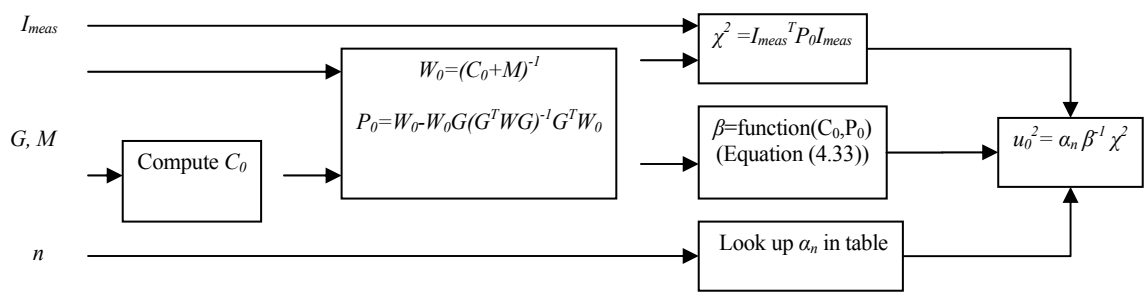

Figure 4.6: Real time inflation factor determination

Once we have this inflation factor, we use the process noise whose covariance is $C=u_0^2 C_0$. Using this covariance, we can now apply the kriging method outlined in the first section to compute the coefficients and the estimation variance.

4.3 THE THREAT MODEL

Section (4.1) described how to compute the kriging estimate assuming a certain model (known u). Section (4.2) showed how to take into account the model uncertainty by inflating the nominal model by u_0. The two previous sections have thus determined the

ionospheric estimation algorithm as if the model (parameterized by u) derived in Chapter 3 always held. However, as we saw at the end of Chapter 3, there is evidence that the model might not always hold: large gradients sometimes divide quiet regions (low u) from disturbed regions (large u). The biggest threat to the system can happen when the measurements are all made in the quiet region, so the system is blind to the coming gradient. As this is an extremely difficult situation to model in probabilistic terms, the method followed here (and used in the current ionospheric estimation algorithm for WAAS) relies on the post-processed data and data deprivation schemes of real data. An extensive description of this algorithm is presented in [Sparks01].

The net effect of this part of the algorithm, called *threat model*, is twofold. The first part to adds an additional term to the error bound that is a function of the geometry of the measurements surrounding the region that is estimated. The second part conditions the computation of an error bound to have the chi-square statistic be below a threshold, as it was outlined earlier. The reason for implementing the chi-square threshold is the fact that violations of stationarity occur much more often during disturbed periods. The idea is to test a given algorithm separating the real data in data used for estimation from the test data (which simulates a user). Then, it is checked whether the error bound computed with the estimation algorithm gives an upper bound of the estimation error. If the error is larger than the estimated error bound, we assume that for a situation with a similar geometry as the one tested, we will need to increase its error bound by as much as the real error outgrew the estimated error bound. In the current WAAS algorithm, the geometry is characterized by two parameters: the number of measurements used for the correction and the distance from the center of gravity of the measurements to the location to be estimated [Sparks01].

In this thesis, we will not give a full account of the threat model because it is still an evolving part of the algorithm and because it can be considered an external part of the algorithm. Furthermore, new ways of computing the threat model show that it is possible to minimize the effect of the threat model [Sparks03]. In Chapter 6, performance will be presented with and without the threat model.

Chapter 5

Adaptation of the Algorithm to the WAAS Message

Now we have a way of computing at any CONUS location of the thin shell an estimate of the vertical ionospheric delay and an error bound. However, this algorithm assumes that the user has all the measurement information. Because of bandwidth limitations and the need to reduce as much as possible the computational load on the user, the ionospheric correction message is in a format described in the WAAS Minimum Operational Standard (MOPS) [MOPS]. This document specifies that the ionospheric information be sent in a grid of 5 by 5 degrees or 10 by 10 degrees in the thin shell at a height of 350 km. At each ionospheric grid point (IGP), the user receives a vertical ionospheric grid delay (IGD) and a grid ionospheric vertical error (GIVE). The user calculates each of the ionospheric delays corresponding to the satellites in sight as well as the confidence bounds from this grid, according to an algorithm which is also set in the WAAS MOPS. Figure 5.1 shows the set of grid points as circles. To be applicable in the short term, an estimation algorithm must be adapted to the ionospheric grid.

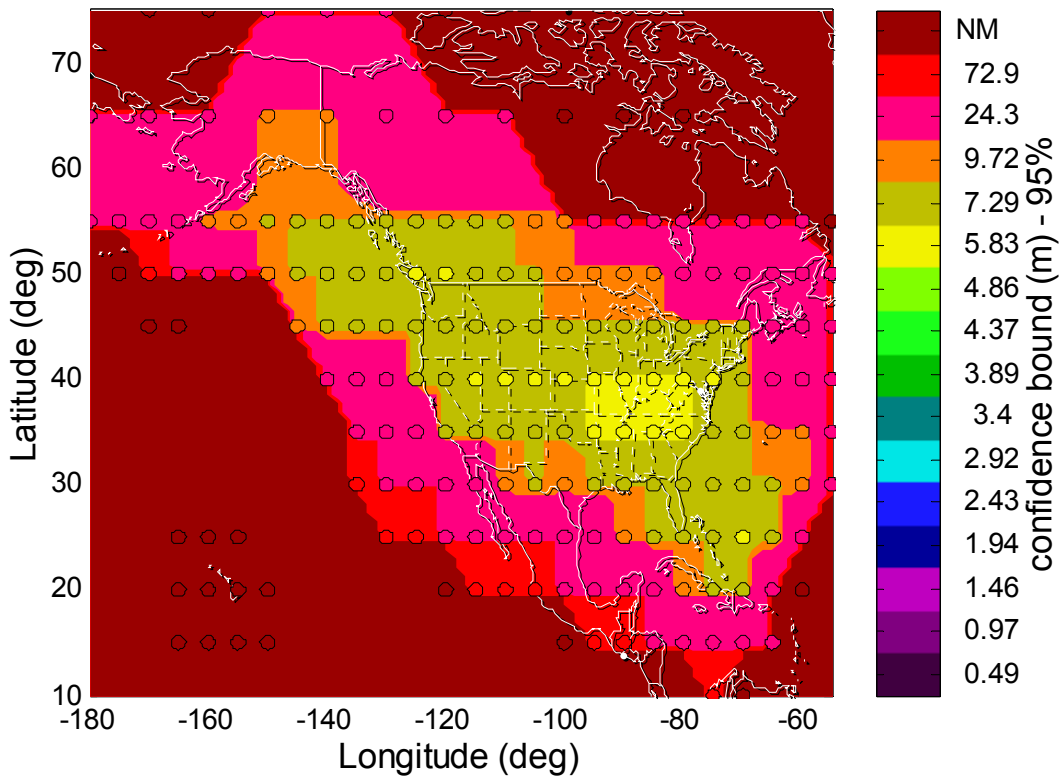

Figure 5.1: Ionospheric grid points as specified in the MOPS

The algorithm as it has been outlined previously cannot be applied directly. Figure 5.2 illustrates the problem in a two-dimensional example. In this diagram, the blue dots represent the IPP measurement location on the horizontal and the value of the vertical delay on the vertical. The estimate computed using kriging at all locations would give a curve that loosely follows the measurement (the exact shape is determined by the variogram). With the estimate, there is also a confidence bound, shown in red. As we said, the user only receives the vertical ionospheric delay at the IGPs, i.e., at the two extremes of the diagram. The estimate computed by the user (black line) does not coincide with the kriging estimate. As a result, the error bound sent to the user at the grid points needs to be such that the interpolation error is accounted for. Fortunately, the situation is not as bad as it is shown in this diagram, because the grid spacing is not as large compared to the ionospheric decorrelation, but it still needs to be evaluated. The first decision to make is what vertical

delay to send at the IGP. The most natural choice is to send the value obtained using kriging at that location, and that is what we will do.

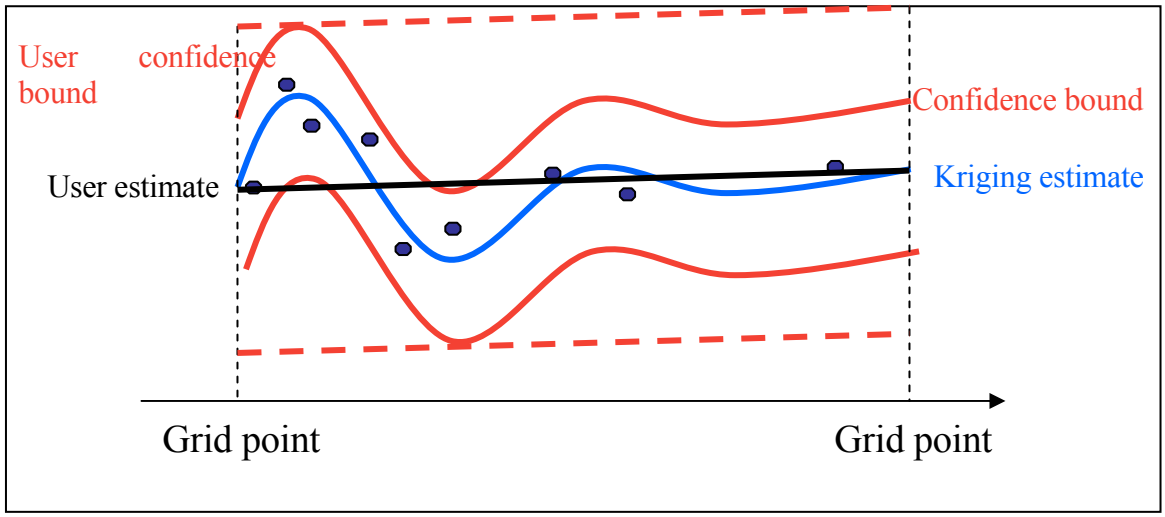

Figure 5.2: Difference between kriging estimate and user estimate

To treat this problem, we will first review how the user forms both the estimate and the error bound (also called confidence bound). Then we will compute an overbound of the true user error bound. Finally we will see how to compute the GIVE such that the user always computes an overbound of the true error bound.

5.1 THE IONOSPHERIC GRID: USER ALGORITHM

For each of its IPPs, the user determines the box of IGPs in which the IPP is contained. Then the user interpolates the data of the four IGPs which form the box:

$$\hat{I}_{user} = \sum_{i=1}^{4} \mu_i \hat{I}_{IGP,i} \tag{5.1}$$

Similarly, the user confidence bound, $K\sigma_{UIVE}$, is found by interpolating the four surrounding σ_{GIVE}:

$$\sigma_{UIVE}^2 = \sum_{i=1}^{4} \mu_i \sigma_{GIVE}^2$$

The μ coefficients are computed by the receiver according to a bilinear interpolation scheme specified in [MOPS]. This scheme has an essential property that will be used later in the section. It is unbiased in the sense that:

$$G_{IGP}\mu = X$$

The matrix G_{IGP} is the geometric matrix corresponding to the IGPs (it is defined as G is for the measurements). X was defined earlier.

5.2 USER ERROR BOUND

The user error bound depends on how the estimate is formed at the IGP by the master station. Since the kriging estimate is optimal, given the assumed covariance, the WAAS message will broadcast the delays at the IGP computed using kriging (as described in Section 5.1.) In this subsection, we will only assume that each IGD estimate is unbiased (in the sense that the expectation of the estimate over all possible residual fields is the same as the expectation of the true vertical ionospheric delay). We have:

$$\hat{I}_{IGP,i} = \lambda_i^T I_{meas}$$

Here λ_i is the vector of coefficients and I_{meas} is the vector of measurements. Since the estimate formed from the IGPs to the user is also unbiased, we have:

$$E\left(\hat{I}_{user}\right) = E\left(\sum_{i=1}^{4} \mu_i \hat{I}_{IGP,i}\right) = E\left(I_{true}\right)$$

We can therefore calculate the estimation variance. (The reader is reminded that, because of the inflation factor, we can treat the problem as if there was no model uncertainty.) We have:

$$E\left(\left(\hat{I}_{user}-I_{true}\right)^2\right)=C\left(x,x\right)-2\sum_{i=1}^{4}\mu_i\lambda_i^T C\left(x,x_k\right)+\left(\sum_{i=1}^{4}\mu_i\lambda_i\right)^T\left(C\left(x_l,x_k\right)+M\left(x_l,x_k\right)\right)\left(\sum_{i=1}^{4}\mu_i\lambda_i\right)$$

This is the expression for the true estimation variance at the user IPP when interpolating the four surrounding IGPs (although it would be true for any number of IGPs).

5.3 GIVE COMPUTATION

In this section we present the method to compute the GIVE at each IGP. We will assume that at each grid point, the vertical ionospheric delay broadcast is the kriging estimate obtained by the algorithm outlined in Chapter 4. After showing the principle of the calculation, we will give the details.

