

Ionospheric Correction Using Tomography *

Andrew J. Hansen

Todd Walter

Per Enge

Stanford University

ABSTRACT

The most demanding application of wide area differential corrections to GPS is vertical positioning of aircraft on precision approach. Here the Wide Area Augmentation System (WAAS) combines accuracy requirements on the order of ones of meters with safety of life integrity requirements which specify that any vertical position errors greater than the Vertical Protection Limit be enunciated to the flight crew within six seconds. The ionosphere is the foremost impediment to satisfying these requirements.

Stanford, as a member of the National Satellite Test Bed (NSTB), is developing techniques for estimating the ionosphere in real-time. Previous research has established a connection between ionospheric error and vertical positioning error within the framework of modal decomposition. Ionospheric tomography is a natural extension of modal decomposition to the estimation of the ionosphere's three-dimensional electron density.

We present a tomographic estimation algorithm and its implementation over the NSTB network. This estimator supplies not only corrections to the user but also appropriate confidence information for predicting the accuracy of those corrections in the aircraft. The tomographic approach to ionospheric correction obviates the troublesome obliquity factor associated with typical gridded vertical delay algorithms.

The capability of ionospheric tomography is demonstrated by a time series of 3D electron density reconstructions over the Conterminous United States (CONUS). The accuracy, integrity, and availability afforded the user by this approach is quantified through application on live NSTB observations.

INTRODUCTION

The ionosphere induces a propagation delay over and above the free space delay on radiowaves propagating through the associated plasma. Although this is nuisance parameter for ranging applications, at L-band the additional delay is well modeled as a linear operator. In addition, since the ionosphere is a dispersive medium this delay can be observed using radiowaves at two different carrier frequencies. In the case of GPS dual frequency receivers a direct measure of the ionospheric delay along the line of sight to each GPS satellite in view can be made. The premise of the Wide Area Augmentation System (WAAS) ionospheric correction process is to form an estimate of the ionosphere using such measurements [1]. The resulting estimate may then be transmitted to single frequency receivers in order to correct the ionospheric range error.

We have developed two distinct ionospheric estimators for use in the WAAS correction process. The first is comprised of a two-dimensional grid over latitude and longitude which models the ionosphere as a thin shell fixed at an altitude in the neighborhood of 350 (km) [2]. Vertical delay is estimated at each grid point from dual frequency reference station (TRS) measurements. The single frequency receiver then employs a fixed function of elevation, the so-called obliquity factor, to map an interpolated vertical delay into a range delay prediction along the line-of-sight.

The second estimator centers around the tomographic inversion of the linear phase delay operator which is simply the Radon transform. Using the known TRS and satellite locations the three-dimensional phase delay operator is an observation matrix formed via some basis. The ionospheric electron density estimate is the inner product of the "inverted" observation matrix and the TRS measurements. A single frequency receiver can then predict its ionospheric range error by dotting the electron density estimate into its own ionospheric observation matrix.

*Supported by FAA Grant 95-6-005.

