

Ionospheric data assimilation methods for geodetic applications

Paul S. J. Spencer, Douglas S. Robertson and Gerald L Mader
National Geodetic Survey, NOS/NOAA and
Cooperative Institute for Research in Environmental Sciences (CIRES)
and NOAA Space Environment Center
University of Colorado/NOAA,
Boulder, Colorado 80309-0216, USA.

Abstract—One of the major limiting factors in geodetic applications of the Global Positioning System (GPS) is lack of knowledge of the propagation delays imposed by the ionosphere. Single frequency, differential carrier phase measurements are limited to baselines with lengths less than the correlation size of the ionosphere (typically 10-20 km). Extending these measurements to longer distances requires accurate estimates of the slant total electron content (TEC) from a receiver to all observable GPS satellites. While dual frequency carrier phase measurements permit an ionosphere-free linear combination, accurate estimates of the double difference in integrated TEC between pairs of satellites and receivers provide an important constraint for accurate and rapid carrier phase ambiguity resolution. To achieve these accuracy requirements various approaches to the assimilation of ground-based GPS data from the CORS network and the mathematical representation of the ionospheric electron density field have been studied. The model presented uses a Kalman filter algorithm to assimilate data in various forms and an optional mapping function to alter the representation of the state vector in terms of a set of discrete radial empirical orthonormal functions (EOF's). Initial results from local networks show agreement with ambiguity-fixed double-differenced ionosphere delays of a few tenths of a TEC. The advantages of the various approaches and additional results will be discussed

I. INTRODUCTION

The problem of imaging the Earth's ionospheric electron density field has been studied for a number of years [1]. Early work applied tomographic

methods to data obtained from TRANSIT satellites. In this case two-dimensional reconstructions were obtained using spaced-station networks of receivers. More recent work uses a Kalman Filter approach which combines data from various sources with prior model estimates to obtain inversions in four-dimensions [2] [3] [4] [5] [6]. The aim of these various approaches is to find an optimal solution to a poorly constrained linear least squares problem. In the case of tomographic methods, the regularization of the inverse problem often takes the form of continuity relationships defining the smoothness or entropy of the solution over a range of neighboring pixels. Another approach used extensively in ionospheric tomography is to map the ionospheric representation to a set of orthogonal functions. For data assimilation methods the regularization is provided in the form of a prior model estimate of the solution along with an estimate of its covariance. The method presented in this paper is properly referred to as a Kalman filter based data assimilation method. However, it is shown that certain approaches more commonly used in tomographic methods can be applied to the Kalman filter approach to increase accuracy and performance.

Single frequency GPS applications require estimates of the integrated electron content (TEC) along slant paths. Results are presented for both geomagnetically quiet and disturbed cases. A comparison between results obtained from the model and an independent set of results obtained from the US geodetic survey is also presented.

II. IMPLEMENTATION OF THE KALMAN FILTER

This section provides an introduction to the Kalman filter assimilation method. The problem is to optimally update the solution to a linear least squares

problem given time dependent observations and a prior model estimate of the solution. The unknowns, which in this case, represent the ionospheric electron density field, are stored in a state vector, x . Associated with the state is a covariance matrix, P , which is updated by the filter each iteration. The sequence of steps in updating the filter may be defined as follows: First the state vector, x , is projected into the future (the minus superscript implies prior estimates)

$$x^- = Ax^- + B$$

For the method presented in this paper the matrices A and B are generated from prior model estimates of the electron density field, M , as follows;

$$A_{i,j} = (1 - \alpha)G_{i,j} M_i^- / M_j$$

$$B_i = \alpha M_j$$

The matrix G represents the correlation between neighboring voxels in the image area and is defined by a Gaussian function. The parameter α is a constant between zero and one. It can be seen that with α set to zero the forward projection of the state is defined by the relative spatial-temporal gradients in the model and with α of unity the forward projection of the state becomes the model estimate. Typically, α is set to a value of 0.1 such that the projection of the state is dominated by the relative spatial and temporal gradients in the model. The aim of this approach is to provide a form of mathematical regularization which is insensitive to systematic errors in the background model.

