An Ionosphere Estimation Algorithm for WAAS Based on Kriging

Juan Blanch, *Stanford University*

ABSTRACT

GPS alone cannot provide the integrity needed for air navigation. Several error sources deteriorate the precision of the position estimate. One of the largest and more unpredictable sources of error for single frequency users is the ionosphere. For this reason, ionospheric behavior drives the performance of the Wide Area Augmentation System (WAAS). At any given time, the only information we have of the ionosphere is a limited amount of Total Electron Content (TEC) measurements. As a consequence, in order to estimate the ionospheric delay and get a confidence bound on such an estimate, we need to understand the spatial structure of the ionosphere over the region of interest. Using the thin shell model framework, where each TEC measurement is identified as a location on the thin shell, labeled the Ionospheric Pierce Point (IPP), the problem is reduced to a 2-dimensional problem. Once we have a good description of a nominal ionosphere, there are two questions that need to be answered before estimating the delay at a given IPP: Are the IPP measurements compatible with the assumed nominal model of the ionosphere? How relevant are the IPP measurements to the location we need to estimate? To answer the first one, an accurate characterization of the ionosphere in nominal conditions is needed. The large observed stationarity violations make this latter question very difficult.

A worst case based method to determine the spatial structure of the nominal ionosphere in terms of the variogram, or, equivalently, the covariance is presented. The technique called 'kriging' produces at each location an estimate and a confidence bound on the estimate, the kriging variance. The particular behavior of the kriging variance at the edge of coverage allows us to intuitively define the 'well sampled' region. We show that a carefully designed estimation algorithm based on kriging could provide confidence bounds on the ionospheric delay corrections allowing WAAS to meet the GNSS Landing System requirements.

INTRODUCTION

WAAS corrects, among other sources of error, for the user's ionospheric delay errors and places strict confidences on those corrections under all conditions [1]. The estimation process is simplified by making the thin shell approximation [2]. Each ray path has a corresponding Ionospheric Pierce Point (IPP), where that path intersects the thin shell. Each ionospheric TEC measurement is transformed in a Vertical Ionospheric Delay. The WAAS correction message specifies the vertical delay values as well as the confidence at each node, labeled σ_{GIVE} [3]. The user interpolates these values to find the ionospheric delays corresponding to its satellites in view. The current algorithm is described in [4]. In order to increase the performance of the system, in particular to be able to meet GNSS Landing System requirements, the confidence bound on the correction would need to be reduced by more than a factor of two: from a σ_{GIVE} of 1.5 m to .6 m.

In fact, under nominal conditions, any reasonable estimation algorithm yields excellent estimates. Unfortunately, the existence of sudden changes in the ionosphere, from quiet to stormy conditions [5], [6], [7], together with the fact that the ionosphere is irregularly sampled, has forced the system to be overly conservative most of the time. A storm detector based on the chi-square test, and a metric measuring the degree of coverage take care of the irregularity threats [4], [8]. It is mainly in the characterization of the undersampled regions that there is room for improvement.

It was shown in [9] that the technique called kriging –a particular minimum mean-square estimator – could mitigate the undersampled problem. Kriging has long been used for mapping purposes, because it provides a natural interpolation. As such it has been applied to map TEC measurements as well as other ionospheric parameters [10], [11]. Kriging has also many attractive features for WAAS [9], for both the definition of the 'well sampled' region and the confidence bound generation. However, a critical step in the estimation when using kriging is the knowledge of the underlying deterministic and random spatial structure [12].

In this work, we are interested in the limitations of ionospheric delay estimation due to ionospheric behavior and due to the location of the IPP measurements. We will therefore omit the difficulties introduced by the bandwidth limitation [4] by assuming that the user has the same information as the master station, that is, all the IPP locations and measurements. For the validation of the proposed solutions, we will also consider that the measurements have very low noise, as it is the case in WAAS 'supertruth' data. In this set of data, the redundancy of receivers and the post-processing allows the isolation and removal of faults [5]. However, it will be explained how to include the higher level of noise that real time measurements have.