5.3.1 PRINCIPLE

The idea of the GIVE computation is to consider each quadrant, that is, each group of four IGPs, and to find four quantities, σ_i (one for each IGP), such that for any user IPP located in the quadrant:

$$E\left(\hat{I}_{user}-I_{true}\right)^2 \le \sum_{i=1}^{4}\mu_i\sigma_i^2 \tag{5.2}$$

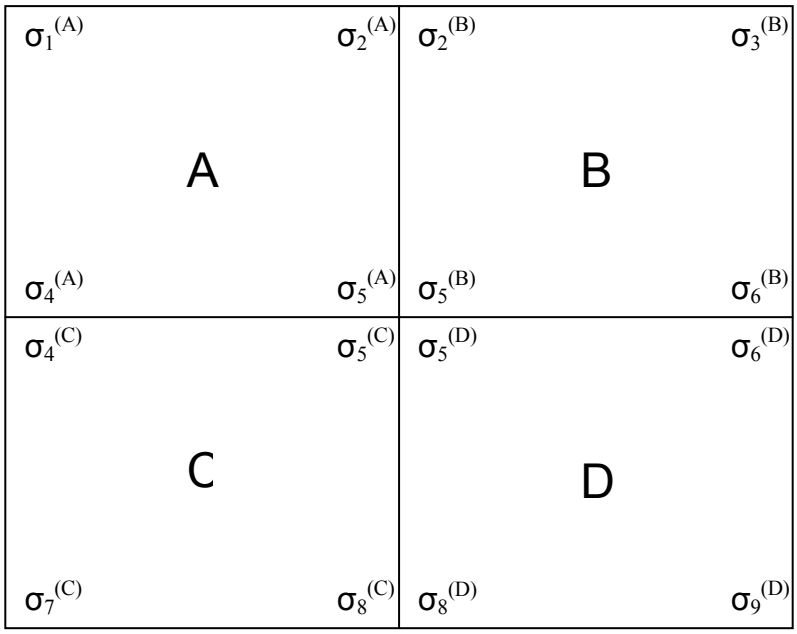

Figure 5.3: GIVE computation

For each quadrant, we will assign a σ_i to each of the four IGPs. Since each IGP belongs to four quadrants, we will have 4 different σ_i for each IGP. σ_{GIVE}^2 will be the maximum of these four quantities. Figure 5.3 illustrates the situation. This diagram shows nine IGPs, and for each quadrant, the four σ_i corresponding to Equation (5.2). Here, we want to compute σ_{GIVE}^2 at the IGP in the center (Number 5). First, for each quadrant, A, B, C and D, we compute the four σ_i. Because we want the broadcast value at IGP number 5 to be safe for any user having an IPP in A, B, C or D, we will need to have:

$$\sigma_{GIVE} = \max\left(\sigma_5^{(A)}, \sigma_5^{(B)}, \sigma_5^{(C)}, \sigma_5^{(D)}\right)$$

5.3.2 DECOMPOSING THE ESTIMATION VARIANCE IN THREE TERMS

We now focus on getting the four σ_i for each quadrant. In order to find these four terms, we write:

$$E\left(\hat{I}_{user} - I_{true}\right)^2 = C(x,x) + \mu^T S \mu + 2\sum_{i=1}^{4} \mu_i \lambda_i^T \left(\sum_{i=1}^{4} \mu_i C(x_i, x_k) - C(x, x_k)\right) \qquad (5.3)$$

The expression for S is:

$$S = \Lambda^T \left(C\left(x_k, x_l\right) + M\left(x_k, x_l\right) \right) \Lambda - \Lambda^T C\left(x_i, x_l\right) - C\left(x_i, x_l\right)^T \Lambda$$

The matrix Λ is formed of the coefficients used to estimate each delay at the IGP:

$$\Lambda = \begin{bmatrix} \lambda_1 & \cdots & \lambda_4 \end{bmatrix}$$

If the expression given by (5.3) was affine in μ, the problem would be solved. But it is not: the second term is quadratic, and the third has an even more complicated dependency on μ.

The problem now is to find an upper bound linear in μ for each one of the three terms. For the first term, we have:

$$C\left(x, x\right) = \sum_{i=1}^{4} \mu_i C\left(x_i, x_i\right)$$

This is the case because we know that the components of the vector μ have a sum of one and *C(x,x)* does not depend on the location. The other two terms require more work.

5.3.3 BOUNDING THE QUADRATIC TERM

The second term is quadratic in μ. Let us assume for now that we can write $S = S^{(+)} + S^{(-)}$, where the first term is a positive semi-definite matrix and the second one is a negative semi-definite matrix. Because of the convexity of positive semi-definite quadratic forms it is easy to see that:

$$\mu^T S^{(+)} \mu \le \sum_{i=1}^{4} \mu_i S_{ii}^{(+)}$$

For the definite negative term, we use the fact that a tangent plane of the graph is an upper bound. For any μ_0 we have:

$$\left(\mu - \mu_0\right)^T S^{(-)} \left(\mu - \mu_0\right) \le 0$$

After developing the equation we have:

$$\mu^T S^{(-)} \mu \le 2\mu^T S^{(-)} \mu_0 - \mu_0{}^T S^{(-)} \mu_0 = b^T \mu$$

The choice of the point μ_0 is arbitrary for each quadrant. We will take μ_0=[1 1 1 1]/4. Figure 5.4 illustrates what is done in the case of a definite negative quadratic form. In blue we have the negative quadratic form, in red the tangent at μ_0.

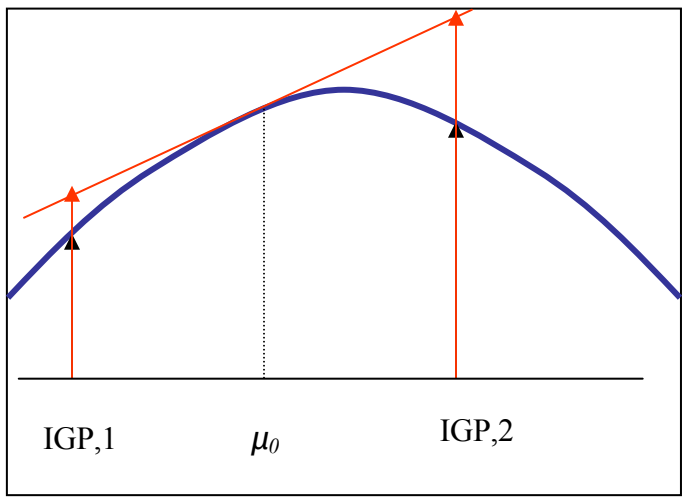

Figure 5.4: Bounding the definite negative term

How do we find the decomposition $S = S^{(+)} + S^{(-)}$? We first show a general method and then show that when we use the kriging estimate at each grid point, there is a natural decomposition. For the general method, we find D diagonal and V orthogonal such that $S = V^T D V$. We then write that $D = D^{(+)} + D^{(-)}$ where the first term has all the positive eigenvalues and the second one the negative ones. We then have $S^{(+)} = V^T D^{(+)} V$ and $S^{(-)} = V^T D^{(-)} V$. Notice that S is only a 4 by 4 symmetric matrix, so this computation is easy to realize in real time. This computation is valid for coefficients λ, such that the estimate is unbiased. There is no need to assume that the coefficients are computed using kriging.

For the case in which the coefficients are computed using the kriging method using the same inflated covariance, there is a natural decomposition that does not require the search for eigenvalues. We have:

$$S = \left[X_i - G^T WC\left(x_i, x_k\right)\right]^T \left(G^T WG\right)^{-1} \left[X_i - G^T WC\left(x_i, x_k\right)\right] - C\left(x_i, x_k\right)^T WC\left(x_i, x_k\right)$$

It is clear that the first term is positive semi-definite and the second one is negative semi-definite.

5.3.4 BOUNDING THE LAST TERM

The last term is the most problematic due to its dependence on the covariance from the IPP user location to the IPP measurements, which is not known by the user. The idea here is to find an upper bound for:

$$\lambda_i^T \left(\sum_{j=1}^{4} \mu_j C\left(x_j, x_k\right) - C\left(x, x_k\right) \right)$$

We first show that there exists ε and δ positive such that:

$$-\delta \le \sum_{j=1}^{4} \mu_j C\left(x_j, x_k\right) - C\left(x, x_k\right) \le \varepsilon \tag{5.4}$$

These inequalities exploit the particular shape of the covariance (the expression for it was given in Chapter 3.) We have:

$$C\left(x_j, x_k\right) - C\left(x, x_k\right) \ge -u_0^2 c \frac{d_{\max}}{a} = -\delta$$

where $d_{\max}$ is the maximum distance between x_i and x_k, which is smaller than 700 km in the case of a 5 by 5 grid. This inequality is simply the fact that the difference between two values of a function is smaller than the maximum of the derivative times the distance between the two values.

For the other inequality, we use the fact that within a quadrant, the covariance from a measurement to any other point in the quadrant is very close to concave. This function has the shape of an inverted cone whose vertex is at x_k and whose generator is:

$$C(d) = u_0^2 c e^{-\frac{d}{a}}$$

for d positive. This function is almost affine over a distance of 700 km (since a=32000 km). As a result, it is easy to find – but cumbersome to write – a concave function such that within the four IGPs the following is true:

$$0 \le C'(x, x_k) - C(x, x_k) \le \varepsilon$$

with ε very small (a value is given in Appendix A). Figure 5.5 illustrates how to define C' and the resulting ε.

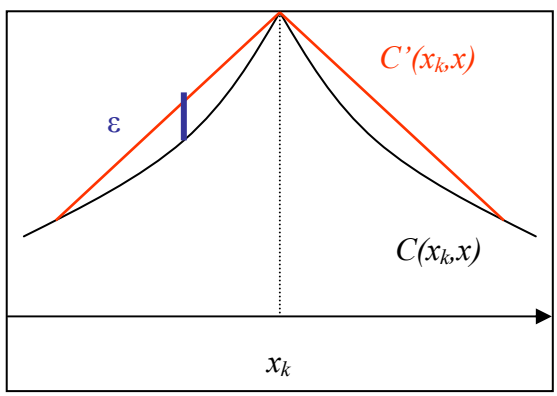

Figure 5.5: Diagram of the covariance and a concave approximation

Since $x = \sum_{j=1}^{4} \mu_j x_j$ and C' is concave we have:

$$\sum_{j=1}^{4} \mu_j C'(x_j, x_k) - C'(x, x_k) \le 0$$

As a result we have:

$$\sum_{i=1}^{4} \mu_j C\left(x_j, x_k\right) - C\left(x, x_k\right) \le \varepsilon$$

Now we can write that:

$$\lambda_i^T \left(\sum_{j=1}^{4} \mu_j C\left(x_j, x_k\right) - C\left(x, x_k\right) \right) \le \delta \sum_{\lambda_{ik}<0} \left|\Lambda_{ik}\right| + \varepsilon \sum_{\lambda_{ik}>0} \left|\Lambda_{ik}\right| = \eta_i$$

We finally have:

$$E\left(\hat{I}_{user} - I_{true}\right)^2 \le \sum_{i=1}^{4} \mu_i C\left(x_i, x_i\right) + \sum_{i=1}^{4} \mu_i S_{ii}^{(+)} + \sum_{i=1}^{4} \mu_i b_i + \sum_{i=1}^{4} \mu_i \eta_i$$

This means that we can take:

$$\sigma_i^2 = C\left(x_i, x_i\right) + S_{ii}^{(+)} + b_i + \eta_i \tag{5.5}$$

As said earlier, for each IGP there are four possible σ_i^2 corresponding to each quadrant where the IGP has influence (A, B, C and D). The computed σ_{GIVE}^2 will be the maximum of these four values.

5.3.5 GIVE COMPUTATION: SUMMARY

We summarize here the steps that are necessary to compute the GIVE for a given IGP (let us index it by *i*). First, we determine the quadrants that are influenced by IGP_i. Usually, this includes the four 5 by 5 degree quadrants that have IGP_i as a common edge (However, at the edge of coverage the pattern will be slightly more complicated). For each of these regions, we need to compute all σ_i^2 associated with IGP_i, $\sigma_i^2\left(A\right)$, $\sigma_i^2\left(B\right)$, $\sigma_i^2\left(C\right)$, and $\sigma_i^2\left(D\right)$. There are therefore four calculations to complete, one for each quadrant.

The first and fourth terms in (5.5) are common to all four of them, so they do not need to be computed 4 times. For the second and third term we need to compute the matrix *S* for each quadrant. For this purpose we need to determine the matrix Λ for each quadrant, that is, the

kriging coefficients at each one of the eight surrounding IGPs. Once we have W, this step only involves matrix products.

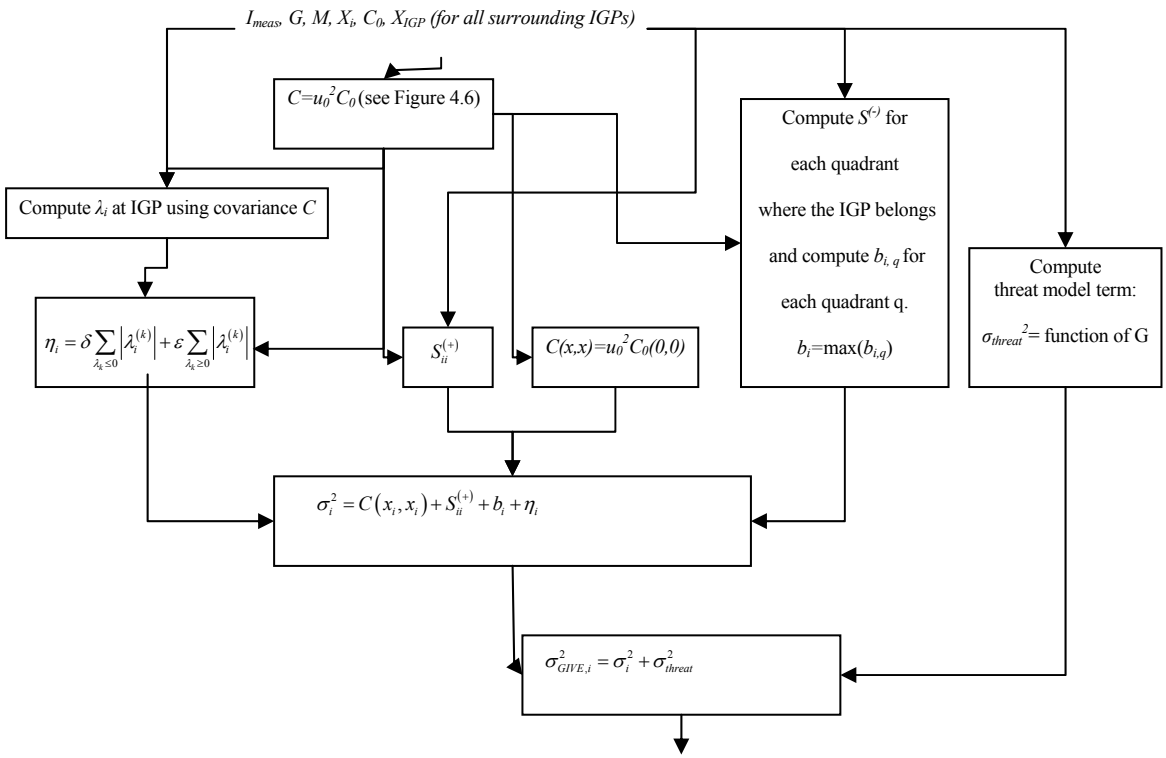

Figure 5.6: GIVE computation flowchart

Figure 5.6 summarizes the calculation of the GIVE. This flow chart is valid for one IGP: first the inflation factor is computed (see Chapter 4); Once this done, we compute the kriging coefficients corresponding to the IGP considered, the term $S_{ii}^{(+)}$, $C(x,x)$, and the term b_i. We sum them and add the threat model term, which only depends on the geometry of the IPPs.

Chapter 6

Performance Analysis

For a safety critical system like WAAS or any other satellite based augmentation system whose main purpose is to provide integrity to users, the first requirement is the ability to give a hard bound on the estimation error. For this reason, the first thing to check in the ionospheric estimation algorithm is that the computed error bounds are really bounding the possible estimation error. Only when we are confident that the error bounds are always an upper bound of the inaccuracy can we begin to evaluate the benefits of the algorithm in terms of availability.