The next stage in the update of the filter involves projecting the error covariance matrix

$$P = AP^-A^T + Q$$

Where the process noise matrix Q is defined as,

$$Q_{i,j} = kG_{i,j}q_iq_j$$

$$q = (M + Ax^-) / 2$$

Where, k is a constant, and G , as before, defines the correlation between spatially separated terms. The Q matrix defines the variance as being a constant fraction of an average of the background model and projected state estimates.

Given a set of line integral observations, z , with covariance R , and path integrals defined by, H , the Kalman Gain is given by

$$K = PH^T(HPH^T + R)^{-1}$$

Finally, using the Kalman gain, the state vector and its covariance are updated

$$x^+ = x + K(z - Hx)$$

$$P^+ = (I - KH)P$$

The simplest representation of the three-dimensional electron density field is to define the state as the electron density in a piece-wise constant or tri-linearly interpolated grid of voxels. For the tomographic problem a number of authors have used an approach which maps this state vector into a space defined by sets of orthogonal functions[7]. Typically these basis sets consist of spherical harmonics for the surface variations and a set of empirical orthogonal functions (EOF's) for the radial profile. As well as regularizing the problem, the number of unknowns is also drastically reduced. For the method presented in this paper an optional mapping of the state vector has also been implemented. In this case, however, the mapping is only applied to the radial profile. Consequently, the ionosphere is constructed from a set of EOF's which are discrete in latitude and longitude. Typical examples of the first three EOF's are shown in Fig. 1.

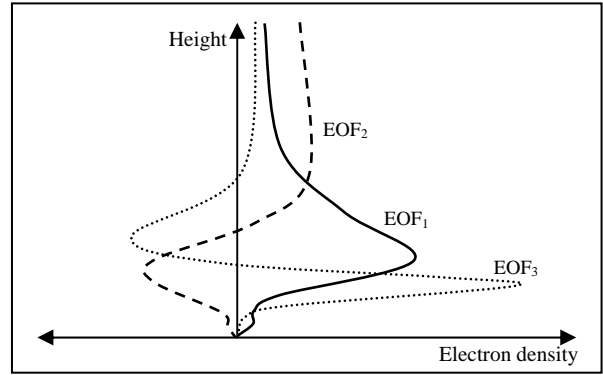


Fig.1. Example empirical orthogonal functions (EOF's).

These functions were generated by applying a singular value decomposition to a set of model profiles obtained using the IRI95 model which is also used as the background model in the Kalman filter. The dominant term, EOF₁, represents a mean ionospheric profile. The higher order EOF's, which gradually decrease in significance, allow the profile to depart from the mean. Mathematically, the mapping of the state from a voxel based representation to a set of discrete EOF's can be written as

$$x \rightarrow M^T x$$

$$A \rightarrow M^T AM$$

$$P \rightarrow M^T PM$$

$$Q \rightarrow M^T QM$$

$$H \rightarrow MH$$

III. ASSIMILATION OF GPS OBSERVATIONS

The phase observable, ϕ_i , on frequency i , measured by a GPS receiver can be written as

$$\phi_i = \rho/\lambda_i - I/\lambda_i f_i^2 - T/\lambda_i - n_i$$

The four terms here represent geometrical distance, ionosphere, troposphere and the integer ambiguity. Various linear combinations of the two frequencies are of use in GPS data processing. The geometry free combination, L4, is given by

$$\phi_1 \lambda_1 - \phi_2 \lambda_2 = -I(1/f_1^2 - 1/f_2^2) - (\lambda_1 n_1 - \lambda_2 n_2)$$

L4 provides an estimate of the ionospheric slant TEC along the satellite-receiver path to within an unknown constant. When calibrated to the P-code observations, an estimate of the slant TEC subject to satellite/receiver clock biases and multi-path is obtained. Given an estimate of the satellite clock biases the receiver biases can be solved for in the filter. It is assumed that with sufficient observations the error in calibration of the phase to the pseudo range due to multi-path is negligible.