In the first part, we will show a method to refine the spatial characterization of the ionosphere in nominal conditions, and come up with a conservative model. In the second part we will describe the algorithm that is optimal under the characterization given in the first part; we will also show how to heuristically generate the 'well sampled' region from the kriging variance. We finally will present the cross validation results over several days worth of supertruth data, which includes the most extreme ionospheric conditions observed during the current solar cycle [7].

SPATIAL STRUCTURE OF THE VERTICAL IONOSPHERIC DELAY

A useful and intuitive way of characterizing the spatial structure of a random field is the variogram [13]. The variogram measures the loss of confidence of a random function as we depart from the measurements, assuming that there is no deterministic underlying trend, and that the random function is stationary. It is defined as:

$$\gamma\left(|h|\right)=\frac{1}{2}E\left\{\left(I\left(u+h\right)-I\left(u\right)\right)^2\right\} \qquad (1)$$

where u is a location, h is a vector and I is the random function that we want to estimate; in our case, it is the vertical ionospheric delay. At each epoch the variogram was estimated by computing all the difference between pairs of measurements $I_v(u_i)$-$I_v(u_j)$ and the corresponding distance $|u_i-u_j|$ [5], [9]. The pairs were classified according to their distance in different categories. In this study we used the bins [0 300 km], [300 km 600 km] etc. For each bin we estimated the variance of the resulting empirical distribution. Since we wanted to get a conservative estimate of the variance, we used the gaussian overbound [14] of the empirical distribution.

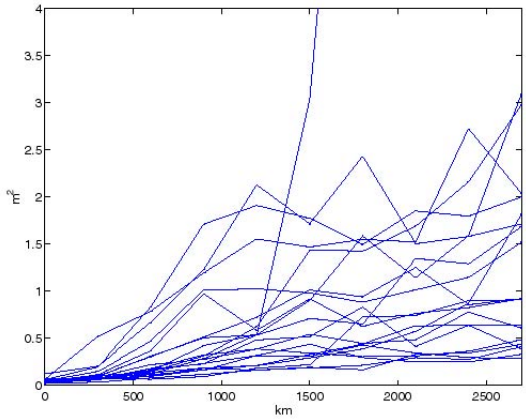

Figure 1. *Variograms for a quiet day*

Figure 1. shows several experimental variograms for a quiet day. The flat variograms correspond to the ionosphere during the night, and the more variable correspond to daytime. In fact, the ionospheric behavior is well approximated by a planar trend [5]. It is the planar trend that is responsible for most of the variability, so it does not correspond to random variability. We can have a better idea of the range of the trend by examining a variogram obtained through the classical formula –instead of computing the gaussian overbound in each bin, we compute the mean of the squares- [9], [12] which is smoother than the variogram obtained through the gaussian overbound. Figure 2 shows such a variogram.

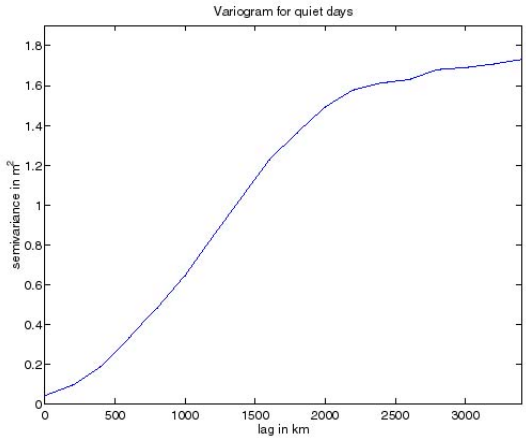

Figure 2. *Variogram for a quiet day obtained with the classical formula.*

Although we cannot use this variogram to bound the ionospheric behavior, we can extract some qualitative information. The overall shape is determined by the presence of a trend, which produces the parabolic behavior from 0 to 2000km. However, we see that for lags larger than 2500 km, the planar trend is not as clear. The other very important characteristic of the variogram is the behavior at the origin [12]. In this variogram, at small lags

we have a non zero derivative. This means that even if we remove the planar trend there is random spatial structure left [12], [13]. We can express this idea by writing that, locally, the vertical ionosphere delay is given by:

$$I_v(x,y) = a_0 + a_1 x + a_2 y + R(x,y) \qquad (2)$$

where the first three terms define the planar trend and R(x,y) is a random function with zero mean and a given variogram. Now, we would like to characterize the random behavior of R, that is, to find a variogram that describes it conservatively.