This chapter starts by checking the integrity of the algorithm outlined in Chapter 5. For this purpose we will describe the different possible schemes to test integrity, which are all based in cross-validation and data deprivation [Sparks01]. Once we have shown that the estimation algorithm is safe (within reasonable data deprivation schemes), we will explain the different metrics that are commonly used to measure the performance of an SBAS. We will continue by introducing the Matlab Algorithm Availability Simulation Tool (MAAST), the service volume analysis tool used to evaluate the performance of the ionospheric estimation algorithm once it is integrated in the system. Then, the results of the MAAST simulations will be plotted and compared to the simulations generated using the current WAAS ionospheric estimation algorithm. In the last section we evaluate the performance of the algorithm without considering the bandwidth limitations.

6.1 INTEGRITY CHECK

In order to check the integrity of the algorithm offline, we want to test whether the computed error bounds are larger than the actual error. As explained in Chapter 2, the requirement is more precise than that, and we will go back to it. There is only one way to do this, which is by using cross-validation [Hastie]. Cross-validation is widely used for estimating prediction error. Here we use it to test whether the predicted error is larger than the actual error. The idea is to put aside a validation set and use it to assess the capacity of the algorithm to produce safe error bounds. That is, we compute the estimate and the error bounds based on the set of remaining measurements. To be exhaustive, the validation set should be any possible set (any subset of the measurements should be tested as a validation set). The large amount of data make this approach infeasible, so we limit the cross-validation to certain types of sets. First we will consider *leave-one-out* cross-validation or single cross-validation, which consists of excluding each measurement one at a time and performing the estimation with the remaining measurements. Then we will consider a more severe data-deprivation scheme.

The data used for these integrity checks is the post-processed data corresponding to sixteen quiet and stormy days over CONUS (February 12, 2000; March 26, 2000; April 6-8, 2000; May 25, 2000; June 6, 2000; July 2, 2000; July 15-16, 2000; March 31, 2001; April 1, 2001; September 4, 7, 8, 11, 2002). Among this set there are many more days corresponding to stormy days. In reality, the percentage of storm days is small (5%, even during solar maximum). The total number of measurements tested was 1,592,603. This is not enough to test our algorithm to 10^{-7}, but there were not more measurements available.

Because we use post-processed data, the measurement noise is assumed to be zero. This integrity check does not test the final algorithm; rather, it tests that the assumed model leads to conservative error bounds. In addition, we are not including the effects of the interpolation error when using the grid (we use the measurements directly to get the kriging estimate at the location to be estimated) and the threat model term is not included in the confidence bound (we want to test how well the algorithm would protect the user without any threat model).

6.1.1 LEAVE-ONE-OUT CROSS-VALIDATION

In this first section we present the results of single (or leave-one-out) cross-validation. For each IPP measurement we first select the measurements that are within a certain radius of the IPP location. (This radius search replicates the search in the real algorithm, and was explained in subsection 5.1.5). Then, we transform the geodetic coordinates to East-North-Up coordinates based on the location to be estimated. After that we estimate the inflation factor by computing the chi-square estimate based on the nominal model and multiplying it by α_n (the table for it was determined in Subsection 5.2.4). Because there is no measurement noise, we do not need to determine the ratio β (it is one). Once we have the inflation factor, we can compute the kriging estimate and the error bound from the inflated model. For the noiseless case, the kriging estimate is the same for any inflation factor because the matrix C is a multiple of the nominal covariance, C_0. This is not the case in general. In this analysis, the chi-square threshold was not implemented. Figure 6.1 shows a histogram of the ratio:

$$\frac{\hat{I}_{est} - I_{true}}{\sigma_{est}}$$

In this equation, $\hat{I}_{est}$ is defined in Equations (4.3) and (4.13), and σ_{est} in Equation (4.14). We will call this ratio the *normalized residual* (here the word residual no longer designates the process noise and instead refers to the difference between true value and estimated value). The first requirement of the normalized residuals is that they need to always be below the factor K =5.33. Another requirement, mentioned in Chapter 2, is that the distribution of the normalized residuals must always be overbound by a unit Gaussian in the sense that for any s we need to have:

$$\text{Prob}\left(\left|\frac{\hat{I}_{est} - I_{true}}{\sigma_{est}}\right| \le s\right) \le \int_{-s}^{s} \frac{1}{\sqrt{2\pi}} e^{-\frac{w^2}{2}} dw = 1 - 2Q(s)$$

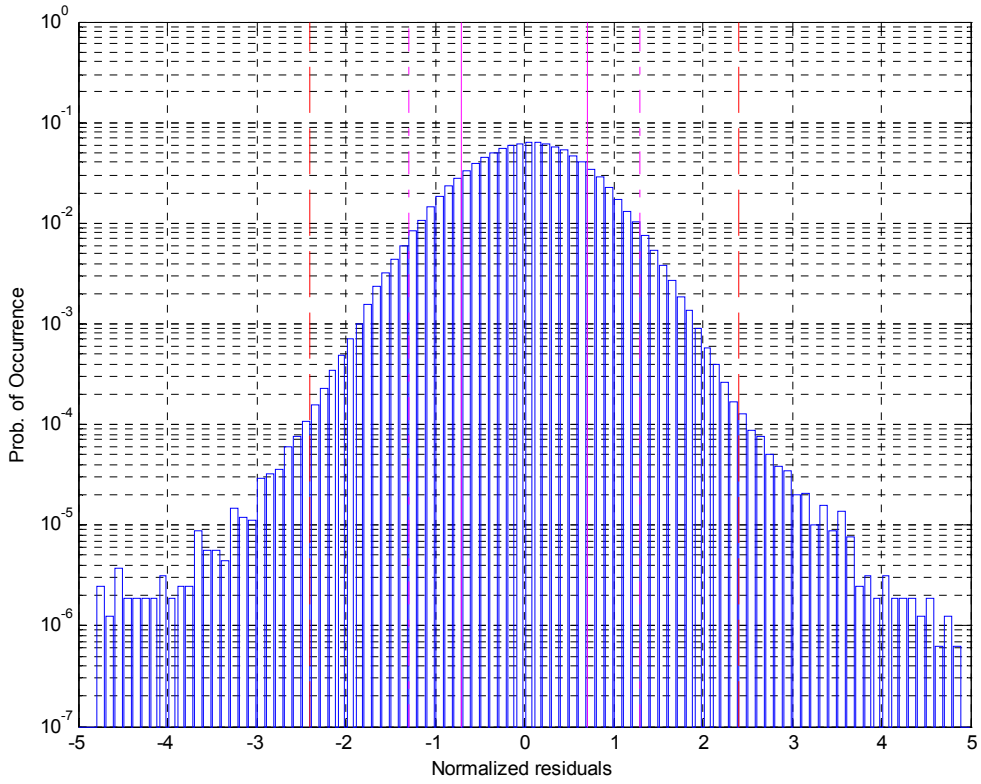

Figure 6.1 Histogram of normalized residuals with single cross-validation

We see from Figure 6.1 that the first requirement is met. None of the normalized residuals is above K (in fact they are well below). The second requirement can be checked by plotting a quantile-quantile plot which is shown in Figure 6.2.

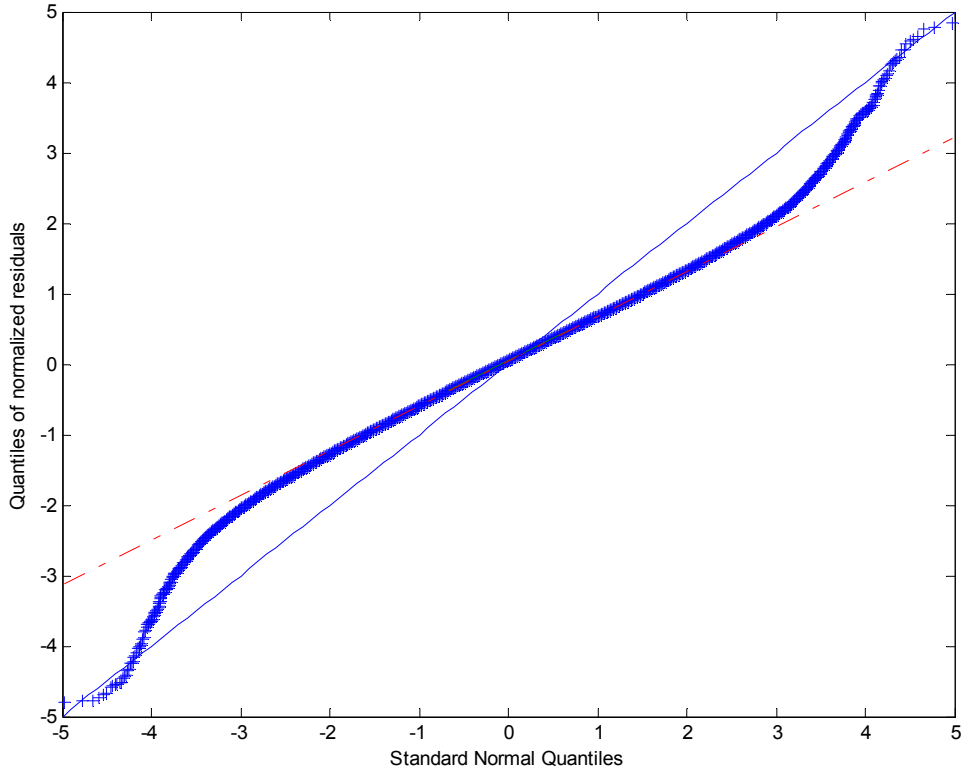

Figure 6.2: Q-Q plot of the normalized residuals

We see that the empirical quantile corresponding to the normalized residuals is below the corresponding unit Gaussian quantile (blue line), except for a few quantiles for which it is still very close. Upon closer examination of the large normalized residuals, we find that they come from data on April 6, 2000, July 16, 2000 and March 31, 2001 at locations where the test measurement was poorly surrounded by measurements, and a large gradient is affecting the test measurement but not the remaining measurements. It is not a consequence of the randomness of the residual field but rather of a large undetected feature. It is in this type of situation that the threat model recognizes a poorly sampled location and inflates the error bound according to what has been observed. As a result, these large residuals would be mitigated in the real system.

6.1.2 DATA DEPRIVATION

In this subsection, we apply data deprivation. That is, instead of only excluding the data to be estimated, we exclude a whole set. As we said earlier, ideally the algorithm should be able to handle any deprivation scheme. But we are limited by our assumption about stationarity. Situations like the one pictured in Figure 4.11 are likely to cause problems when we suppose that all the IPPs below 25 degrees of latitude are not observed. There are many possible data deprivation schemes. Here, we choose the deprivation scheme that best simulates a user: we exclude all the data coming from the station where the measurement was taken. This method more closely mimics the estimation error of a user, because there is no ionospheric information in similar lines of sight as the one tested. Figure 6.3 shows the empirical distribution of the normalized residuals.

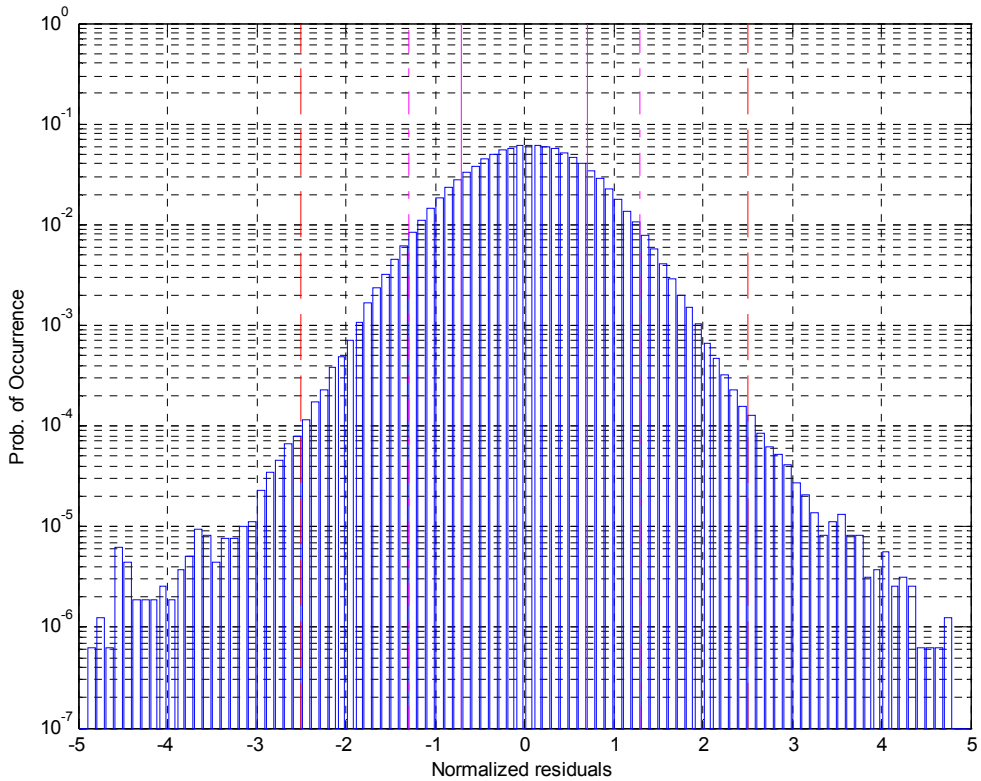

Figure 6.3 Histogram of normalized residuals with station removal

Again, there is no residual above 5.33. The fact that the residuals have not increased with less data indicates that the algorithm is taking into account correctly the smaller data set and the increasing confidence bounds. Figure 6.4 shows the quantile-quantile plot of the residuals. Again, the plot is similar to the previous q-q plot.