For precise geodetic applications it is necessary to resolve the integer nature of the phase ambiguity. The first stage in this processing strategy is to resolve the wide-lane, L5, ambiguity.

$$\phi_1 - \phi_2 = \rho(f_1 - f_2)/c - I(1/f_1 - 1/f_2)/c - T(f_1 - f_2)/c - (n_1 - n_2)$$

The value of the wide-lane combination is in its large wavelength of 0.86 m which corresponds to approximately 4 TECU.

One of the goals of this work was to provide an accurate estimate of the ionosphere for geodetic applications. For precise geodetic positioning a double difference of the L5 observable is generated for a pair of receivers and satellites. For this reason an accurate determination of the double difference in ionospheric slant TEC was necessary. To achieve this, a method was developed of assimilating double difference observations that uses the integer nature of the phase ambiguity to refine the ionospheric accuracy. The wide-lane double differences between a pair of receivers and satellites can be written

$$\nabla \Delta L_5 = \nabla \Delta \rho + \nabla \Delta T + \nabla \Delta I + \nabla \Delta n$$

Given estimates of the distances from satellites to receivers, the ionospheric term and the tropospheric term, it is possible to estimate the double difference wide-lane ambiguity

$$|\nabla \Delta n|_{est} = \text{int}(\nabla \Delta L_5 - \nabla \Delta \rho - \nabla \Delta T - |\nabla \Delta I|_{est})$$

Rounding this ambiguity to the nearest integer enables a refined estimate of the un-ambiguous double difference ionosphere to be made

$$\nabla \Delta I = \nabla \Delta L_5 - \nabla \Delta \rho - \nabla \Delta T - |\nabla \Delta n|_{est}$$

This information may then be assimilated into the filter by double differencing terms in the observation matrix, H, as

$$\nabla \Delta H \cdot x = k \nabla \Delta I$$

Using the prior solution obtained from the Kalman filter, the ambiguous slant TEC double-differences are relaxed to the nearest integer ambiguity. The improved double difference estimates are then fed into the filter along with geometry free data. It should be noted here that the assimilation of wide-lane double differences provide a local minimum only. Provided the ionospheric estimate is close to a true solution (within 2 TECU) then solution accuracy will be improved using this method. To obtain a global minimum the geometry free data must also be assimilated at the same time. Double-difference input data has been implemented in our code but has not yet been used in the results presented in section IV.

III. DATA

The data used in this study were obtained from the U.S. National Geodetic Survey's (NGS's) Continuously operating Reference System (CORS) network of GPS stations. As the name implies, this network is a set of several hundred GPS stations that operate continuously throughout the United States and Central America. The NGS provides this data in RINEX files that can be freely downloaded from the NGS website: <http://www.ngs.noaa.gov/>. Fig. 2 shows a map that gives the locations of the CORS stations that were operational in early 2004. The number of CORS stations is increasing steadily as new stations are added to the network.

There are two ways that are commonly used to express the ionosphere model values. The simplest parameter is simply the density of free electrons, usually designated N_e . The other parameter that is commonly displayed is the integrated electron density along a line, referred to as the total electron content (TEC). The line of integration is typically either a raypath to the satellite (sometime termed a slant TEC) or a vertical line above a point on the ground. The TEC is expressed in terms of TEC Units (TECU). One TECU is 10^{16} electrons/square meter.

A handy rule-of thumb is that one TECU causes roughly one cycle of phase delay at GPS frequencies.