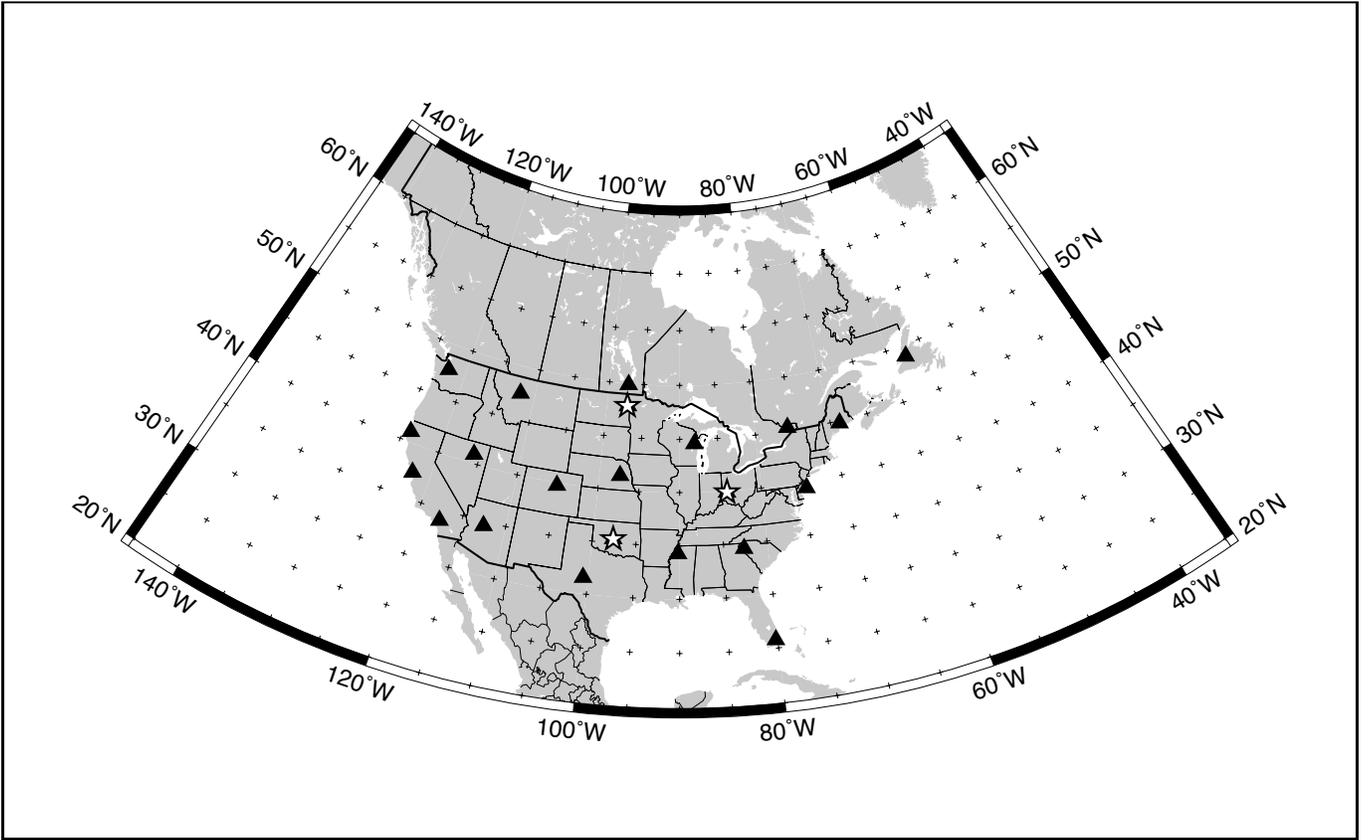


Figure 1: The NSTB reference network geometry is comprised of 23 dual frequency receivers connected in real-time to a T1 backbone. Stations marked with triangles were used as reference observers and those with stars were monitors not incorporated into the WAAS correction

Below we give a description of the tomographic inversion process and identify it as an extension of the modal decomposition concept presented in previous work [3]. The reader is referred to the literature [2, 4, 5] for an exposition of grid estimation techniques. A quantitative comparison between Stanford's grid and tomographic implementations is made on live NSTB observations in both the pseudo-range and position domain for a static user within the context of accuracy, integrity, and availability metrics. We close with some comments on immediate implications and future work.

PROBLEM FORMULATION

Radiowave propagation delay is governed by the phase relation

$$\Delta\Phi = -2\pi\frac{f}{c} \int_{R(r)}^{SV(r)} n(r, f, t) dl(r) \quad (1)$$

where f is the frequency, c is the speed of light, SV is the transmitter location, R is the receiver location, n is the index of refraction, and r is a four dimensional

position vector (x, y, z, t) . For the purpose of L-band ranging (1) is a linear relation since the phase path is very nearly the line of sight between the satellite (SV) and receiver (R) [6]. The effect of the ionosphere is captured in the index of refraction, $n(r, f)$, which is a function of both radiowave frequency and position along the phase path. The full expression for the index of refraction in a plasma such as the ionosphere is given by the Appleton-Hartree equation [7]. We will utilize the standard L-band approximation [8]

$$n(r, f, t) = 1 - \frac{e^2 N_e(r)}{4\pi\epsilon_0 m f} + \mathcal{O}(1/f^3) \quad (2)$$

to the Appleton-Hartree equation, where e is the charge on an electron, $N_e(r)$ is the local electron density, ϵ_0 is the permittivity of freespace, and m is the rest mass of an electron. Substituting (2) into (1) and subtracting out the free space delay

$$\Delta\Phi = \frac{\kappa}{f} \int_{R(r)}^{SV(r)} N_e(r) dl(r) \quad (3)$$

where κ is a constant, we are left with an expression for the phase advance induced on the carrier by the ionosphere.

If we instead insert the group index of refraction

$$\begin{aligned} n'(r, f) &= \frac{\partial n f}{\partial f} \\ &= n + f \frac{\partial n}{\partial f} \\ &= 1 + \frac{e^2 N_e(r)}{4\pi\epsilon_0 m f} + \mathcal{O}(1/f^3) \end{aligned} \quad (4)$$

into (1) we find a similar expression for the corresponding code phase delay.

$$\Delta\Theta = -\frac{\kappa}{f} \int_{R(r)}^{SV(r)} N_e(r) dl(r) \quad (5)$$

where Θ is the code phase.

Since GPS is a ranging application it is convenient to cast (3) and (5) in terms of time rather than phase

$$\Delta\phi = -\frac{K}{f^2} \int_{R(r)}^{SV(r)} N_e(r) dl(r) \quad (6)$$

and

$$\Delta pr = \frac{K}{f^2} \int_{R(r)}^{SV(r)} N_e(r) dl(r) \quad (7)$$

where ϕ is carrier, pr is code, and K is the appropriate constant. To first order, the code delay is equal but opposite the carrier phase advance.

The frequency dependence of the phase delay equation defines the dispersion relation which intern provides a measurement mechanism. The total electron content (TEC) along the line of sight can be observed by any receiver with sufficient frequency diversity. In the case of L_1/L_2 dual frequency TRSs the code delay observation equation is

$$\begin{aligned} \text{TEC} &= \int_{R(r)}^{SV(r)} N_e(r) dl(r) \\ &= \frac{f_{L_1} f_{L_2} (pr_{L_1} - pr_{L_2})}{K(f_{L_2}^2 - f_{L_1}^2)} \end{aligned} \quad (8)$$

which can be carrier smoothed [9] in real-time. Conversely the TEC induced delay is a ranging error source for single frequency users since it is unknown.