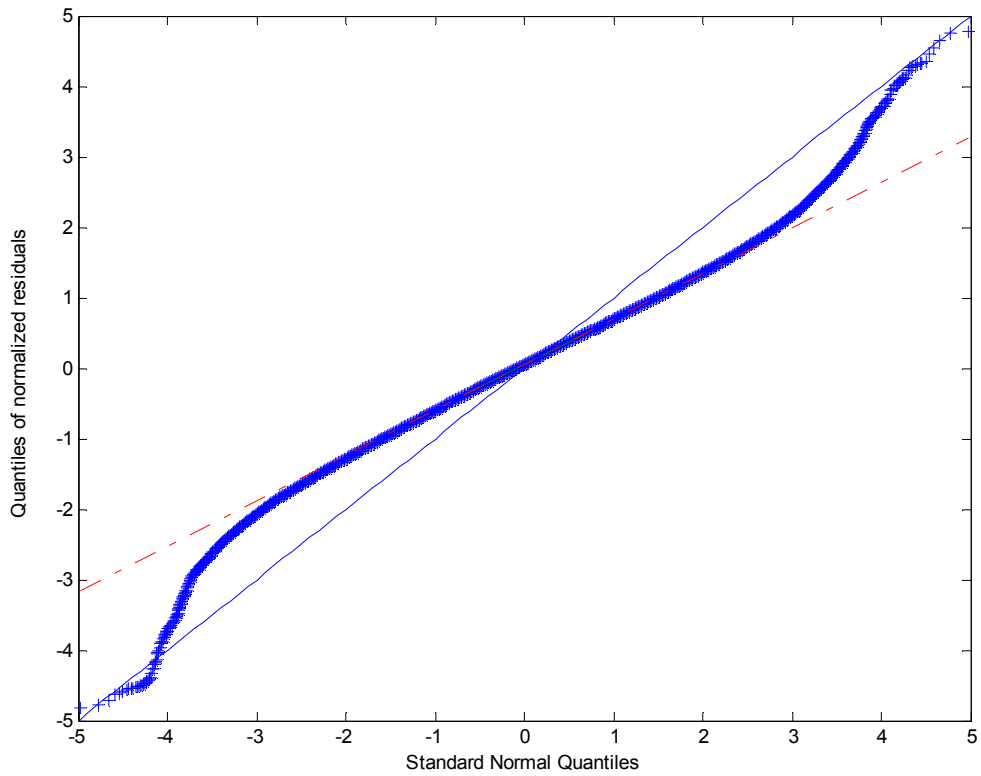

Figure 6.4: Q-Q plot of the normalized residuals with station removal

6.2 PERFORMANCE METRICS

Once we are confident that the designed algorithm is safe, we want to know how small the calculated error bounds are and how this translates to the user's bound on inaccuracy. In this section we explain the different ways of measuring the performance of an ionosphere

estimation algorithm. Before introducing the two metrics used, we need to present the simulation tool used to generate the results.

6.2.1 MAAST

The MAAST is a service volume analysis tool. The goal of this tool is to predict the availability of WAAS. For this purpose, it computes all the configurations over a period of time (typically 24 hours) long enough to be representative of all typical satellite configurations. For each one MAAST computes the message that would be sent to the users, in particular, the UDRE and the GIVE. After that, it computes the Vertical Protection Level that a user would compute for different locations over CONUS. A full description of MAAST can be found in [Jan].

All terms except the GIVE depend only on the geometry of the satellites (and the history of the geometry). As shown, the GIVE is dependent on the actual behavior of the ionosphere through the inflation factor u_0. MAAST has two ways of modeling the effect of the inflation factor on the GIVE, one with and the other without real data. The GIVE was computed using the process illustrated in Figure 5.6.

Without real data, we can evaluate the performance of the algorithm when the ionosphere is described conservatively by the nominal model. Assuming this model, we know that the chi-square statistic will be above a threshold, T, with probability P_{fa} (probability of false alarm). If we set this probability low enough, we know that u_0^2 will be below αT with high probability. Let us take $P_{fa}=10^{-4}$. For a given day with nominal ionospheric conditions we have an upper bound of u_0 for any situation. There are two main drawbacks in this option. First, it is slightly pessimistic in the sense that the chi-square statistic is usually much lower than the threshold by a factor of two or three. Second, it does not allow us to simulate real situations where the ionosphere is slightly disturbed.

MAAST can also be used with archived real data. Ideally, we would use archived data that replicates exactly the data that is fed to the estimation algorithm in real time. As said earlier, this data is noisier than the post-processed data. Unfortunately, access to the rea; data being fed to the real time algorithm (the noisy data) is limited. The idea is to use the

post-processed data. There is a large amount of it and it covers a broad range of ionospheric behavior. Because it has low measurement noise, as opposed to the real situation, we need to compute the chi-square as if there was no measurement noise. If we did not do this, the chi-square statistic would be lower than it should be. However, we still consider that there is measurement noise when computing the estimation variance. It is for this reason that in this thesis real data is used. This method will allow us to more accurately evaluate the inflation factor during nominal and disturbed ionospheric behavior. The drawback is that the post-processed data filters out a large amount of measurements because their noise it too large. As a consequence, the data appears to be scarce in regions where it would not be in real time.

6.2.2 GRID IONOSPHERIC VERTICAL ERROR AND VPL

The Grid Ionospheric Vertical Error is the quantity that is sent to the user at each IGP. For historical reasons, the GIVE is defined as:

$$GIVE = 3.29\sigma_{GIVE}$$

The first and most direct way of evaluating the performance of the algorithm is by computing the error bounds, or equivalently, the Grid Ionospheric Vertical Error over a period of time. This period of time should be such that the set of satellite configurations is representative of any situation. Since the period of a satellite is approximately 12 hours, a period of 24 hours is enough. The next question is how to display the results. It is not enough to compute the mean of the error bounds and see how this mean compares to the mean of another estimation algorithm. It could be that the error bounds are larger where we do not really care but smaller where we want them to be small. Showing the mean of the error bound per location is still not enough because the extreme values may be averaged out and it is those extreme values that can hurt availability. Figure 6.5 is an example of how we will plot the GIVE values over CONUS. As indicated earlier, the circles correspond to the location of the IGPs.

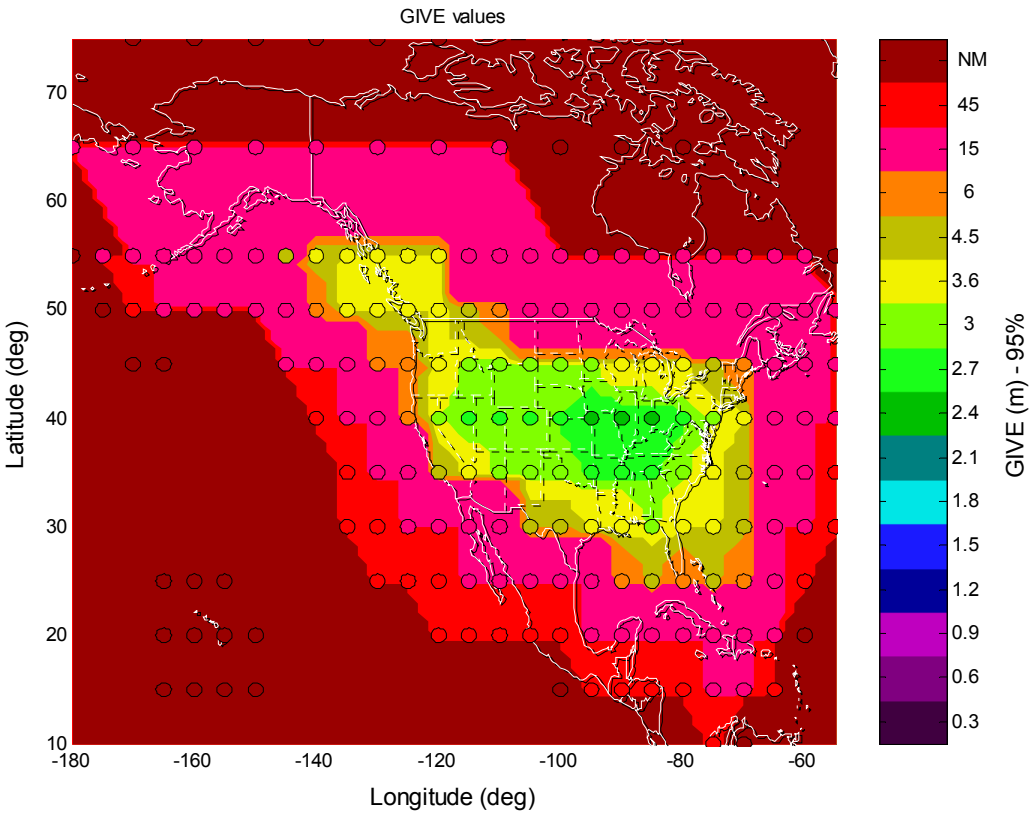

Figure 6.5: GIVE map over CONUS

To obtain this plot we first run MAAST. At each location we can create a histogram of the GIVE on the grid. At each IGP, the color code corresponds to the magnitude of the 95th percentile of the GIVE. Similarly, we can plot VPL quantiles, where the map gives the position of the user.

6.3 RESULTS WITHOUT THREAT MODEL

Here we show the results without including the threat model term. These represent an upper bound of the performance. We will use here real data from a quiet day (July 2,2000)

and a disturbed day (September 8,2002). The threat model is not used, and the storm detector is turned off.

In order to have an idea of the performance of the algorithm designed in this thesis, we will compare the results given by MAAST with kriging with the results given by the proposed Final Operation Capability (FOC) estimation algorithm for WAAS proposed by Raytheon. This other algorithm has several common features with the algorithm proposed in this thesis [Cormier]. The Raytheon proposed algorithm is based on a planar fit and a dynamic inflation factor of the GIVE based on the chi-square statistic.

6.3.1 QUIET DAY

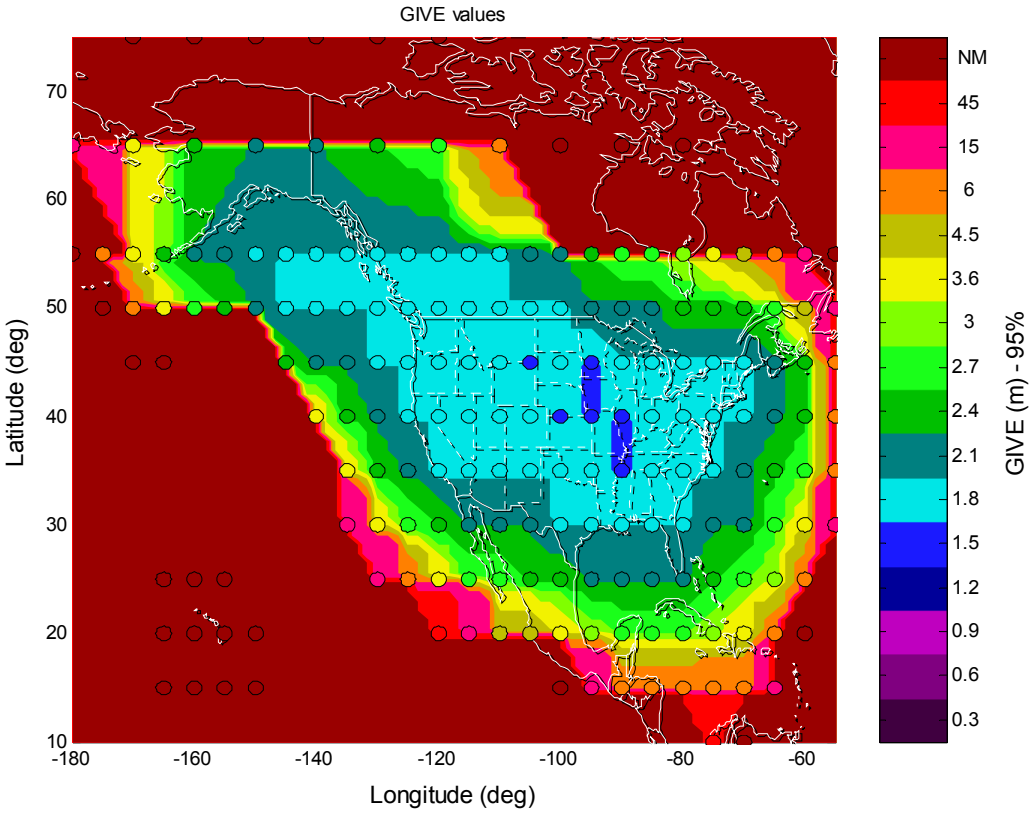

Figure 6.6: GIVE map for July 2, 2000 using kriging

In Figure 6.6, the 95 percentile of the GIVE at each location is plotted. These GIVEs combined with the other errors (in particular the UDRE) as explained in Chapter 1 produce the VPLs plotted in Figure 6.7. This map can be compared to the VPL produced by the Raytheon proposed FOC algorithm and plotted in Figure 6.8. We see that the region with VPLs below 20 meters has increased considerably. This will not increase the level of service immediately because the critical VAL is 50 meters, which is already attained on quiet days with the current algorithm. It would however put WAAS closer to the next level of service, which requires VPLs below 20 meters 99.9% of the time. Also, it would make it more robust to satellite outages.