De-trending the data is a classical problem in spatial statistics [15]. It is problematic to just fit a plane to the data and compute the variogram of the residuals. Some measurements might have high leverage on the fit, either due to their large value or their geometry [16], 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 if such outliers do exist, and we 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 [12], [15]. This way we ensure that the original data is not affected by the fit. For our purposes, there are two problems in this approach. First, we end up with a very small amount of measurements, thus getting less 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 decided to proceed in the following way: for each pair of measurements, we did a planar fit on the remaining measurements (up to a certain radius) and assumed this planar fit would describe the trend. 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 correctly the trend. However, this is likely to bias the variogram in a conservative way. We mitigated this effect by considering only the densest regions in terms of IPP measurements.

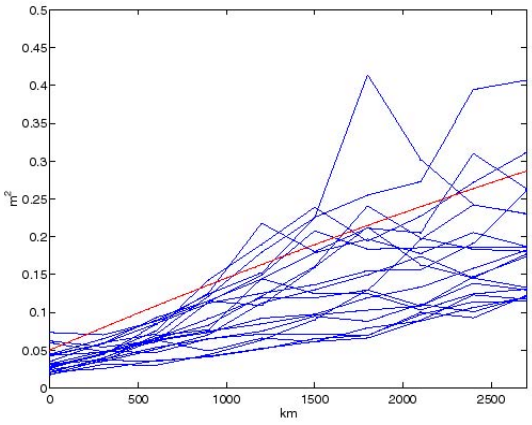

Figure 3. *Experimental residual variograms for a quiet day.*

Figure 3 shows a series of experimental residual variograms generated using this method. We used only pairs of measurements with latitudes 30 deg and 45 deg, and longitudes between -105 deg and -85 deg. The radius of search used was 1500 km. The nominal day used was July 2nd, 2000. If the random function R had no spatial dependency, we would expect those variograms to be flat. We see instead that there is a clear spatial structure once the trend is removed.

Since we are going to be using the variogram for estimation, we need to find an admissible variogram [11], [13] that is close enough to the experimental variogram. The simplest we can find is a linear model with a non zero intercept at the origin. However, such a model cannot be described by a covariance, because the variogram has no finite limit as the separation becomes infinite [12] –we will see later that we need the covariance. We chose instead a model variogram with a linear behavior at the origin and a finite limit at infinity –a sill. An admissible variogram having these properties is given by:

$$\gamma(h) = c\left(1 - e^{-\frac{h}{a}}\right) + \nu \;,\; \gamma(0) = 0 \qquad (3)$$

This model is called exponential model [13]. For a model variogram we always have by definition $\gamma(0)=0$. So there is a discontinuity at the origin, which is called the nugget effect [12]. The different parameters were chosen visually so that the function given by (3) would approximately overbound the different experimental variograms for a nominal day. The parameters retained were: $c=1m^2$, $a=10000km$, and $\nu=.05\ m^2$.

Because of the existence of a sill in this variogram model, it is equivalent to say that the residuals have the covariance [12], [13]:

$$C(h) = c + \nu - \gamma(h) = ce^{-\frac{h}{a}},\ C(0) = c + \nu \qquad (4)$$

We now have a model of the ionosphere under nominal conditions.