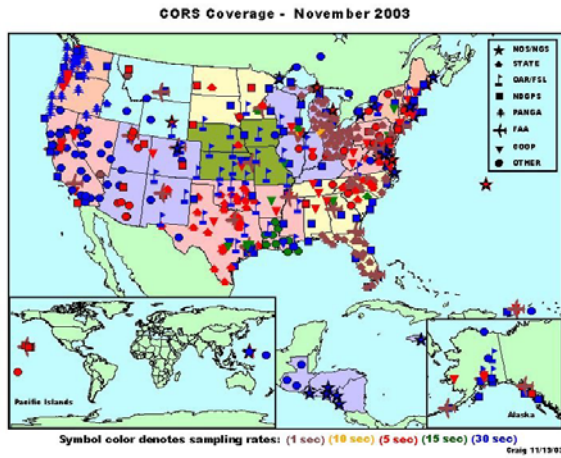


Fig. 2. Map of the CORS station locations.

Fig. 3 shows a schematic representation of some of the CORS GPS data. The plot displays the intersection points where a set of CORS data raypaths intersect a hypothetical shell located at an altitude of 350 km. All of the raypath intersection points within a 15-minute interval are shown. The points are coded by a grayscale that illustrates the line-of-sight TEC value along each raypath as determined by the calibrated phase differences. The key to the grayscale values is shown in the bar at the right in TEC units. Raypaths for stations in coastal Alaska and Central America as well as the continental United States can be seen. This plot covers a time interval that includes the height of the ionosphere storm of Oct 29, 2003. The anomalously high TEC values in the southern regions of the United States resulting from this storm are clearly visible in the plot.

Estimated TEC at 350km shell, 29-Oct-2003 21:00:00UT

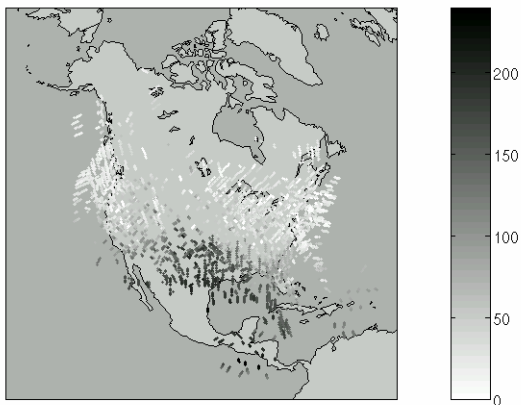


Fig 3. Plot of raypath intersections with a 350 km high shell.

IV. RESULTS

There are several problems involved in presenting and discussing the results from ionosphere modeling. The quantity of information is so vast that it is not easy to devise effective methods of displaying it in print; it is a classic “drinking out of a fire hose” problem. The ionosphere problem is inherently four dimensional even when only the spatial distribution of electron density is considered, and even more dimensions are needed if other ionosphere properties such as ion composition were to be displayed. Perhaps the only effective method of presenting four-dimensional data is to show a variety of two-dimensional sections through the data. The other major problem in presenting ionosphere model results centers on the need to determine and display the quality of the model values, *i.e.*, the error levels of the estimates of ionosphere model parameters.

We can begin to examine the modeled ionosphere by looking at some projections of the model results. Fig. 4 shows a two-dimensional section that displays the vertical TEC in TEC units for a quiet ionosphere day (Oct 28, 2003) at this point in the solar cycle from a solution that used data from CORS GPS network across the continental United States. The solution used phase data calibrated (“leveled”) to the pseudorange values; it estimated receiver biases in addition to ionosphere parameters. The observed TEC values in the range of about 50 TECU are fairly typical for late afternoon of a quiet ionosphere day. In contrast, Fig. 5 shows the ionosphere at about the same time on the following day, in the middle of one of the strongest ionosphere storms ever recorded. From the southwestern U.S. down across Mexico the TEC values exceeded 200 TECU. A complete .avi motion picture file that displays the ionosphere behavior through this time period can be found on the web at the National Geodetic Survey website: <http://www.ngs.noaa.gov>.