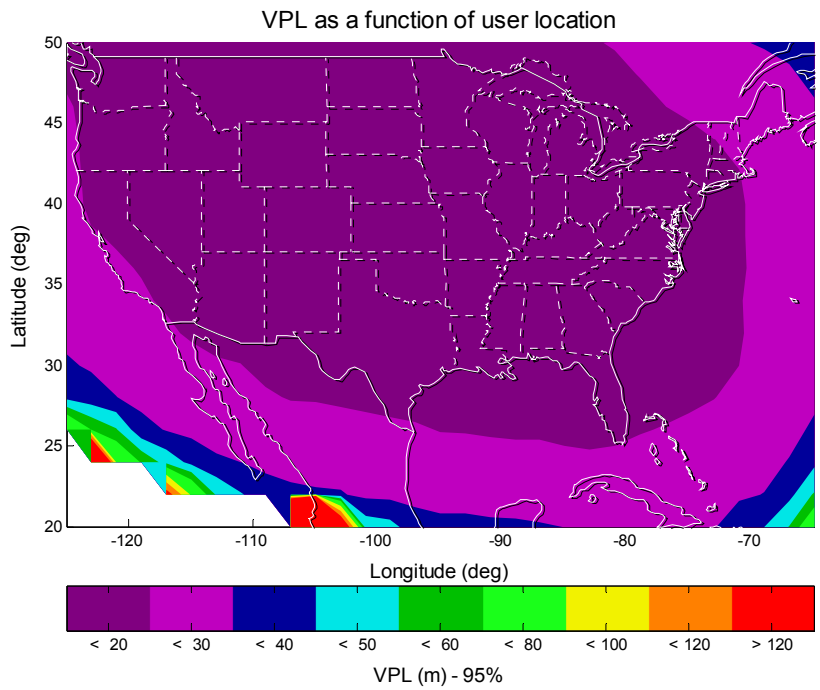

Figure 6.7: VPL map for July 2, 2000 using kriging

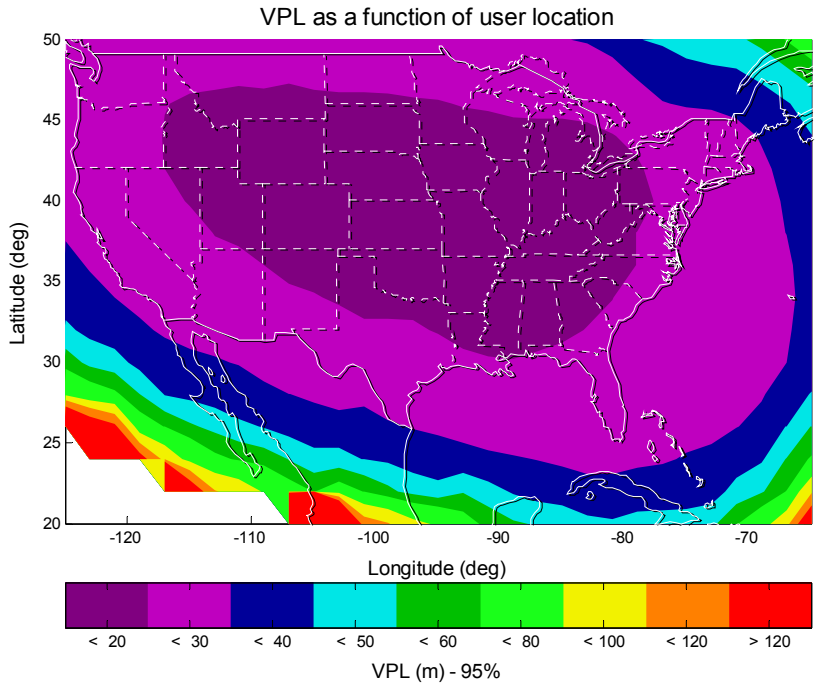

Figure 6.8: VPL map for July 2, 2000 using Raytheon's algorithm

6.3.2 DISTURBED DAY

In this section we examine the performance of the algorithm during a mildly disturbed day which did, however, considerably limit WAAS performance. The same settings as in the previous section were used. As expected, the GIVE values (Figure 6.9) are larger than on a quiet day. This is due to the larger decorrelation of the ionosphere, which is captured in the algorithm through the inflation factor u_0^2, whose determination is the subject of Chapter 4.

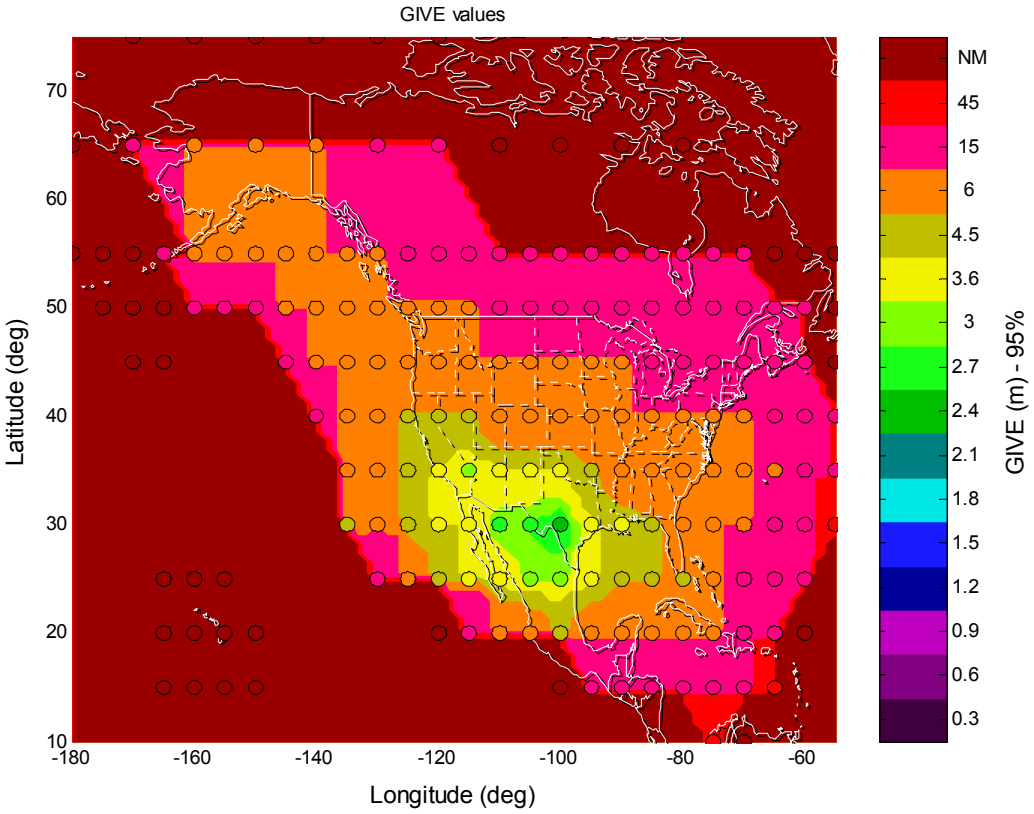

Figure 6.9: GIVE map for September 8, 2002 using kriging

Figure 6.10 shows the corresponding VPLs, which are larger than on a quiet day and Figure 6.11, the VPLs obtained with the Raytheon proposed FOC algorithm. The difference now is dramatic: the VPLs have been reduced by almost 40%.

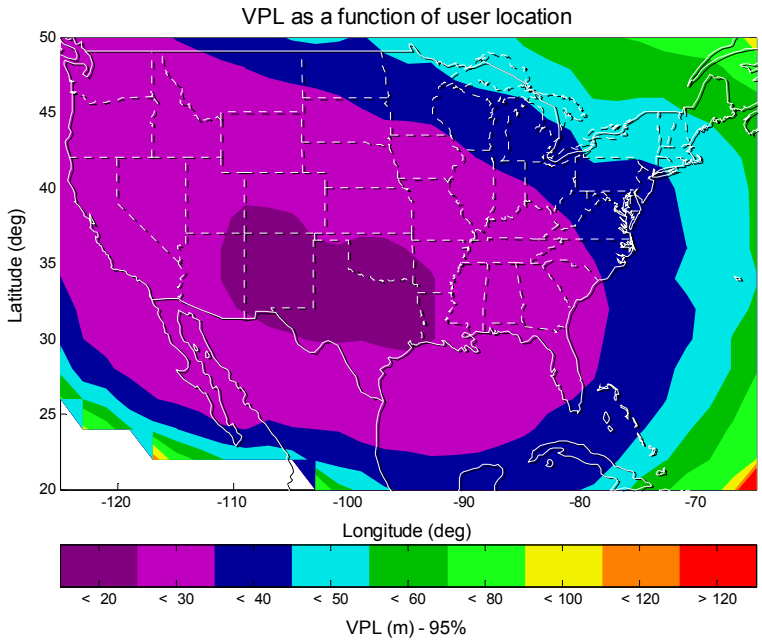

Figure 6.10: VPL map for September 8, 2002 using kriging

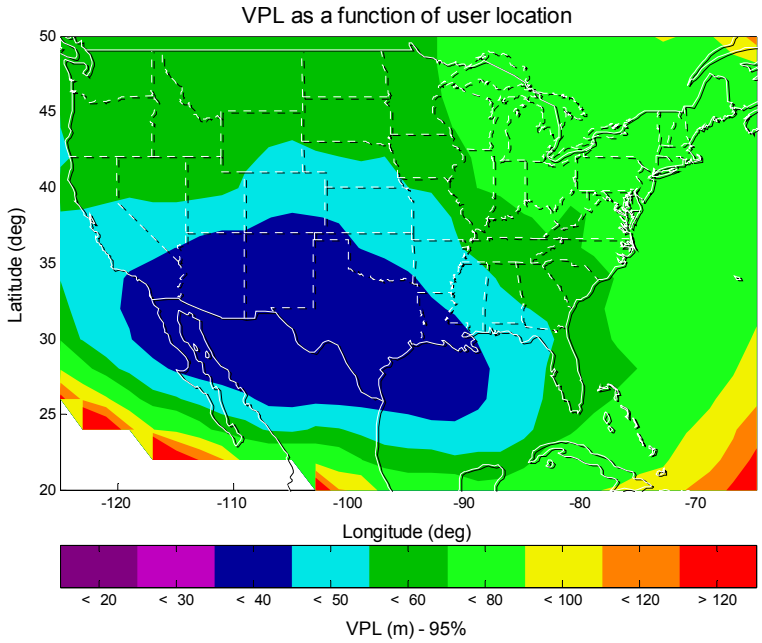

Figure 6.11: VPL map for September 8, 2002 using Raytheon's algorithm

This difference in performance comes mainly from the different characterizations of the ionosphere. In both algorithms, the ionosphere is characterized as a planar trend with a residual random field (see Chapter 3). The difference is that in our algorithm we assume that the residual field has a distance dependent covariance, that is, IPPs that are closer are more likely to be correlated, whereas in the other algorithm, the residual field is uncorrelated from one point to another. While that might be a good approximation on quiet days (although there is already a degradation during quiet days as seen in the previous subsections) it does not work as well during mildly disturbed days, when the ionosphere, though not clearly following a planar trend, is still correlated with distance.

6.4 RESULTS WITH THREAT MODEL

In this section we show the results including the threat model term. As said in Section 4.4, the threat model term is dependent on the algorithm. A different threat model is used for the algorithm based on kriging and the algorithm used for comparison. We will show the same plots as before where we have included the threat model term.

As mentioned earlier, the threat model is meant to protect users against large ionospheric gradients. For this reason, the GIVEs are mainly going to be degraded at the edge of coverage.

6.4.1 QUIET DAY

Figure 6.12 shows how Figure 6.6 is modified once we include the threat model term. Now, the region with very low GIVEs is smaller but still covers most of the CONUS region.

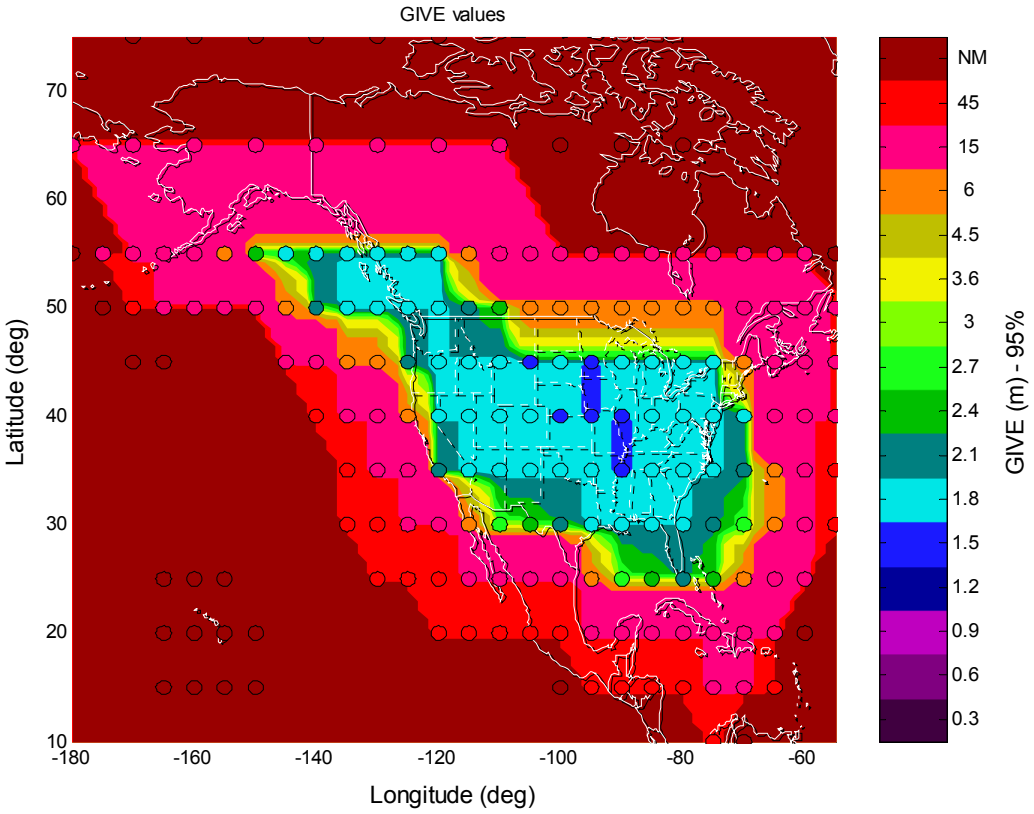

Figure 6.12: GIVE map for July 2, 2000 using kriging with threat model

In Figure 6.13 we show the corresponding VPL map. As expected, the performance is degraded, but most of CONUS is still below 20 meters VPL 95% of the time. The performance of the Raytheon proposed FOC algorithm (Figure 6.14) is also degraded such that the relative improvement is similar to the previous one.