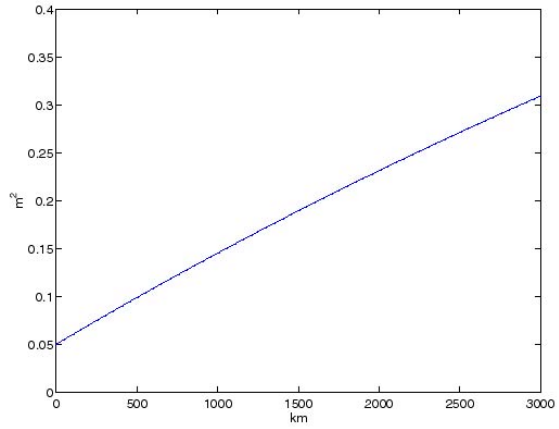

Figure 4. *Model variogram for nominal ionosphere.*

If the ionosphere was stationary over the CONUS region at all times, even in storm conditions, we would not need any additional characterization –except for the temporal variability, a problem not treated in this study. The extreme ionospheric conditions observed during the current solar cycle show that there can exist sharp limits between regions of the ionosphere that are quiet and regions that are stormy [6], [17]. For example, during the first hours of July 16th, 2000, for the IPPs above 37 deg latitude, the variance is 10 cm, whereas below the variance is 100 times larger. There is no stationary model that can describe this case without being overly pessimistic in the quiet region. Instead, we need to evaluate, given a set of IPPs compatible with a spatial structure, which locations will be well described by the same spatial structure. In [6] it was suggested that in the regions well surrounded by IPP measurements, that is, well inside the convex hull of the measurements, we could safely assume we had the correct spatial structure. While being easy to assess visually, this is a difficult parameter to compute. The uncertainty maps associated with kriging appear to naturally account for insufficient sampling [9]. Where the coverage in IPPs is low the kriging variance is high. We can therefore use these maps to assess the degree of coverage. We will come back to this problem after explaining how to generate the map of variances.

ESTIMATION ALGORITHM

Now that we have a statistical description of the nominal ionosphere behavior, we can shape the estimation algorithm. The first step of is to make sure the measurements are compatible with the assumed model. The measurements are assumed to be of the form:

$$I_{v,meas}(x_k, y_k) = a_0 + a_1 x_k + a_2 y_k + R(x_k, y_k) + N_k \qquad (5)$$

with N_k the measurement noise, for which the covariance is well known [4]. Now let us assume that both R and N are multivariate normal, and that R and N are independent. From the measurements we can form a quantity that will be chi-square distributed with n-3 degrees of freedom [16]. To do that, we filter the deterministic component of (5), i.e., we find combinations of the measurements that will have zero mean. We then decorrelate those combinations so that we have a sum of independent gaussian random variables.

We can then design a test, in the same way it was done in [4]: given the degrees of freedom and an allowable false alarm rate, P_{fa} we can calculate a threshold value for the chi-square using its known distribution. Using the same failure model as in [4], which is that the variogram increases uniformly by the same factor, we obtain that if we pass the chi-square then we need to increase the final confidence bound by:

$$R^2_{irreg}(P_{fa}, P_{md}) = \frac{\chi^2_{1-P_{fa}}}{\chi^2_{P_{md}}} \qquad (6)$$

where P_{md} is the allowable misdetection rate. The details of the rationale can be found in [4]. We took $P_{md}=10^{-3}$ and $P_{fa}=10^{-3}$. What is important is that we have a test that evaluates whether the measurements are compatible with the assumed model, and that indicates by how much we need to inflate the confidence bound as a function of the number of measurements.