Inversion TEC(TECU) 28-Oct-2003 21:15:00UT

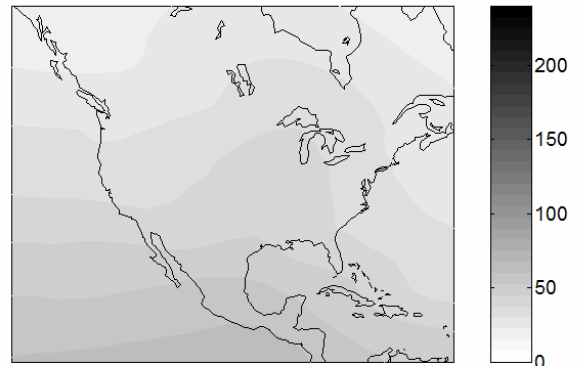


Fig. 4. Vertical TEC in TEC units for a quiet day.

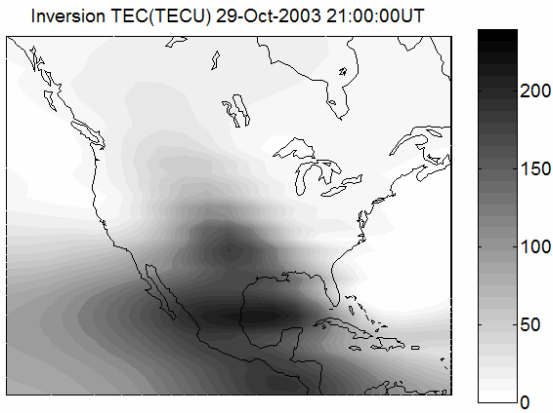


Fig. 5. Same as Fig. 4 for an ionosphere storm day.

Notice that during this storm the TEC values across the northeastern U.S. are extremely low, comparable to nighttime conditions. This is an example of a recurrent feature of ionosphere storms that is well known among ionosphere researchers but not always appreciated among user groups, that an ionosphere storm can drive TEC values in negative directions as well as positive, *i.e.*, to TEC values that are lower than normal as well as higher. The reasons for the regional decreases in ionization during a storm are related to changes in the neutral composition of the ionosphere [8] [9] [10].

Figs. 6 and 7 show the line-of-sight slant TEC values calculated from this model for a CORS station in Northhampton, MA on both the storm day and a typical non-storm day. The TEC values on the storm day are about four times higher than on the non-storm day, even for a station that is far away from the main storm. Notice also that during the storm, many of the ray-paths exhibit very low TEC values, with a sharp break downward at about 19:00 UTC, corresponding to the reduction in TEC values seen in Fig. 5.

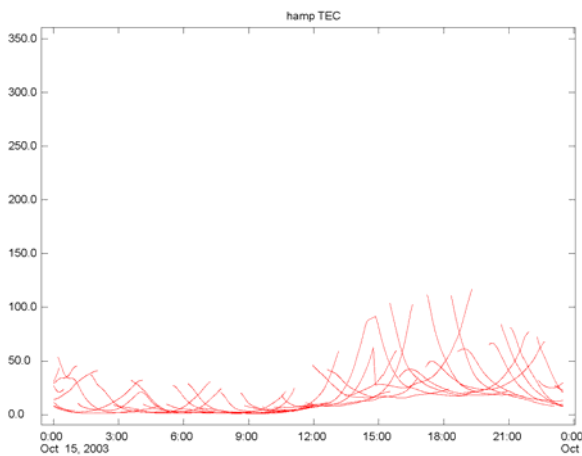


Fig. 6 Line-of-sight TEC values in units of L1 phase for Northhampton, MA on a quiet day.

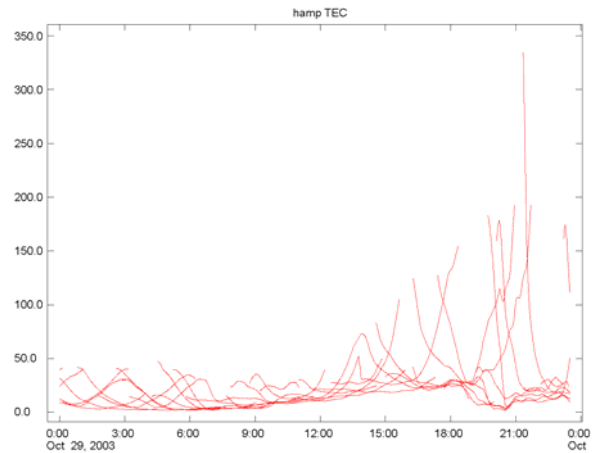


Fig. 7. Same as fig. 6, for an ionosphere storm day.