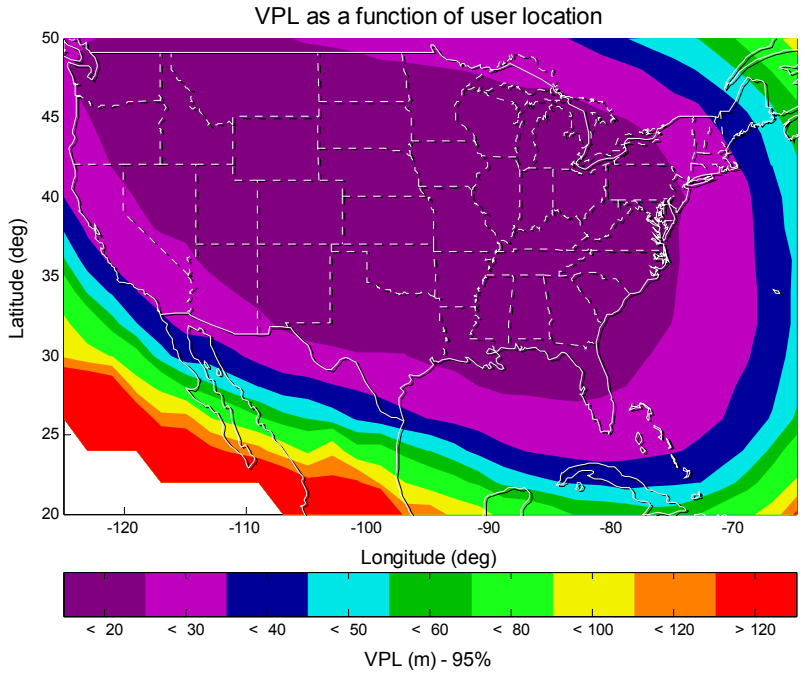

Figure 6.13: VPL map for July 2, 2000 using kriging with threat model

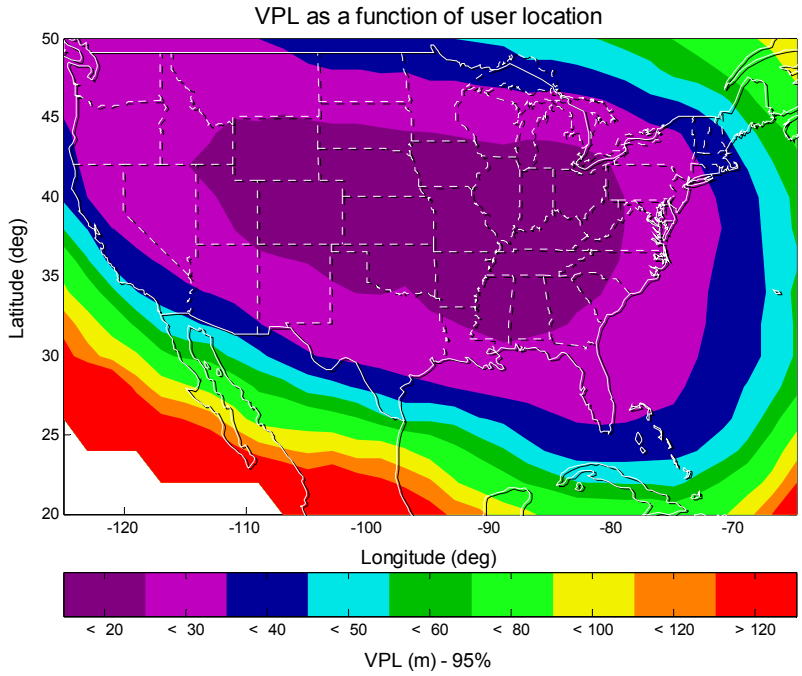

Figure 6.14: VPL map for July 2, 2000 using Raytheon's algorithm with threat model

6.4.2 DISTURBED DAY

In this subsection we have added the threat model term in the GIVE on September 8, 2002. We see again in Figure 6.15 that the GIVE is degraded on the edge of coverage. The effect is not as large as for a quiet day due to the quantization of the GIVE, which is much wider at larger values. The effect on the VPL is shown in Figure 6.16. Again the performance is slightly degraded compared to the case without the threat model. As a comparison, we plot the VPL for the Raytheon proposed FOC algorithm with threat model in Figure 6.17.

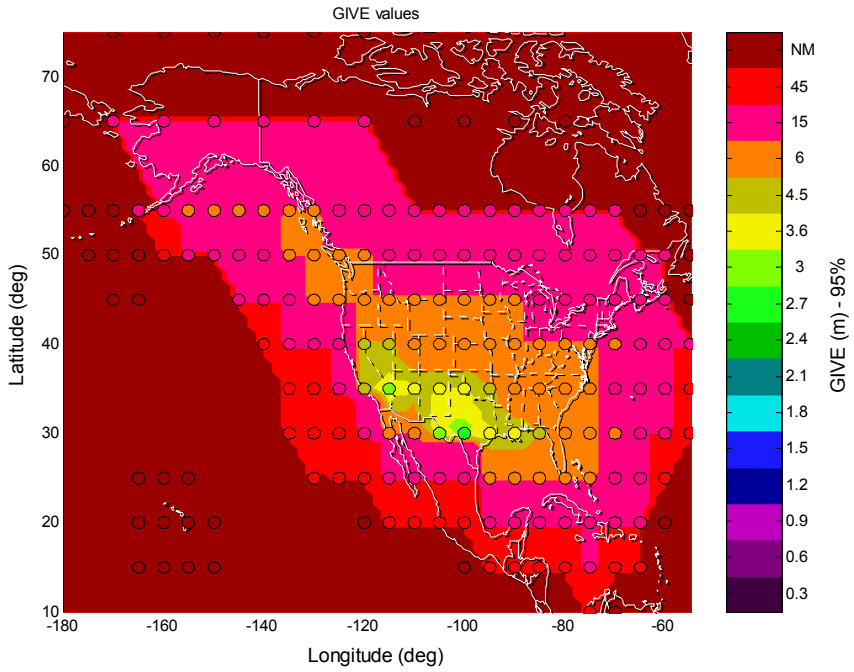

Figure 6.15: GIVE map for September 8, 2002 using kriging

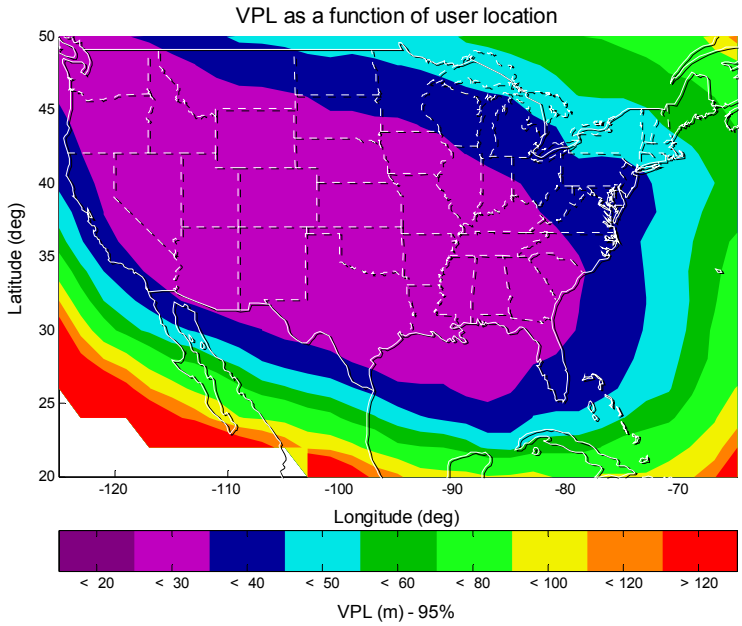

Figure 6.16: VPL map for September 8, 2002 using kriging

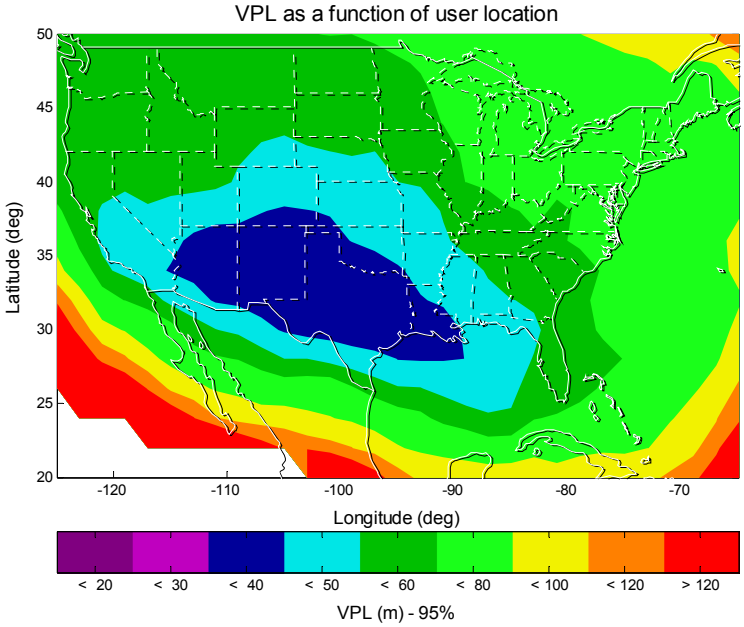

Figure 6.17: VPL map for September 8, 2002 using Raytheon's algorithm

6.5 LOSS OF AVAILABILITY DUE TO THE MESSAGE

At the end of Chapter 4 we ended up with an ionospheric estimation algorithm that could work if the user had all the measurement information. However, at that stage the algorithm was not useful with the current WAAS standards. The algorithm was modified in Chapter 5 to take into account the current WAAS standards. There are two steps involving bandwidth limitations that degrade the performance that are due to bandwidth limitations: the use of the ionospheric grid, which requires an additional inflation of the error bound (as we saw in Chapter 5) and the quantization of the Grid Ionospheric Vertical Error (instead of sending the actual value, the master station sends an index corresponding to a tabulated value). In the current algorithm, the quantization is indicated in the color bar of Figure 6.5. In this section, we evaluate the performance of the algorithm without bandwidth limitations. Figure 6.18 shows the VPL map if we suppose that there is no quantization and there is no degradation due to the grid.

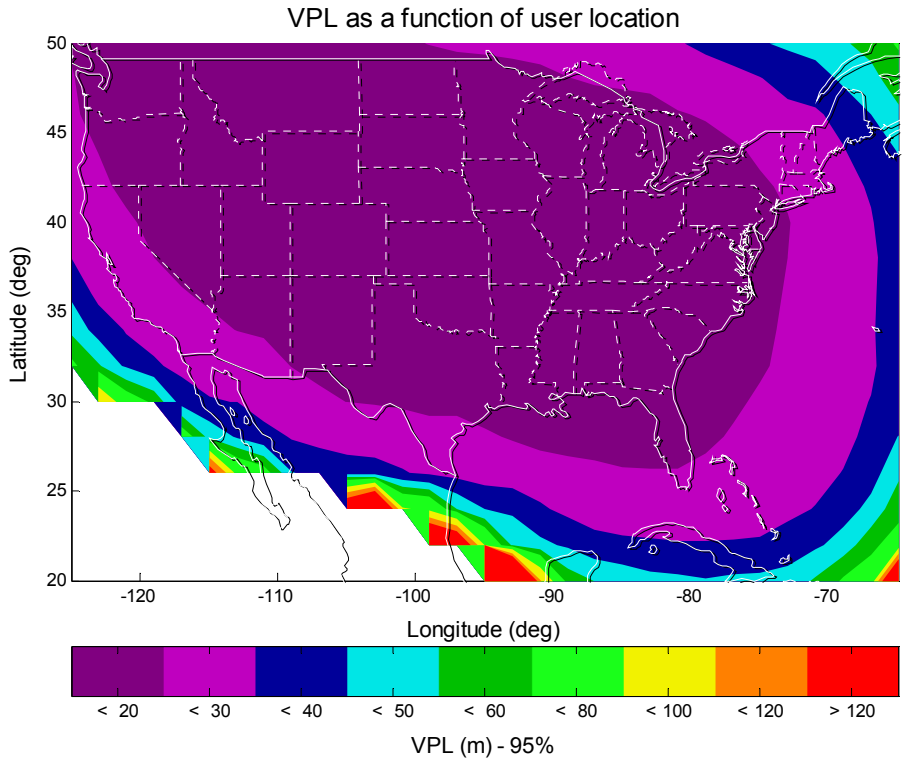

Figure 6.18: VPL map for July 2, 2000 using kriging without bandwidth limitations

This map can be compared with the VPL map shown in Figure 6.13. We see that we gain some performance, but the difference is small. To know whether the (small) performance loss of the full algorithm is due to the quantization or the inflation due to the grid, we plot in Figure 6:19, the VPL without quantization but with the grid.

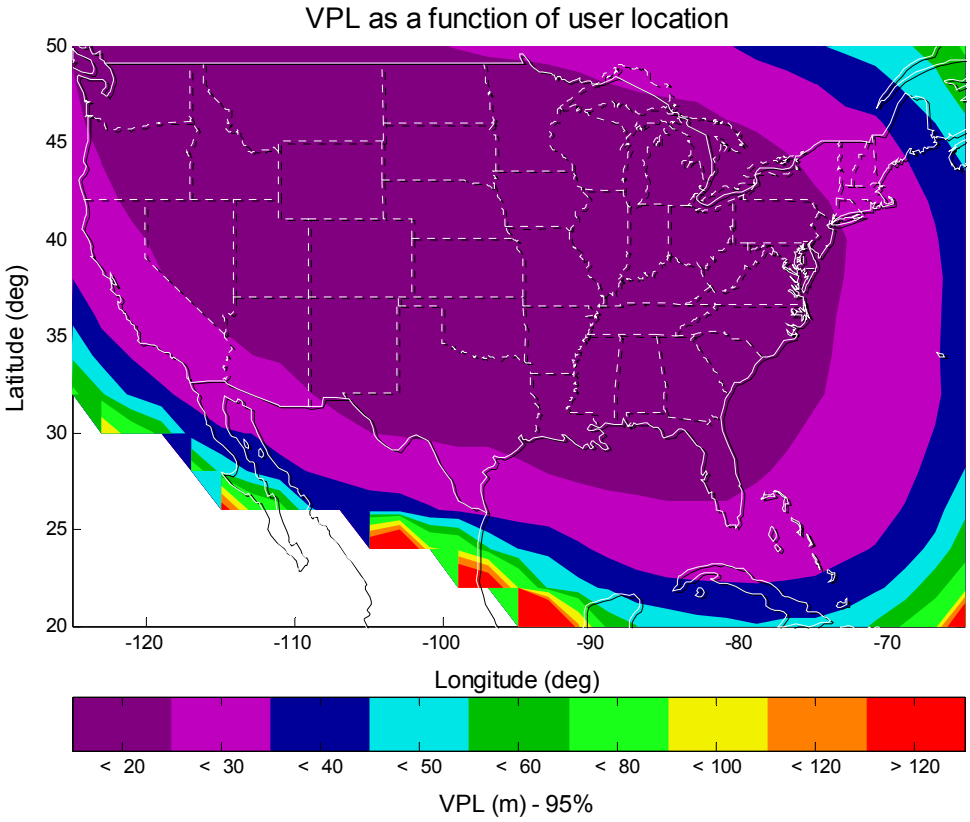

Figure 6.19: VPL map for July 2, 2000 using kriging without quantization

This map is virtually identical to the previous one, so the grid does not introduce a significant degradation.

Chapter 7

Conclusion

7.1 SUMMARY OF RESULTS

The goal of this thesis was to design an algorithm for satellite based augmentation systems that, from GPS measurements collected in a network of reference stations, estimates the delay caused by the ionosphere at any location within the reference stations, and provides an error bound. Several problems needed to be addressed. First, the measurements are irregularly scattered in the region to be estimated. Second, the measurements are noisy. Third, the process noise due to the ionosphere cannot be characterized by a single process noise and must be bounded in real time. Finally, for such an algorithm to be useful, it has to fit within the current standards of satellite based augmentation systems. We review now the different contributions of this thesis to solve this problem.