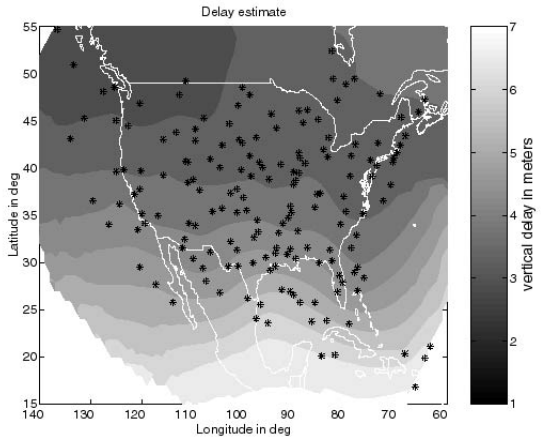

Figure 5. *Kriging map over CONUS of the vertical ionospheric delay for July 2nd 2000.*

We now need to give the best estimate and generate the confidence bound on the estimate, given the assumed model (the model that passed the chi-square test). We will use here the technique known as kriging with a trend (KT), or universal kriging [12], [13], which minimizes the mean-square estimation error given that the spatial structure is given by (2). In fact, the only change with respect to ordinary kriging [9],[12] is the addition of two constraints expressing the fact that there is a deterministic trend described by two additional parameters. The equations for KT are:

$$\hat{I}_v(u_0) = \left[I_{v,meas}(u_1) \quad \ldots \quad I_{v,meas}(u_n) \right] \lambda \tag{7}$$

$$\sigma^2_{kriging} = C(\infty) - b\lambda \tag{8}$$

where λ is the solution of the equation:

$$\begin{bmatrix} C(|u_i - u_j|) + V & G \\ G^T & 0 \end{bmatrix} \begin{bmatrix} \lambda \\ \mu \end{bmatrix} = b \tag{9}$$

with $G = \begin{bmatrix} 1 & x_1 & y_1 \\ \vdots & \vdots & \vdots \\ 1 & x_n & y_n \end{bmatrix}$ and $b = \begin{bmatrix} C(|u_0 - u_1|) \\ \vdots \\ C(|u_0 - u_n|) \\ 1 \\ x_0 \\ y_0 \end{bmatrix}$

As in ordinary kriging, the estimation produces a kriging variance $\sigma^2_{kriging}$, and we take the confidence bound to be the square root of the kriging variance. Figure 6 shows a map of the confidence bound generated by KT (where we have not yet multiplied by R_{irreg}).

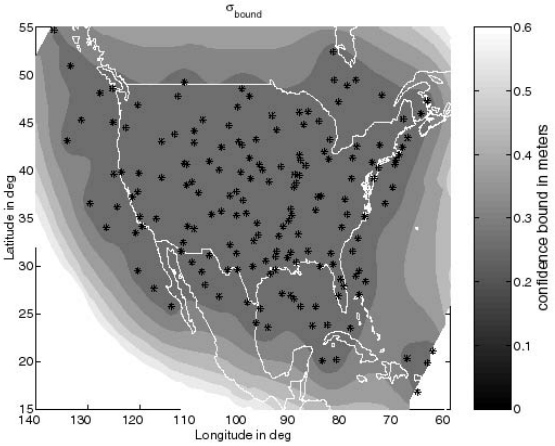

Figure 6. *Map of kriging variance for July 2nd 2000.*

The shape of the kriging confidence bound suggests a way of generating the stationary region. Outside the convex hull of the measurements the confidence bound grows very quickly. We can for example decide that all the locations that have a kriging confidence bound above a certain threshold are not estimable, because we do not have enough confidence on the underlying structure at that location. If we take for example a threshold of .30 m, the well sampled region will be as shown in Figure 7.

Summarizing, the estimation algorithm does the following. First, we find all the IPP measurements within a radius of 2000 km. Then, we check that the measurements are compatible with the assumed model. We apply KT to obtain the kriging estimate and the kriging confidence bound; compare the kriging confidence bound to the threshold above which we have decided not to monitor and, if it passes, we multiply the kriging confidence bound by R_{irreg}.