The task at hand is to make dual frequency TEC measurements at TRS locations shown in Figure 1 and combine them in real-time to form an estimate of the electron density and an appropriate confidence interval on that estimate. The estimate and confidence is then transmitted to the user's receiver to correct the TEC induced error on each single frequency pseudorange measurement.

There are three key points in our application of tomographic inversion to ionospheric estimation. The first is the equivalence of the observation equation (8) to the Radon transform expressing a linear operator. The second is the understanding that this inverse problem is under-determined. The third is the appreciation of both $N_e(r)$ and TEC as random fields.

TOMOGRAPHIC INVERSION

Application of tomography to the field of ionospheric estimation is relatively new, with first appearance in the literature around 1988 [10, 11, 12]. A general survey of the static two-dimensional tomographic implementations to date can be found in [13]. Recently, time dependent ionospheric tomography has been applied in three dimension on simulated data [14]. Here we present an implementation of real-time three-dimensional tomographic inversion applied to live GPS observations.

As a point of departure we define the operator, $G : N_e(r) \rightarrow \text{TEC}$, we wish to invert

$$G \equiv \int_{R(r)}^{SV(r)} (\cdot) dl(r) \quad (9)$$

from the observation equation. The operator has finite dimension $\text{TEC} \subseteq \mathcal{R}(G)$ in ‘‘row’’ space given a fixed number of TRS/SVs and presumably infinite dimension $\mathcal{D}(G) \subseteq N_e(r)$ in ‘‘column’’ space.

From here we need to cast the space of the electron density. The physical ionosphere is a three dimensional distribution singularly valued in time. As such it has a well defined inner product and a bounded norm implying that $N_e(r)$ is a Hilbert space. This affords us tremendous theoretical leverage most importantly the *Riesz Representation Theorem* and the *Decomposition Theorem*, definitions and proofs for which the reader is referenced to the literature [15, 16]. The implication of the former is that the electron density may be transformed to any basis which spans the space of the bounding linear functional. The latter provides the machinery for truncating the infinite basis set needed to span the domain of G to a finite dimensional basis suitable for computer implementation.

Presuming there is no closed analytic expression for $N_e(r)$ from which to form a complete basis (a pretty safe assumption), a necessary step for computer implementation is to expand the column space of G . There are many important considerations when choosing a basis. From the operational perspective we feel the most significant are the mathematical framework necessary to quantify the confidence interval rigorously,