7.1.1 VERTICAL IONOSPHERIC DELAY STRUCTURE

The first step (Chapter 3) addressed the characterization of the vertical ionospheric delay, that is, the ionospheric delay projected on the thin shell. We showed that it is possible to interpret the vertical ionospheric delay as the combination of a planar trend and a random residual field correlated with distance. This residual field can be characterized by a variogram with a non-zero intercept and a certain slope. We determined the variogram for

a quiet day (or nominal day) from real data, and showed that for more disturbed ionospheric behavior we could assume that the nominal variogram multiplied by a certain factor, u, can be used. Because this model assumes stationary stochastic processes it cannot capture correctly the occasional non-stationarity of the ionosphere. In particular, we noted that large gradients pose a very real threat.

7.1.2 IONOSPHERIC ESTIMATION

An ionospheric estimation algorithm was designed in Chapter 4 based on the characterization given in Chapter 3. The algorithm is based on the well known estimation method called kriging, which is a type of minimum mean square estimator. The classical application of kriging requires the knowledge of the process noise covariance. Since the process noise covariance is not known in our problem, a covariance that conservatively describes the real covariance needed to be estimated. This was realized by inflating the nominal covariance (given in Chapter 3) by a factor dependent on the number of measurements, a chi-square statistic computed from the measurements and the measurement noise. The dependence on measurement noise results in the error bound being largely dependent on the quality of the measurements. This is a different situation to one where the process noise is well known (in this last case, the measurement noise has little impact on the final error bound because it is averaged out), and the successful treatment of this problem in this thesis might be its most relevant contribution. (To fully protect the user, the algorithm includes a term coming from the *threat model*, an external part of the algorithm not developed in this thesis. The threat model protects users against severe gradients and non-stationarity.)

Another contribution of this thesis is the adaptation of kriging to the WAAS ionospheric grid. Typically, kriging requires the user to know all measurements to compute the estimate and the error bound. We have shown that the user can interpolate the values broadcast by WAAS in the ionospheric grid (computed by the master station using kriging) if the error bounds at the grid points are carefully computed. This was done in Chapter 5 using the specific properties of the covariance and the interpolation scheme used by the receivers (and specified in the MOPS).

7.1.3 PERFORMANCE

In Chapter 6 we showed that the algorithm designed in this thesis can provide a significant improvement compared to the current design. It is difficult to give a single number reflecting the improvement; the maps shown in Chapter 6 might summarize the situation better. Additionally, the relative improvement depends on the day that is considered. We have shown that on a quiet day there is a sizable relative improvement: the region with a VPL lower than 20 meters covers most of CONUS. However, this improvement does not have an impact on availability on a quiet day, since the availability is already maximal (although it does bring the system closer to the next level of performance). The improvement has much more effect on mildly disturbed days, and the plots shown in Chapter 6 corresponding to September 8, 2002 suggest that the VPLs could be lowered as much as $0% with this algorithm, bringing VPLs that were as high as 60 meters to 40 meters. As an average over several days, we expect this algorithm to provide at least a 15% to 20% reduction in VPL.

7.2 SUGGESTIONS FOR FUTURE WORK

The threat model has not been covered in this thesis. We have seen that there is a large difference between performance with and without the threat model. That means that our model is not accounting correctly for the violations of stationarity. This is not unexpected as we are using stationary random models. An ideal estimation algorithm would close the gap between the results with threat model and without; in other words, the threat model would not be needed. Future work should focus on integrating explicitly the possible threats in the ionospheric model, instead of relying on the threat model. (This was one of the goals of kriging, since kriging recognizes undersampled situations , but was not achieved because the threats turn out to be much more catastrophic).

Another area of future research is the adaptation of this algorithm to equatorial latitudes. The algorithm presented here was designed for mid-latitude regions (all data used comes

from CONUS), where the ionosphere is well behaved most of the time, and where the thin shell model is a surprisingly good approximation. This is not the case in equatorial regions. There, the ionosphere has much more structure and variability, and the thin shell model does not perform as well [Komjhaty]. Although much of the analysis done in this thesis relies on the thin shell model, it is possible to relax it to improve the performance in equatorial regions. For example, one could assume that instead of having all the electron content concentrated at one height, we could assume a certain (fixed) vertical profile. This way, we would preserve the two-dimensionality of the estimation without ignoring the three-dimensional character of the ionosphere.

Appendix A

COMPLEMENTARY DEFINITIONS AND PROOFS

This appendix contains the definition of the *Gaussian overbound* used in this thesis, the proof of inequality (4.30) (which is the basis for the computation of parameter β) and the expression of the constant ε introduced in Chapter 5.

A.1 GAUSSIAN OVERBOUND

Let us consider a zero mean distribution over the real axis, $p(x)$. We will say that $q(x)$ is an upper bound of $p(x)$ if we have for any s:

$$\int_0^s p(x)\,dx \geq \int_0^s q(x)\,dx$$

If the distribution q is Gaussian, we say we have a Gaussian overbound. This definition of overbound is interesting because of the result shown in [DeCleene], which states that if the distributions are unimodal and symmetric the convolution of the overbounds of two distributions is an overbound of the convolution of the two original distributions. This result allows us to use covariance propagation to compute probabilities without strictly having Gaussian distributions.

A.2 INEQUALITY IN SECTION 4.2

In this section we show a proof of the inequality (4.2) which is:

$$\left(1+\left(\prod_{k=1}^{p} b_k\right)^{\frac{1}{p}}\right)^{p} \leq \prod_{k=1}^{p}\left(1+b_k\right)$$

It might be possible to derive this inequality from classic inequalities. We give here a short proof. It is easy to check that the function:

$$\ln\left(1+e^{x}\right)$$

is convex. As a consequence we have the inequality

$$\ln\left(1+e^{\frac{1}{p}\sum_{k=1}^{p} a_k}\right) \leq \frac{1}{p}\sum_{k=1}^{p}\ln\left(1+e^{a_k}\right)$$

We obtain our inequality by doing the change of variable $b_k = e^{a_k}$.

A.3 BOUNDING ADDITIONAL TERM IN GIVE

In section 5.3.4 we needed a positive constant such that:

$$0 \leq C'\left(x, x_k\right) - C\left(x, x_k\right) \leq \varepsilon$$

for any x within the quadrant we are trying to protect. $C'\left(x, x_k\right)$ is a concave function that is an upper bound of the covariance and is as close to the covariance as possible. The expression for the covariance is:

$$C\left(x_k, x\right) = u_0^2 c e^{-\frac{|x_k - x|}{a}}$$

Let us define:

$$d_{\min} = \min\left\{\left|x_k - x\right| \mid x \text{ is in quadrant}\right\}$$
$$d_{\max} = \max\left\{\left|x_k - x\right| \mid x \text{ is in quadrant}\right\}$$

We define *C'* as:

$$C'\left(x_k, x\right) = \frac{u_0^2 c}{d_{\max} - d_{\min}}\left(d_{\max} e^{-\frac{d_{\min}}{a}} - d_{\min} e^{-\frac{d_{\max}}{a}} - \left|x_k - x\right|\left(e^{-\frac{d_{\min}}{a}} - e^{-\frac{d_{\max}}{a}}\right)\right)$$

We want to show that there exists ε such that for any $d = \left|x_k - x\right|$ where *x* is in the quadrant we have:

$$0 \le \frac{u_0^2 c}{d_{\max} - d_{\min}}\left(d_{\max} e^{-\frac{d_{\min}}{a}} - d_{\min} e^{-\frac{d_{\max}}{a}} - d\left(e^{-\frac{d_{\min}}{a}} - e^{-\frac{d_{\max}}{a}}\right)\right) - u_0^2 c e^{-\frac{d}{a}} \le \varepsilon$$

The first inequality is easy to verify because *C'* is affine, coinciding with *C* at $d_{\min}$ and $d_{\max}$ and *C* is a convex function of *d* (but not of *x*). Let us now define $\delta = d - d_{\min}$ and $\delta_{\max} = d_{\max} - d_{\min}$. The expression above is now:

$$0 \le u_0^2 c e^{-\frac{d_{\min}}{a}}\left(\left(1 - \frac{\delta}{\delta_{\max}}\left(1 - e^{-\frac{\delta_{\max}}{a}}\right)\right) - e^{-\frac{\delta}{a}}\right) \le \varepsilon$$

We now find the maximum of the function:

$$f\left(\delta\right) = \left(1 - \frac{\delta}{\delta_{\max}}\left(1 - e^{-\frac{\delta_{\max}}{a}}\right)\right) - e^{-\frac{\delta}{a}}$$

over [0,δ]. The maximum is attained when:

$$\frac{\delta}{a} = -\ln\left(\frac{a}{\delta_{\max}}\left(1-e^{-\frac{\delta_{\max}}{a}}\right)\right)$$

We end up with:

$$f(\delta) \le 1 + \frac{a}{\delta_{\max}}\ln\left(\frac{\delta_{\max}}{a}\left(1-e^{-\frac{\delta_{\max}}{a}}\right)\right)\left(1-e^{-\frac{\delta_{\max}}{a}}\right) - \frac{a}{\delta_{\max}}\left(1-e^{-\frac{\delta_{\max}}{a}}\right)$$
$$= 1 + \frac{a}{\delta_{\max}}\left(1-e^{-\frac{\delta_{\max}}{a}}\right)\left(\ln\left(\frac{\delta_{\max}}{a}\left(1-e^{-\frac{\delta_{\max}}{a}}\right)\right)-1\right)$$

This is a function of the parameter *a* (a parameter defining the variogram which is set to 32000km (see Chapter 3)) and $\delta_{\max}$ which in the 5 by 5 degree cells will typically be smaller than 700km. With these constants we end up with:

$$f(\delta) \le 1.25 * 10^{-4}$$

As we saw, *c*=2 so:

$$\varepsilon \le u_0^2 2.5 * 10^{-4}\ \mathrm{m}^2$$

This term will result in a very small contribution to the GIVE, and could potentially be ignored.

A.4 CORRECTION TO THE PHMI EQUATION

This section was added after the submission of this work. The changes introduced here have a very limited impact on the final GIVE. This note is included here for completeness purposes.

Equation 4.19 assumes that the estimation error is independent of the chi-square statistic. In the general case, they are not independent. This note provides an upper bound of the true PHMI as a function of the expression provided in Equation 4.19.

x is the estimation error:

$$x = \lambda^T I_{meas} - I_{unknown}$$

y is the vector of reduced measurements and has the covariance S(u), which will be shortened as S.

The covariance of $[x\ y]^T$ is given by:

$$\operatorname{cov}\left(\begin{bmatrix} x & y \end{bmatrix}^T\right) = \begin{bmatrix} s & \rho^T \\ \rho & S \end{bmatrix}$$

Where we have:

$$s = u^2 \sigma^2_{process}(\lambda) + \sigma^2_{meas}(\lambda)$$
$$S = \Gamma\left(u^2 C_0 + M\right)\Gamma^T$$
$$\rho = \Gamma\left(u^2 C_0 + M\right)\lambda - \Gamma C_v$$

Under our model, P(HMI|u) is given by:

$$P(HMI \mid u) = \int_{K^2 \alpha y^T y\left(\sigma^2_{process}(\lambda)\right) + K^2 \sigma^2_{meas}(\lambda) \le x^2} p_\rho(x, y)\, dx\, dy$$

The density is indexed by the correlation between X and Y.

We now do the change of variables:

$$\begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} s^{-\frac{1}{2}} & 0 \\ 0 & S^{-\frac{1}{2}} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

The expression for the PHMI is now:

$$P(HMI\,|\,u) = \int_{K^2\alpha Y^T SY\left(\sigma^2_{process}(\lambda)\right)+K^2\sigma^2_{meas}(\lambda)\le sX^2} p_\rho(X,Y)\,dXdY$$

$$= \int_{Y^T\Sigma Y+\ell\le sX^2} p_\rho(X,Y)\,dXdY$$

Where we have:

$$\Sigma = K^2\alpha\left(\sigma^2_{process}(\lambda)\right)S$$

$$\ell = K^2\sigma^2_{meas}(\lambda)$$

And the covariance of [X Y] is given by:

$$\operatorname{cov}\left(\begin{bmatrix} X \\ Y \end{bmatrix}\right) = \begin{bmatrix} 1 & b^T \\ b & I_{n-3} \end{bmatrix} = \Delta_b$$

where:

$$b = S^{-\frac{1}{2}} s^{-\frac{1}{2}} \rho$$

Notice that we always have $b^T b < 1$, because the matrix above is a covariance matrix.

We now prove that:

$$\int_{Y^T\Sigma Y+\ell\le sX^2} p_\rho(X,Y)\,dXdY \le \frac{1}{\sqrt{1-b^T b}} \int_{Y^T\Sigma Y+\ell\le sX^2} p_{\rho=0}(X,Y)\,dXdY$$

We have:

$$\int_{Y^T\Sigma Y+\ell\le sX^2} p_\rho(X,Y)\,dXdY = \int_{X=-\infty}^{X=+\infty}\left(\int_{Y^T\Sigma Y+\ell\le sX^2} p_\rho(X,Y)\,dY\right)dX$$

We now examine the integral over Y:

$$\int_{Y^T\Sigma Y+\ell\le sX^2} p_\rho(X,Y)\,dY = \frac{1}{(2\pi)^{\frac{n-2}{2}}}\left|\Delta_b\right|^{-\frac{1}{2}} \int_{Y^T\Sigma Y+\ell\le sX^2} e^{-\frac{1}{2}\left[X \quad Y^T\right]\Delta_b^{-1}\left[X \quad Y^T\right]^T} dY$$

We have:

$$\left|\Delta_b\right| = 1 - b^T b$$

We develop the PHMI expression:

$$\frac{1}{(2\pi)^{\frac{n-2}{2}}}\left|\Delta_b\right|^{-\frac{1}{2}} \int_{Y^T \Sigma Y + \ell \le sX^2} e^{-\frac{1}{2}\left[X \quad Y^T\right]\Delta_b^{-1}\left[X \quad Y^T\right]^T} dY$$

$$= \frac{1}{(2\pi)^{\frac{n-2}{2}}\sqrt{1-b^T b}} \int_{Y^T \Sigma Y \le sX^2 - \ell} e^{-\frac{1}{2}\left(Y^T\left(I_{n-2}+\frac{bb^T}{1-b^T b}\right)Y - 2Y^T \frac{b}{1-b^T b}X + \left(1+\frac{b^T b}{1-b^T b}\right)X^2\right)} dY$$

Let us now do the change of variable:

$$Z = Y - \left(I_{n-3} + \frac{bb^T}{1-b^T b}\right)^{-1} \frac{b}{1-b^T b} X = Y - c$$

$$\frac{1}{(2\pi)^{\frac{n-2}{2}}\sqrt{1-b^T b}} \int_{Y^T \Sigma Y \le sX^2 - \ell} e^{-\frac{1}{2}\left(Y^T\left(I_{n-2}+\frac{bb^T}{1-b^T b}\right)Y - 2Y^T \frac{b}{1-b^T b}X + \left(1+\frac{b^T b}{1-b^T b}\right)X^2\right)} dY$$

$$= \frac{1}{(2\pi)^{\frac{n-2}{2}}\sqrt{1-b^T b}} \int_{(Z+c)^T \Sigma (Z+c) \le sX^2 - \ell} e^{-\frac{1}{2}\left(Z^T\left(I_{n-2}+\frac{bb^T}{1-b^T b}\right)Z + X^2\right)} dZ$$

For this step of the calculation, we have used the fact that:

$$\left(I_{n-3} + \frac{bb^T}{1-b^T b}\right)^{-1} b = \left(1 - b^T b\right) b$$

which implies that:

$$\frac{b^T}{\left(1-b^T b\right)}\left(I_{n-3} + \frac{bb^T}{1-b^T b}\right)^{-1} \frac{b}{\left(1-b^T b\right)} = \frac{b^T b}{1-b^T b}$$

So far we have only proceeded by equalities, we now start taking upper bounds. The first upper bound is:

$$\int_{(Z+c)^T \Sigma (Z+c) \le sX^2 - \ell} e^{-\frac{1}{2}\left(Z^T\left(I_{n-2}+\frac{bb^T}{1-b^Tb}\right)Z+X^2\right)} dZ \le \int_{Z^T \Sigma Z \le sX^2 - \ell} e^{-\frac{1}{2}\left(Z^T\left(I_{n-2}+\frac{bb^T}{1-b^Tb}\right)Z+X^2\right)} dZ$$

This step can be easily proven and is very conservative.