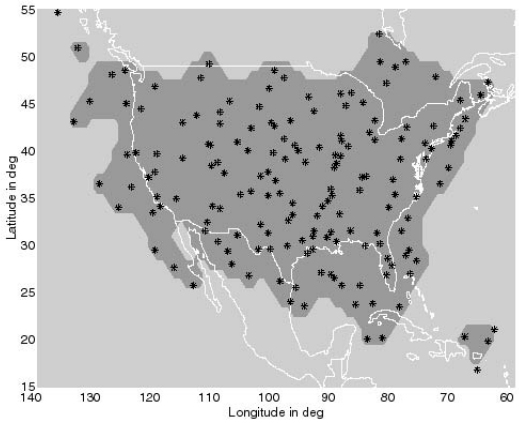

Figure 7. *Covered region using the kriging variance. The darker gray indicates that the region is well sampled.*

CROSS-VALIDATION RESULTS

There are two different aspects of the algorithm that need to be evaluated: the integrity and the availability both in terms of coverage and level of accuracy. For the coverage, Figure 7 shows that the metric based on the kriging variance covers the CONUS region. The most important is the integrity. We need to make sure that the true ionospheric delay experienced by the user falls within the bounds sent. This was checked using cross-validation: for each measurement we compute the estimate and the confidence bound produced by the algorithm at the same location, excluding the test measurement. The normalized residuals were then formed:

$$K = \frac{\hat{I}_v - I_{v,meas}}{\sqrt{\sigma^2_{bound} + \sigma^2_{meas}}} \quad (10)$$

In the denominator we have the difference between the predicted vertical ionospheric delay and the true vertical ionospheric delay. σ^2_{bound} is either the kriging variance or the kriging variance corrected by R^2_{irreg} (we examined both) and σ^2_{meas} is the measurement noise, which here is very low because we are using supertruth data. If the predicted delays and bounding variances were safe, the resulting distribution should be overbounded by a zero mean unit-variance gaussian, that is, the empirical probability of having a residual larger than a given threshold should always be below the probability given by the normal distribution.

Figure 8 shows the normalized residuals for July 2nd 2000. Here the residuals were normalized by the kriging variance only. We can see that the confidence bound is very conservative, even before taking into account R^2_{irreg}. Instead of displaying every distribution for every day we computed some important characteristics. The most interesting are the maximum residual and the gaussian overbound. These quantities were computed for the following days: April 6th-8th 2000, July 15th-16th 2000 for the storm days, February 12th 2001 for a mildly disturbed day and May 25th 2001, July 2nd 2000 and March 26th 2000 for nominal days. They include the worst ionospheric conditions experienced in the current solar cycle.

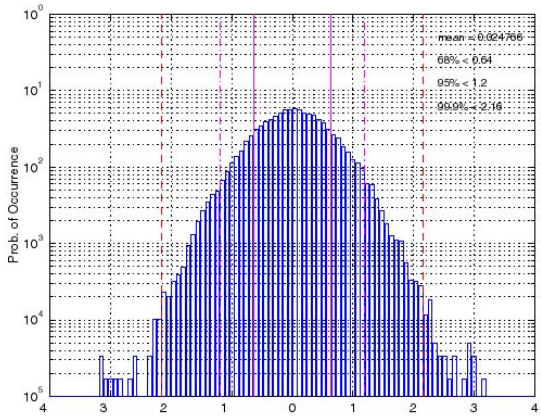

Figure 8. *Histogram of results for July 2nd 2000, before multiplying the confidence bound by R_{irreg}.*

	Mean of σ_{bound}	Gauss. overbound	Max. norm. residual
July 15th -16th	.28 m	1.25	5.13
April 6th-8th	.28 m	1.22	4.63
February 12th	.27 m	1.2	4.9
May 25th	.27 m	1.05	4.45
Mar 26th	.27 m	1.05	4.38
July 2nd	.27 m	.74	3.18

Table 1. *Result of the cross-validation for several quiet days, before multiplying by R_{irreg}*

In Tables 1 and 2, the first column represents the mean of the confidence bounds, the second column gives the

standard deviation of the gaussian overbound -of the normalized residual distribution-, and the third gives the magnitude of the maximum residual.

Table 1 shows the results using only the kriging variance. While being enough on July 2nd, there is not enough protection against irregularities in storm days: the gaussian overbounds are larger than 1. This is due to the fact that our confidence on the model depends on the number of measurements. The loss of confidence is taken into account by $R_{irreg.}$.