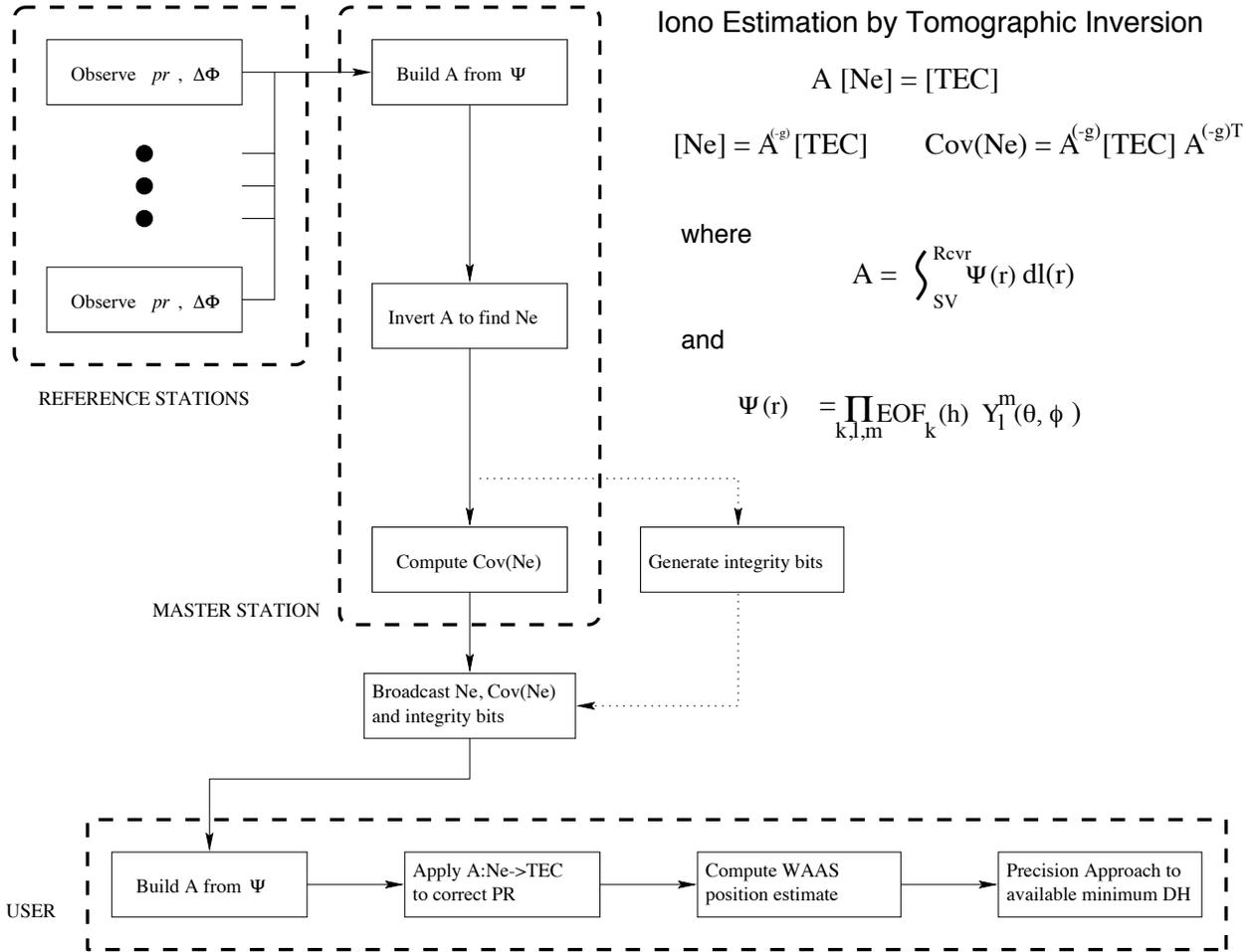


Figure 2: The tomographic TEC delay correction process has three main sections: reference observations, ionospheric estimation, user range correction.

transmission bandwidth efficiency, and bounded versus global support which defines the sparsity of G . Secondary considerations include functional form and codification effort versus computational effort.

This being said we have selected spherical harmonics, $(Y_l^m(\theta, \phi))$ in the latitude and longitude dimensions and empirical orthogonal functions, $\Gamma(h)$, in the radial dimension for the first implementation leaving a fuller evaluation to subsequent work. The complete basis, $\Psi(h, \theta, \phi)$, is then the series of products

$$\Psi_{klm}(h, \theta, \phi) = \prod_{k,l,m=0}^{\infty} \Gamma_k(h) Y_l^m(\theta, \phi) \quad (10)$$

which has global support hence G is dense.

We now make our first assumption, namely that the ionosphere has a red power spectrum [3, 14], and invoke the second theorem to truncate $\Psi(h, \theta, \phi)$ to finite order. For spectral bases such as $\Psi(h, \theta, \phi)$ the cutoff

amounts to a Nyquist criteria on the wave number of the latitude/longitude and height functionals. This bound is deterministic in the geometry of the TRS network/SV constellation and sets a lower bound for the basis order.

The viewing geometry of a ground reference network is inherently weaker at resolving vertical variations than those in the horizontal since the line of sight ray is never perpendicular to the vertical dimension. Indeed there is a theoretic upper bound on the wave number any ground based estimator can attain in the vertical dimension. On the contrary, the horizontal wave number bound can be increased indefinitely simply by adding more TRSs. With this in mind, we treat $\Psi(h, \theta, \phi)$ as a finite basis in $K \times L \times M$ functions for a fixed observation network but defer the specification of K, L, M to the section on error prediction below. In any case, G now operates in a discrete column space and can be expressed simply as an observation matrix,

that is

$$\begin{aligned} \text{TEC} &= \langle G, N_e(r) \rangle \\ &\approx \langle G, \Psi_{KLM}(h, \theta, \phi) \rangle \cdot x(t) \\ &= A \cdot x(t) \end{aligned} \quad (11)$$

where $x(t)$ is the vector of coefficients we seek to estimate and the inner product of G and $\Psi(h, \theta, \phi)$ is carried out over each measurement line of sight. The explicit time dependence of x will be dropped in the sequel.

For the purpose of this paper we consider $\langle \cdot \rangle_{\ell_2}$ as the inner product for (11). The appropriate inversion in $L_2[a, b]$ is the stochastic inverse

$$A^{-g} = \Sigma_m A^T (A \Sigma_m A^T + \Sigma_d)^{-1} \quad (12)$$

where Σ_m in the absence of an *a priori* is the red spectrum covariance presumed above and Σ_d is the measurement covariance. This choice of inner product places the inversion implicitly in the larger framework of a Kalman filter. In fact as mentioned in [14], the stochastic inverse of the observation matrix is nothing more than the Kalman gain matrix.

Although $L_2[a, b]$, is a rather “large” space and certainly not the only choice, it is one of the simplest to evaluate, an important consideration in the user receiver where computational power is a premium. More importantly it has the additional benefit that the covariance of the estimate is available in closed form. This is critical to any system such as WAAS which must generate a rigorous confidence interval.

The real-time inversion process at one epoch is now straightforward.