We can also write:

$$\int_{Z^T \Sigma Z \le sX^2 - \ell} e^{-\frac{1}{2}\left(Z^T\left(I_{n-2}+\frac{bb^T}{1-b^Tb}\right)Z+X^2\right)} dZ \le \int_{Z^T \Sigma Z \le sX^2 - \ell} e^{-\frac{1}{2}\left(Z^T Z+X^2\right)} dZ$$

This proves the inequality (1.1).

$$\int_{Y^T \Sigma Y + \ell \le sX^2} p_\rho(X,Y)\, dXdY \le \frac{1}{\sqrt{1-b^Tb}} \int_{Y^T \Sigma Y + \ell \le sX^2} p_{\rho=0}(X,Y)\, dXdY$$

This equation shows that the PHMI is $\frac{1}{\sqrt{1-b^Tb}}$ larger than the one predicted by Equation 4.19.

We now provide an overbound of the parameter $B = \frac{1}{\sqrt{1-b^Tb}}$ that can be studied offline.

The expression for b^Tb in the case of the planar fit for a given *u* is:

$$b^Tb = \frac{\rho^T S^{-1} \rho}{s} = \frac{\left(\lambda^T\left(u^2C_0+M\right) - u^2 C_v{}^T\right)\Gamma^T\left(\Gamma\left(u^2C_0+M\right)\Gamma^T\right)^{-1}\Gamma\left(\left(u^2C_0+M\right)\lambda - u^2C_v\right)}{\lambda^T\left(u^2C_0+M\right)\lambda - 2\lambda^T u^2 C_v + u^2\sigma^2_{decorr}}$$

We now look for a bound on this expression for u between u=1 and infinity. To do this, we are going to use the fact that:

$$\Gamma\left(\left(C_0+M\right)\lambda - C_v\right) = 0$$

This allows us to write that:

$$b^T b = \frac{\left(u^2-1\right)^2 \left(\lambda^T C_0 - C_v^{\ T}\right)\Gamma^T \left(\Gamma\left(u^2 C_0 + M\right)\Gamma^T\right)^{-1} \Gamma\left(C_0\lambda - C_v\right)}{\lambda^T \left(u^2 C_0 + M\right)\lambda - 2\lambda^T u^2 C_v + u^2 \sigma_{decorr}^2}$$

As a result, we have:

$$b^T b \le \frac{u^4 \left(\lambda^T C_0 - C_v^{\ T}\right)\Gamma^T \left(\Gamma\left(u^2 C_0 + M\right)\Gamma^T\right)^{-1} \Gamma\left(C_0\lambda - C_v\right)}{\lambda^T \left(u^2 C_0 + M\right)\lambda - 2\lambda^T u^2 C_v + u^2 \sigma_{decorr}^2}$$

We also have:

$$\left(\lambda^T C_0 - C_v^{\ T}\right)\Gamma^T \left(\Gamma\left(u^2 C_0 + M\right)\Gamma^T\right)^{-1} \Gamma\left(C_0\lambda - C_v\right) \le \left(\lambda^T C_0 - C_v^{\ T}\right)\Gamma^T \left(\Gamma\left(u^2 C_0\right)\Gamma^T\right)^{-1} \Gamma\left(C_0\lambda - C_v\right)$$

and:

$$\lambda^T \left(u^2 C_0 + M\right)\lambda - 2\lambda^T u^2 C_v + u^2 \sigma_{decorr}^2 \ge \lambda^T \left(u^2 C_0\right)\lambda - 2\lambda^T u^2 C_v + u^2 \sigma_{decorr}^2$$

Finally:

$$b^T b \le \frac{\left(\lambda^T C_0 - C_v^{\ T}\right)\Gamma^T \left(\Gamma C_0 \Gamma^T\right)^{-1} \Gamma\left(C_0\lambda - C_v\right)}{\lambda^T C_0 \lambda - 2\lambda^T C_v + \sigma_{decorr}^2}$$

This expression is now independent of u. It is now possible to study the actual distribution of B and set the α_n accordingly, such that the true PHMI is below the allocation. It is expected that this correction will have a minimal impact on the inflation factor.

Bibliography

[Angus] John Angus, Raytheon Co. Fullerton, CA, September 2000. Private communication.

[Benedicto] J. Benedicto, P. Michael, J. Ventura-Traveset. "EGNOS, the European Regional Augmentation to GPS and GLONASS," *Proceedings of the Fifth ESA International Workshop on DSP Techniques Applied to Space Communications*, Noorwijk, The Netherlands, September 1998.

[Braasch] M. S. Braasch. "Multipath Effects," *Global Positioning System: Theory and Applications*. B. W. Parkinson, J. Spilker, eds. Washington, D.C.: AIAA, 1996.

[Bradley] P.A. Bradley, M.I. Dick, T. Samardjiev, Lj. R. Cander, I. Stanislawska. "Implementation of the Kriging Technique in Instantaneous Mapping," *Proceedings COST38/PRIME Workshop*, Eindhoven, pp. 41-43, 1994.

[Braff] R. Braff, C. Shively. "GPS Integrity Channel," *Selected Papers on Satellite Based Augmentation Systems*, Vol. III, pp. 258-274, 1986.

[Chao] Y.C. Chao, S. Pullen, P.K. Enge, B.W Parkinson. "Study of WAAS Ionospheric Integrity," *Proceedings of the Institute of Navigation GPS-96*, pp. 781-88, September 1996.

[Chou] H. Chou. *An Adaptive Correction Technique for Differential Global Positioning System*. PhD thesis, Stanford University, 1991.

[Cormier] D. Cormier. "Real Time R_{irreg}," *Raytheon Co. internal document*, September 2002.

[Craig] J.W. Craig. "A New, Simple, and Exact Result for Calculating the Probability of Error for Two-dimensional Signal Constellations," *Proceedings of IEEE Military Communications Conference MILCOM '91*, pp. 571-75, 1991.

[Cressie] N.A.C. Cressie. *Statistics for Spatial Data*, revised edition. John Wiley and Sons, New York, 1993.

[Datta] S. Datta-Barua *et al*. "An Investigation on the LAAS Threat Model," *Proceedings of the Institute of Navigation National Technical Meeting*, San Diego, CA, 2002.

[DeCleene] B. DeCleene. "Defining Pseudorange Integrity – Overbounding," *Proceedings of the Institute of Navigation GPS-00*. Salt Lake City, UT, 2000.

[Dehel] T. Dehel *et al*. "National Satellite Test Bed (NSTB) Observations of the Effects of Ionospheric Storms on a Prototype Wide Area Augmentation System," *Proceedings of the Institute of Navigation National Technical Meeting*, San Diego, CA, 1999.

[Delika] D. Delikaraoglou. "On the Stochastic Modeling of GPS Ionospheric Delays, " Maunuscripta Geodaetica, Vol. 14, pp. 100-109, 1989.

[Enge96] P. Enge *et al*. "Wide Area Augmentation of the Global Positioning System," *Proceedings of the IEEE*, Vol. 84, No. 8, pp. 1063-1088, 1996.

[Enge01] P. Enge, P. Misra. *Global Positionning System: Signals, Measurements, and Performance*. Ganga-Jamuna Press, 2001.

[Gabler] S. Gabler, C. Wolff. "A Quick and Easy Approximation to the Distribution of a Sum of Weighted Chi-square Variables," *Statistische Hefte,* vol. 28, pp. 317-325, 1987.

[Graas] F. van Graas. "Code Noise and Multipath Real Time Estimator," *WAAS Ionospheric Performance Panel Meeting*, June 2000.

[Hansen00] A. Hansen *et al.* "Ionospheric Correlation Analysis for WAAS: Quiet and Stormy," *Proceedings of the Institute of Navigation GPS-00*, 2000.

[Hansen02] A. Hansen. *Tomographic Estimation of the Ionosphere Using Terrestrial GPS Sensors*. PhD thesis, Stanford University, 2002.

[Hargreaves] J. K. Hargreaves. *The Solar-Terrestrial Environment*. Cambridge University Press, 1992

[Hastie01] T. Hastie, R. Tibshirani, J. Friedman. *The Elements of Statistical Learning*. Springer, 2001.

[Hastie02] T. Hastie. *Lecture notes for the Stanford course, "Introduction to Statistical Modeling,"* 2002.

[ITM] J. Blanch, T. Walter, P. Enge. "Ionospheric Threat Model Methodology for WAAS," *Proceedings of the Institute of Navigation Annual Meeting*, Albuquerque, NM, 2001.

[Jan] S. S. Jan, W. Chan, T. Walter, P. Enge. "Matlab Simulation Toolset for SBAS Availability Analysis," *Proceedings of the Institute of Navigation GPS-01*, Salt Lake City, 2001.

[Journel] A. Journel. *Lecture notes for the Stanford course "Geostatistics for spatial phenomena"*, 2002.

[Kee] C. Kee, B.W. Parkinson, and P. Axelrad, "Wide Area Differential GPS," *Navigation, Journal of the U.S. Institute of Navigation*, Vol. 38, No. 2, Summer 1991.

[Klobuchar] J. Klobuchar. "Ionospheric Effects on GPS," *Global Positioning System: Theory and Applications*. B. W. Parkinson, J. Spilker, eds. Washington, D.C.: AIAA, 1996.

[Komjathy] A. L. Komjathy, L. Sparks, T. Mannucci, X. Pi. "Assessment of WAAS Ionospheric Correction Algorithms in the South American Region," *Proceedings of the Institute of Navigation GPS-02*. Portland, OR, 2002.

[Lejeune] R. Lejeune, M. B. El-Arini, S. Ericson. "Comparison of Performance Results of Ionospheric Tomography and Thin Shell Algorithms for an SBAS in Brazil," *Proceedings of the Institute of Navigation National Technical Meeting*, San Diego, CA, 2004.

[McHugh] T. McHugh *et al*. "WAAS Observed Performance During 60 Day Test," *Proceedings of the Institute of Navigation National Technical Meeting*, Anaheim, CA, 2003.

[MOPS] RTCA Special Committee 159, *Minimum Operational Performance Standards for Airborne Equipment Using Global Positioning System/ Wide Area Augmentation System*, RTCA/DO-229C, November 2001.

[Ross] S. Ross. *A First Course in Probability*. Prentice Hall, Fifth Edition, 1998.

[SARPS] Satellite-Based Augmentation System (SBAS) Standards and Recommended Practices (SARPS), International Civil Aviation Organization (ICAO)

[Schempp] T. R. Schempp, A. L. Rubin. "An Application of Gaussian Overbounding for the WAAS Fault Free Frror Analysis," *Proceedings of the Institute of Navigation GPS-02*. Portland, OR, 2002.

[Shallberg] K. Shallberg *et al*. "WAAS Measurement Processing, Reducing the Effects of Multipath," *Proceedings of the Institute of Navigation GPS-01*, Salt Lake City, UT, 2001.

[Shimamura] A. Shimamura. "MSAS Project Status," *Selected Papers on Satellite Based Augmentation Systems*, Vol. VI, pp. 53-60.

[Sparks01] L. Sparks *et al*. "The WAAS Ionospheric Threat Model," *Proceedings of the Beacon Satellite Symposium*, Boston, MA, 2001.

[Sparks03] L. Sparks *et al*. "The Dependence of WAAS Ionospheric Error Bounds upon the Spatial Distribution of GPS Measurements," *Proceedings of the Institute of Navigation National Technical Meeting*, Anaheim, CA, 2003.

[Spilker] J. Spilker. "Tropospheric Effects on GPS". *Global Positioning System: Theory and Applications*. B. W. Parkinson, J. Spilker, eds. Washington, D.C.: AIAA, 1996.

[Stanis] I. Stanislawska *et al*, "The Kriging Method of TEC Instantaneous Mapping," *Advances in Space Research*, in press, 2003.

[Walter97] T. Walter, P. Enge, A. J. Hansen. "A Proposed Integrity Equation for WAAS MOPS," *Proceedings of the Institute of Navigation GPS-97*, pp. 475-84, September 1997.

[Walter00] T. Walter *et al*. "Robust Detection of Ionospheric Irregularities," *Navigation, Journal of the U.S. Institute of Navigation*, Vol. 48, No. 2, Summer 2001.

[Webster] R. Webster, M. Oliver. *Geostatistics for Environmental Scientists*. John Wiley and Sons, New York, 2001.

[Wild] U. Wild *et al*. "Stochastic Properties of the Ionosphere Estimated from GPS Observations," *Proceedings of the Institute of Navigation GPS-90*, pp 411-427, September 1990.

[Wilson] B. D. Wilson *et al*. "Monitoring The Ionosphere Using A Global GPS Network: Applications and Validation," *Proceedings of the Institute of Navigation National Technical Meeting*, Santa Monica, CA, 1997.