There is yet another way to look at the ionosphere behavior modeled with our tomographic software. The data shown in Figs. 2 and 3 are similar to what could be produced by a shell-model. But the use of EOF's allows us to recover some information about the vertical profiles of the electron density, N_e . Fig 6 shows a cross-section of N_e values along the longitude meridian at 115 degrees west during the storm shown in Fig. 5. In effect, a shell model would “squash” this vertical variation into TEC values at a pre-determined shell height. Yet it is clear from this figure that there are low-angle lines of sight coming in from the ground at the right side of the figure that would intersect such a shell at locations that have low TEC values, but the line would then go on to intersect regions of high electron density that lie above the shell. The ability of the tomographic models to retain some of this vertical density variation information is one of the advantages of this modeling technique.

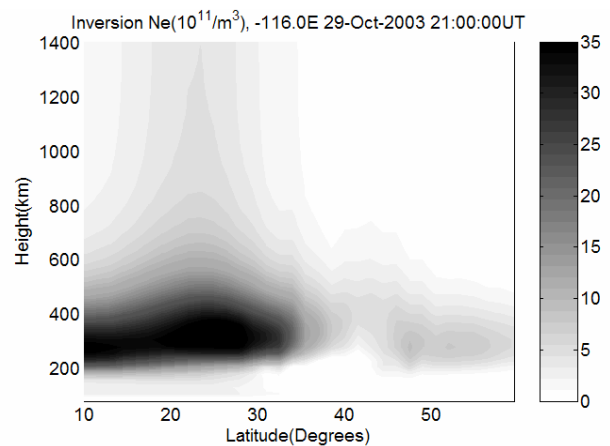


Fig. 8. Vertical distribution of electron density along the 116 degree west longitude during the ionosphere storm of Oct. 29.

V. VALIDATION AND INTERCOMPARISONS

Data assimilations and sophisticated mathematical models are of relatively limited use unless they can be verified and validated so that reasonable estimates of their error levels can be determined. In NASA space missions such validation is sometimes referred to as determining “ground truth.” But ground truth can be difficult to come by for the ionosphere because there are very few in-situ measurements available. There are some satellite-based in-situ measurements of electron density in the ionosphere (from the DMSP satellites), but the spatial coverage of such data is extremely sparse. Nowhere are they dense enough to calculate line-of-site TEC values that could be directly compared to GPS observations.

We have approached this problem by making some preliminary intercomparisons with ordinary GPS data. One of the critical questions for GPS data analysis is whether *a priori* geodetic models can be made accurate enough to allow the resolution of phase ambiguities, or at least reduce the volume of parameter space that must be searched in the process of resolving ambiguities.

To resolve this question we have constructed a program that calculates the magnitude of the wide-lane residuals for double differenced data, *i.e.*, the discrepancies within a wide-lane cycle between the observed delays and our best models that use precision ephemerides and other accurate modeling algorithms. The residuals were calculated with and without model ionosphere corrections for a variety of baseline lengths ranging from about 50 km to about 500 km. If the models were perfect, then these residuals would all be identically zero. They become as large as 0.5 cycle due to model errors, particularly including unmodeled ionosphere effects. The question we are addressing here is whether the modeled ionosphere corrections shift the wide-lane residual magnitudes significantly closer to a value of zero. The closer they are to zero, the easier it is to resolve phase ambiguities.