1. Collect vector of TEC measurements, d , from reference network
2. Build A from TRS-SV geometry (one row per measurement)
3. Form electron density estimate, $\hat{x} = A^{-g}d$
4. Form covariance on that estimate, $\Sigma_x = A^{-g}\Sigma_d A^{-gT} + (I - A^{-g}A)\Sigma_m$

RANGE ERROR PREDICTION

Once the vector of electron density coefficients and its covariance are available to the user’s receiver, the range correction for the unknown TEC delay on single frequency measurements can be predicted by constructing the local observation matrix, A_u , and dotting

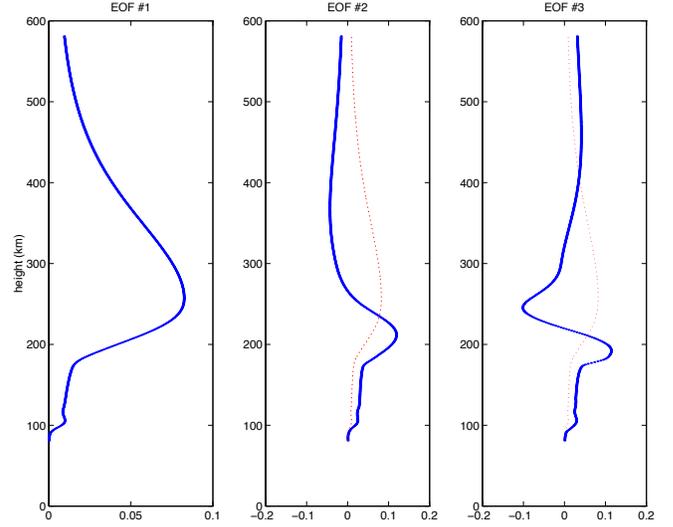


Figure 3: Vertical empirical orthogonal functions from 1990 International Reference Ionosphere electron density distributions.

it into \hat{x} .

$$\hat{d}_u = A_u \hat{x} \quad (13)$$

and the covariance estimate on the linear prediction is given by

$$\Sigma_{d_u} = A_u \Sigma_x A_u^T \quad (14)$$

The range correction is then subtracted from the current pseudorange measurements which in turn are propagated through the navigation equation. The covariance estimate is also propagated through the integrity equation to form a confidence interval. The reader is referenced to [17] for details on this process.

The user’s TEC covariance estimate lays the groundwork for setting the number of basis functions used to approximate the $N_e(r)$ space. Setting the upper bound on the number of modes lower than the resolving power of the reference network aliases observed ionospheric modes onto the user receiver’s so called unobserved modes [3]. There is no theoretic upper bound when employing an optimal filter since unresolved modes will be damped out. However, the work needed to build A is linear in the number of modes, $K \times L \times M$, and the inversion of A is at best quadratic in that same number.

Choosing the actual number of basis functions differs depending on the operational scenario. In scientific purposes with no real-time constraint and a focus on resolving the electron density the number of basis functions should be at the upper end of the spectrum. Given a fixed observation network the cost

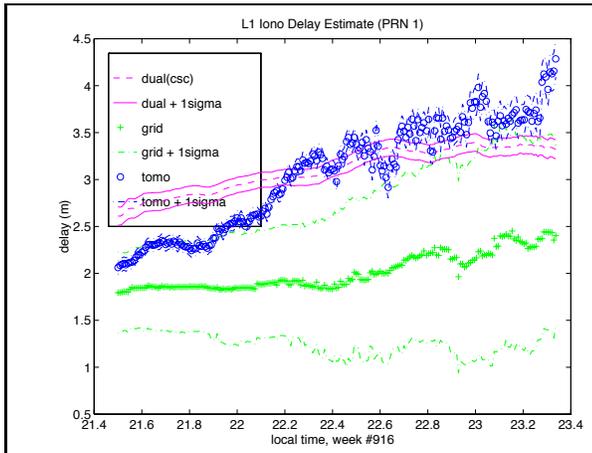


Figure 4: Dual frequency carrier smoothed TEC measurements from a static user are compared against code delay predictions from the grid and tomographic estimators. In this case the monitor station at Grand Forks, ND is viewing PRN #1.

of adding basis functions is two-fold: computational expense, but more importantly increased observation time needed for the geometry to resolve additional modes. This second cost is the critical condition when moving to the real-time arena. The only way to mitigate the time dependence is to have a strong *a priori* estimate and a dynamic model possessing the unusual property of accuracy that increases with wave number. Neither of these are currently available at the level of maturity needed for safety of life integrity requirements demanded in WAAS.

Fortunately in our case we have an escape hatch from this dilemma. Recall that we wish to correct the TEC induce pseudorange error for the single frequency user. The estimate and therefore also its covariance are passed back through the observation equation. If we examine (9) we see that it is a smoothing operator and thus the effective wave number in TEC is reduced from that in $N_e(r)$. Provided the inversion is faithful in covering any biases caused by aliasing with its covariance estimate, the basis functions, $\Psi_{klm}(h, \phi, \theta)$ can be truncated to a number commensurate with the allowable observation time.

We have chosen spherical harmonics to order 5, and 3 EOFs, The EOFs are shown in Figure (3) and extend from 80 to 580(km). As compared with previous work of other authors [14, 18, 19] this number of terms is the low end of the spectrum in concordance with our decision against incorporating dynamics and the presumption of an *a priori* distribution. The observation matrix constructed over this basis is however rotated into the solar-magnetic frame to take advantage of any