	Mean of σ_{bound}	Gauss. overbound	Max norm. res.
July 15th –16th	.46 m	.74	3.07
April 6th-8th	.46 m	.73	2.77
February 12th	.47 m	.74	3.08
May 25th	.46 m	.65	2.78
Mar 26th	.45 m	.63	2.63
July 2nd	.45 m	.50	2.07

Table 2. *Result of the cross-validation for several quiet days, after multiplying by* R_{irreg}

Table 2 shows how we are protecting against irregularities: the gaussian overbound is always well below one and the residuals are never above 3.08. We also should point out that most of the time the ionosphere behaves as in July 2nd 2000. The level of accuracy is below .5 m.

Integrity failures are more likely to happen when there are few measurements. We tested the algorithm under two different data deprivation schemes that correspond to two generic threats [5]: an isolated irregularity in an otherwise quiet region, and an irregularity coming from the edge of coverage. For both deprivation schemes we used a quiet day, July 2nd 2000, and a storm day July 16th 2000. In the first one, we implemented a disk data deprivation scheme: in the estimation process we used only the measurements that are at a distance larger than a given radius. The cross-validation process was carried out for radii going from 0 to 600 km. Figure 9 shows the gaussian overbound both for a quiet day and a storm day as a function of the excluded disk radius. When the gaussian overbound is 0, it means that under this deprivation scheme, all kriging confidence bounds were above .3 m, i.e., the IPP was not well sampled. The fact that the gaussian overbound does not grow, means that we are correctly taking into account the loss of information as we have less IPP measurements.

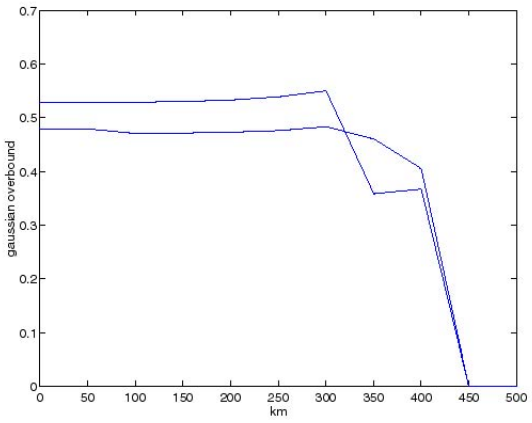

Figure 9. *Gaussian overbound of the normalized residuals as a function of excluded disk radius, for a quiet day and for a storm day.*

In the second deprivation scheme, all IPPs that where under an East West line at a distance D from the location to be estimated were excluded. This deprivation scheme was designed to put the algorithm under severe undetected gradients, like the one observed on July 16th, 2000 [6], [7]. We varied D from –200 km to 800 km. For negative distances we are in fact out of the convex hull of the IPP measurements. For this reason, at –200 km, there are no points for which the kriging confidence bound is under .3 m.

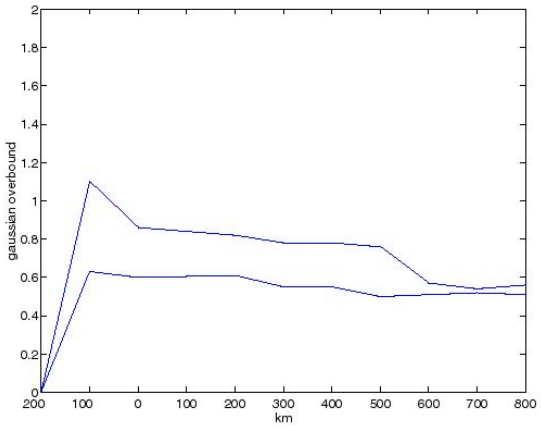

Figure 10. *Gaussian overbound of the normalized residuals as a function of the distance to the uncovered region.*

The worst case corresponds to –100 km, where the overbound is above 1. This is due to the huge gradient experienced on July 16th 2000 mentioned above. We notice however that the effect is small, and there is no real integrity threat: the maximum residual was 4.5. We can conclude that, together, the chi-square test and the kriging variance protect efficiently against irregularities.