Fig. 9 shows two histograms of these residual magnitudes on a quiet ionosphere day for a 53 km baseline in the central U.S.. The bars exhibit the fraction of observations that fall into bins whose width is 0.05 wide-lane cycle. The light gray bars show the fraction of residuals that fell into each 0.05-cycle bin without any ionosphere corrections, and the darker bars to the right show the same fraction of residuals after the ionosphere corrections were applied. The dark bars (residuals with ionosphere corrections) are clearly higher toward the zero end of the scale and lower toward the other end (*i.e.*, the number of residual magnitudes that were close to

zero was larger, and the number that was close to 0.5 was smaller.). The differences that are caused by the modeled ionosphere corrections are small here, as might be expected over a baseline of only 53 km. 53 km is short enough that the ionosphere correction is largely the same at both ends of the baseline and therefore cancels from the differenced observations.

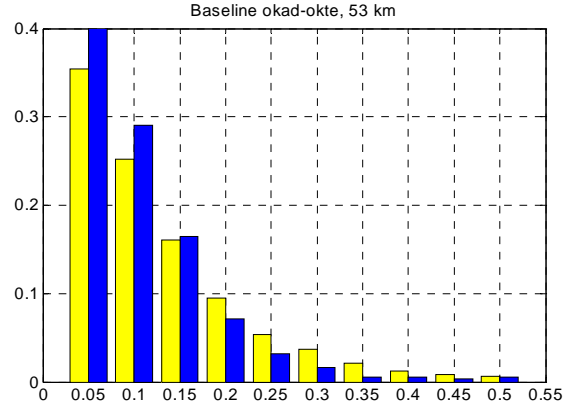


Fig. 9. Histogram of the magnitudes of wide-lane residuals without ionosphere correction (light gray bars) and with ionosphere corrections (dark gray bars) for a 53 km baseline in Oklahoma.

Fig. 10 shows a histogram for data on the same day but for a 123 km baseline:

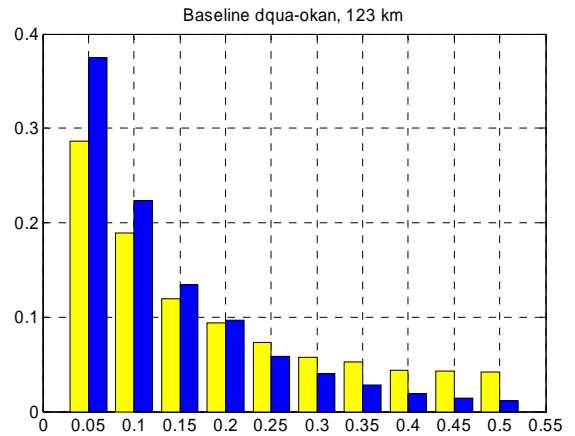


Fig. 10. Same as Fig. 9 for a 123 km baseline.

Here the shifting of residual magnitudes toward zero is substantially larger than in Fig. 9, presumably because there is less cancellation of the ionosphere effects at this baseline length. Fig. 11 shows the same residual histograms for a 500 km baseline:

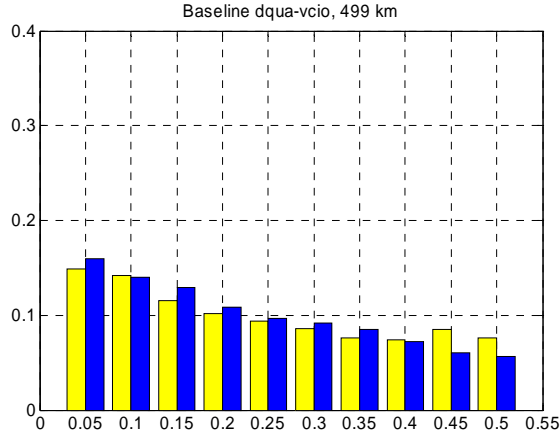


Fig. 11. Same as Fig. 9 for a 499 km baseline.

At 500 km the overall performance of the model is substantially poorer (*i.e.*, fewer residuals are near zero), as might be expected for the longer baseline length, yet still the modeled ionosphere corrections move the residuals systematically toward zero.