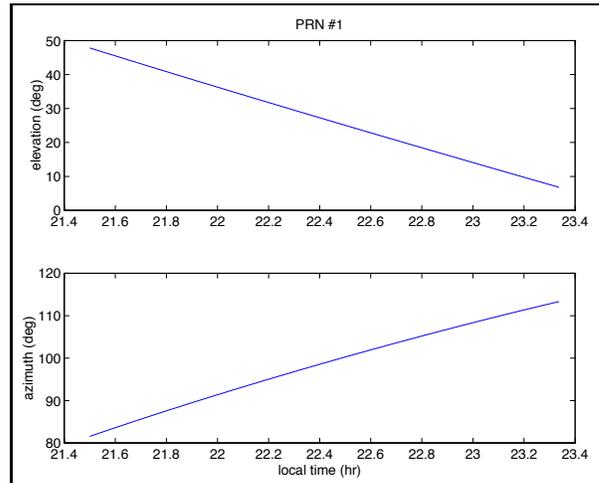


Figure 5: The elevation and azimuth of PRN #1 viewed by the Grand Forks monitor station corresponding to the ionospheric delay curves of Figure 4

stationarity in the physical electron density distribution. This mitigates some of the latency in transmitting the estimate to the user. A schematic of the complete correction process is shown in Figure (2).

We have applied the tomographic estimator described above and the grid estimator of [2] on live NSTB data. We present here results from two particular data sets, one collected on 27 July 1997 and the other on 8 August 1997 from the the reference network shown in Figure 1. These data correspond to portions of the same data analyzed in [20] for clock/ephemeris integrity monitoring and in [17] for user integrity analysis. The receivers marked by stars were not incorporated into the WAAS estimation but rather treated as static users.

RESULTS

We have the capability to monitor the performance of the user's ionospheric correction in both the pseudorange domain and position domain. The former is a direct assessment of the TEC delay correction but is difficult to gauge as even dual frequency measurements do not necessarily constitute truth. The latter is an indirect assessment but in the case of a static user can quantify accuracy, integrity, and availability against truth.

Figure 4 compares the code delay prediction of the two estimators against the dual-frequency carrier smoothed estimate on the satellite range measurements to PRN #1 in view of the static monitor station located in North Dakota. The first thing to note is the

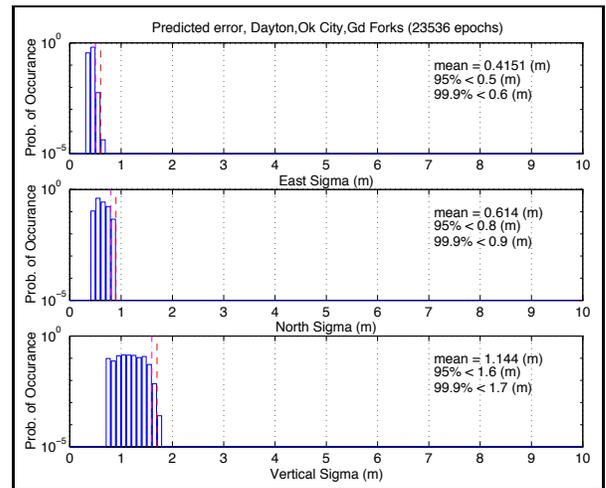
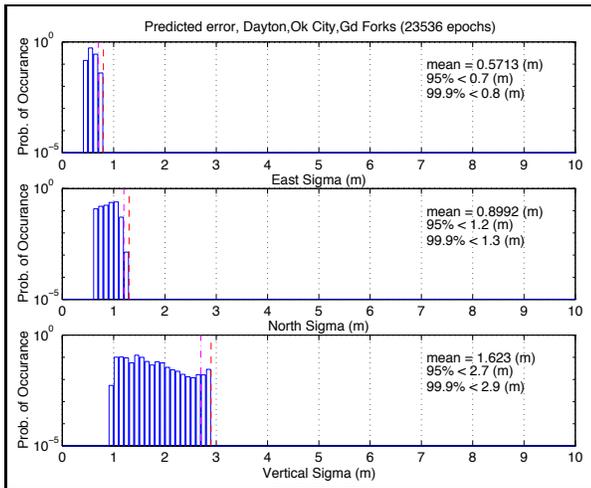
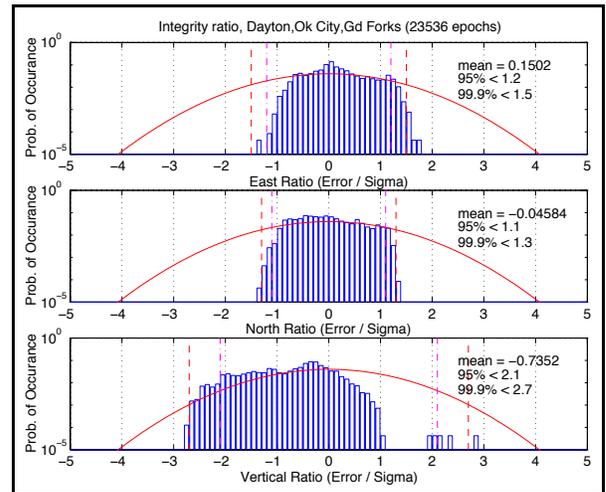
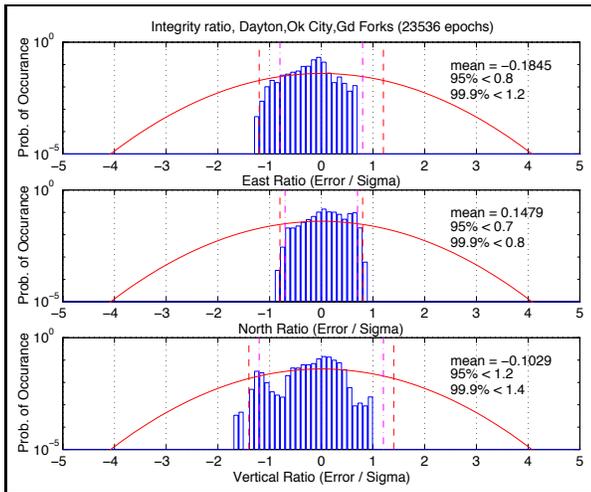
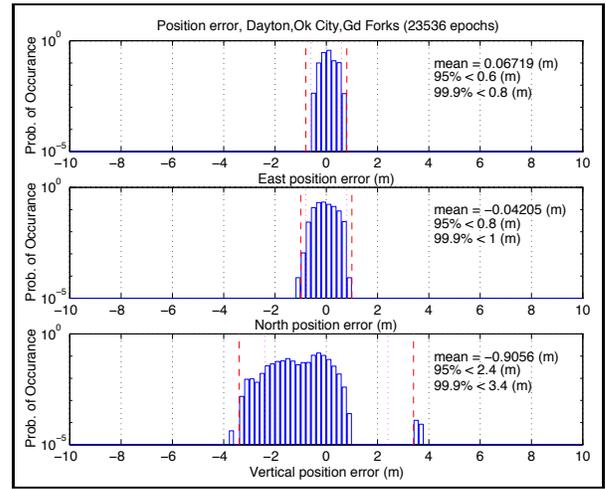
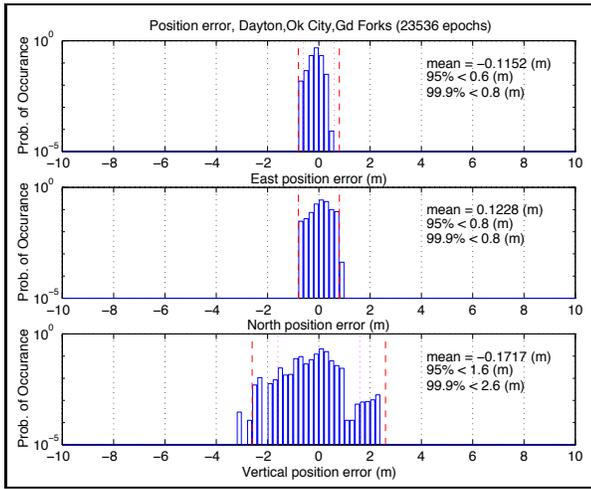