The algorithm presented here can be modified in several ways. The most obvious is to have several models of the ionosphere describing different ionosphere conditions instead of one. For example, if the tightest model did not pass the chi-square test, we could test a covariance structure representing a more disturbed ionosphere. It would be specially useful on mildly disturbed days.

CONCLUSION

We have refined the characterization of the nominal ionosphere over the CONUS region within the thin shell model. We have designed an algorithm that is optimal in the least squares sense under the given characterization. A storm detector makes sure the measurements are compatible with the model, and the kriging variance metric defines the coverage through a threshold. The results show that it is possible to design an ionosphere estimation algorithm for WAAS providing safe confidence bounds under .5 m.

ACKNOWLEDGEMENTS

This work was sponsored by the FAA Product Team (AND-730). The author would like to thank Eric Altschuler at Raytheon and Larry Sparks, Xiaoqing Pi and Tony Manucci at JPL for providing the supertruth data. The author would also like to thank Prof. Andre Journel for his guidance on the use of spatial statistics and Todd Walter for helpful discussions.

REFERENCES

[1] Enge P.,Walter T., Pullen S., Kee C., Chao Y.C., and Tsai Y.-J, ”Wide Area Augmentation of the Global Positioning System,” Proceedings of the IEEE, vol. 84, no 8, pp. 1063-1088,1996.

[2] Global Positioning System Standard Positioning Service Signal Specification, June 1995.

[3] WAAS MOPS, RTCA SC 159 DO-229B.

[4] Walter T., Hansen, A., Blanch, J., and Enge P., ”Robust Detection of Ionospheric Irregularities,” in proceedings of ION GPS, Salt Lake City, UT, September 2000.

[5] Hansen, A., Walter T., Blanch, J., and Enge P., ”Ionospheric Correlation Analysis for WAAS: Quiet and Stormy,” in proceedings of ION GPS, Salt Lake City, UT, September 2000.

[6] Blanch J., Walter T., and Enge P., ”Ionospheric Threat Model Methodology for WAAS,” in proceedings of ION 57th NTM, Albuquerque, NM, June 2001.

[7] Coster, Anthea J., Foster, J.C., Erickson P.J. and Rich F.J., “Regional Mapping of Storm Enhanced Density during the July 15-16 2000 Geomagnetic Storm,” Proceedings of the Beacon Satellite Symposium, Boston College June 6, 2001.

[8] Sparks L., Mannucci A.J., Altshuler E., Fries R., Walter T., Hansen A., Blanch J., Enge P. “The WAAS Ionospheric Threat Model,” in proceedings of the Beacon Satellite Symposium, Boston, MA, June 2001.

[9] Blanch J., Walter T., and Enge P., ”Application of Spatial Statistics to Ionosphere Estimation for WAAS,” in proceedings of ION NTM 2002, San Diego, NM, January 2002.

[10] Stanislawska, I., Junchnikowski, G., Cander Lj. R., "The kriging method of ionospheric parameter foF2 instantaneous mapping," Annali di Geofisica, 39(4), 845-852, 1996.

[11] Stanislawska, I., Junchnikowski, G., Zbyszynski, Z., "Generation of instantaneous maps of ionospheric characteristics," Radio Science, 36(5), 1073-1081, 2001.

[12] Andre Journel, Lecture notes for the Stanford class "Geostatistics for Spatial Phenomena", 2002.

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

[14] DeCleene B.,"Defining Pseudorange Integrity – Overbounding" in proceedings of ION GPS, Salt Lake City, UT, September 2000.

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

[16] Hastie T., lecture notes for the class "Linear Models", 2001.

[17] Datta-Barua S., Walter T., Pullen S., Enge P.," An Investigation on the LAAS Threat Model." in proceedings of ION NTM 2002, San Diego, NM, January 2002.