Figure 6: Accuracy, integrity, and availability results of the end-to-end WAAS correction process using the vertical delay grid of [2] to estimate ionospheric delay.

Figure 7: The same metrics applied to the end-to-end WAAS correction process using tomographic inversion to estimate ionospheric delay.

generally low level of overall delay due to the current minimum of the ~ 11 year solar cycle. Secondly, the

trends follow the elevation dependence implicit in the thin shell model.

The grid estimator’s major deficit is in specifying the range correction confidence interval at low elevation angle. The case shown above is typical, where the variance is artificially inflated to cover the obliquity factor assumption. There is no rigorous way to verify the resulting confidence which is a small but non-negligible integrity risk. Further, the variance can not be closed down where appropriate causing a direct penalty in availability.

The insights of the pseudorange comparison are helpful but, as mentioned above, difficult to apply quantitatively. We can assess the veracity of the static user’s correction and confidence bounds in the position domain and assign numbers to the accuracy, integrity, and availability. The drawback is that all error sources, multipath, troposphere, receiver noise, residual clock and ephemeris terms, and residual interfrequency biases, are rolled into the position solution. Applying the two estimators to the same data provides relative rather than absolute performance improvements, nonetheless this exercise is very informative. WAAS corrected positions using the two estimators were computed at Dayton, OH, Oklahoma City, OK, and Grand Forks, ND, comprising a little more than 6 hours of 1 Hz data.

The top histogram Figure 7 reports the cumulative accuracy for the complete WAAS navigation solution with the grid and tomographic estimators in the East, North, and Vertical local coordinates and tomographic estimators. The effect of using the truncated basis in the tomographic estimator shows clearly in the resulting biased position error. The geometry of the reference network is capable of supporting higher order spherical harmonics. While this is useful as an academic measure, the important information is contained in the bottom two histograms.

The middle chart in Figure 7 addresses integrity by showing the histogram of the the actual-to-predicted error ratio at each epoch in the data set. Using the σ_v equation defined in [17], integrity is maintained so long as the tails of the histogram are bounded by the unit Gaussian. This bound is shown as the solid curve and in both cases integrity was preserved. Although the biased position accuracy of the tomographic estimator does push the interval out noticeably, the shape of the histogram is retained and even slightly curtailed.

The bottom graphic in Figure 7 shows the WAAS predicted one σ error bound histogram. For a navigation sensor error bound of 19.2 (m) at a probability level better than 10^{-7} the maximum allowable bound is 3.6 (m), if σ is larger than this value then the sys-

tem is declared unavailable at that location and epoch. For both ionospheric estimators the WAAS system was available at these three locations more than 99.9% of the duration of this data set. However, the three nines σ value is dramatically lower for the tomographic estimator, without extending the tails of the integrity histogram. While the duration of the data set analyzed here is not statistically significant for an operational scenario, the benefits of extending the WAAS ionospheric correction process to include the vertical dimension have become more lucid.

COMMENTS AND FUTURE WORK

We have put in place the mathematical framework and an early prototype here. The application of tomography to the correction of ionospheric delay affords the possibility of simultaneously improving the availability and integrity of the WAAS system. In the short term, since the first round of WAAS MOPS for precision approach are nearing completion, the tomographic estimator could be used to generate a MOPS compliant grid estimate simply by integrating out the vertical dimension.

In the longer term, additional message types could be developed which leverage the compactness of more sophisticated bases and both the grid and tomographic models could be broadcast. This is particularly appealing when we consider the concept of focusing the correction bandwidth on the regions with highest resolution at any given epoch.

As the NSTB collects more and more data, we can address the choice of basis functions more appropriately and explore other inner products which induce a better norm and subsequently a more accurate inversion than the one presented here.

ACKNOWLEDGMENTS

The authors wish to thank Prof. Bruce Howe from the University of Washington Applied Physics Laboratory for helpful discussions. The FAA Technical Center is acknowledged for maintaining the NSTB network that provided the live data for this work. Our thanks also go to the Satellite Program Office for sponsoring WAAS research at Stanford.

REFERENCES

- [1] P. Enge, T. Walter, S. Pullen, C. Kee, Y.-C. Chao, and Y.-J. Tsai, “Wide area augmentation of the

- the global positioning system,” *Proceedings of the IEEE*, vol. 84, no. 8, pp. 1063–1088, 1996.
- [2] Y.-C. Chao, S. Pullen, P. K. Enge, and B. W. Parkinson, “Study of WAAS ionospheric integrity,” *Proceedings of the Institute of Navigation GPS-96*, pp. 781–788, September 1996.
- [3] A. J. Hansen, Y.-C. Chao, T. Walter, and P. Enge, “Ionospheric estimation and integrity threat detection,” *Proceedings of the Institute of Navigation National Technical Meeting*, pp. 883–889, January 1997.
- [4] A. Draganov, T. Cashin, and J. Murray, “An ionospheric correction algorithm for WAAS and initial test results,” *Proceedings of the Institute of Navigation GPS-96*, pp. 789–797, September 1996.
- [5] R. S. Conker, M. B. El-Arini, T. W. Albertson, J. A. Klobuchar, and P. Doherty, “Description and assessment of real-time algorithms to estimate the ionospheric error bounds for WAAS,” *Navigation*, vol. 44, pp. 77–87, Spring 1997.
- [6] K. G. Budden, *The Propagation of Radiowaves*. Cambridge: Cambridge University Press, 1988.
- [7] J. Hargreaves, *The Solar–Terrestrial Environment*. Cambridge: Cambridge University Press, 1992.
- [8] B. W. Parkinson and J. Spilker, eds., *Global Positioning System: Theory and Application*. Washington, D.C.: AIAA, 1996.
- [9] R. Hatch, “The synergism of gps code and carrier measurements,” Tech. Rep. MX-TM-3353-82, Magnovox, January 1982.
- [10] J. Austen, S. Franke, and C. Liu, “Ionospheric imaging using computerized tomography,” *Radio Science*, vol. 23, pp. 299–307, 1988.
- [11] H. Na, *Orthogonal Tomographic Inversion of the Ionosphere*. PhD thesis, University of Illinois, Urbana-Champaign, 1991.
- [12] E. Fremouw, J. Secan, and B. Howe, “Application of stochastic inverse theory to ionospheric tomography,” *Radio Science*, vol. 27, no. 5, 1992.
- [13] T. Raymund, “Comparisons of several ionospheric tomography algorithms,” *Annales Geophysicae*, vol. 13, no. 12, pp. 1254–1262, 1995.
- [14] B. Howe, “4-D simulations of ionospheric tomography,” *Proceedings of the Institute of Navigation National Technical Meeting*, pp. 269–278, January 1997.
- [15] E. Kreyszig, *Introductory Functional Analysis with Applications*. New York: John Wiley & Sons, Wiley classics ed., 1989.
- [16] R. L. Parker, *Geophysical Inverse Theory*. Princeton, New Jersey: Princeton University Press, 1994.
- [17] T. Walter, P. Enge, and A. J. Hansen, “A proposed integrity equation for WAAS MOPS,” *Proceedings of the Institute of Navigation GPS-97*, p. to appear, September 1997.
- [18] E. Fremouw, J. Secan, R. Bussey, and B. Howe, “A status report on applying discrete inverse theory to ionospheric tomography,” *International Journal of Imaging Systems and Technology*, vol. 5, pp. 97–105, 1994.
- [19] V. Kunitsyn, E. Andreeva, A. Popov, and O. Razinkov, “Methods and algorithms of ray radiotomography for ionospheric research,” *Annales Geophysicae*, vol. 13, no. 12, 1995.
- [20] Y.-J. Tsai, T. Walter, and P. K. Enge, “Integrity monitoring for WAAS clock/ephemeris error estimation,” *Proceedings of the Institute of Navigation GPS-97*, p. to appear, September 1997.