Next, Fig. 12 shows the same residual histograms for the 123 km baseline (as in Fig. 10), but this time on a day with an extremely strong ionosphere storm:

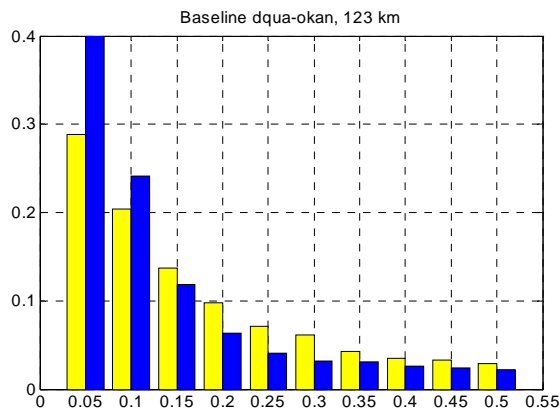


Fig. 12. Same as fig. 10, but for an ionosphere storm day, Oct 29, 2003.

Notice that the non-ionosphere residuals are worse on the ionosphere storm day than they are on the quiet day shown in figure 8, but only by a slight amount. This may be a result of the fact that many of the raypaths are going through regions of the ionosphere toward the northeast direction where the ionization levels have been destroyed by the storm. Still the modeled ionosphere corrections shift the residuals substantially toward zero.

The final histogram in fig. 13 shows the same baseline and day as fig. 10, but restricted to only the data during the period of the storm, from roughly

15:00 UTC to the end of the day. In this time period the residual statistics are markedly worse than over the whole day, but even here the modeled ionosphere corrections improve the residuals slightly.

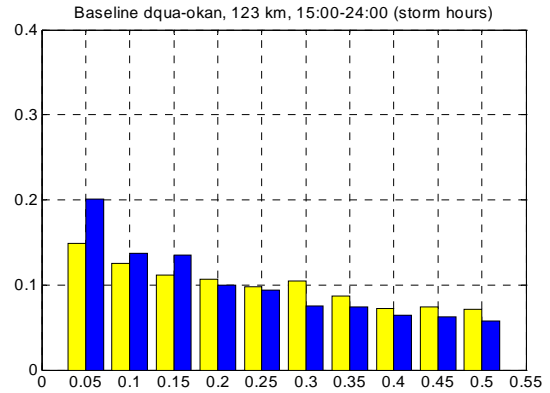


Fig. 13. Same as Fig. 12, but only for data during the time of the storm, after 15:00 UT.

We plan to do further testing using single-frequency observations with and without ionosphere corrections. The degree to which the ionosphere corrections move us closer to the dual-frequency determinations of the positions of the same stations will give us another measure of the quality of the ionosphere model.

VI. CONCLUSIONS

Modeling the ionosphere well enough to improve GPS data processing is an exceedingly complex problem. But substantial progress has been made in implementing Kalman filter data assimilation algorithms that are able to process the vast quantities of new data that have recently become available. These algorithms rely heavily on effective exploitation of enormous recent advances in computer power as well as the use of modern computer graphics to help understand the results.

There are at least three major areas in which we can expect substantial additional improvements for ionosphere modeling the near future. The first area involves improvements in the input data. Additional stations could make the GPS data more dense and more uniformly distributed. This would be particularly important if the models were to be made global rather than regional. Also the quality of the data can be improved by exploiting new technologies for decreasing multipath effects, and through the use of improved algorithms for ambiguity resolution, for example

A different class of data improvements involves exploiting entirely new data types, including GPS observations by LEO satellites, which would give

much greater sensitivity to vertical profiles than can be obtained with ground based data. Other novel data types include in-situ measurements of electron density by satellites such as DMSP satellites. Yet another novel data type exploits satellite observations of the dispersion of VHF signals generated by lightning [11].

The mathematical models used here could be improved in a number of ways. For example we could include more EOF terms, or we could increase the density of the horizontal grid. Serious improvements along these lines would require major increases in computer power, both RAM size and processor speed. Conversely, improvements in computer power alone would allow significant improvements in our models even without any increase in the quantity or the quality of the input data. In other words, the models are significantly compute-bound at present. We also need to explore the parameter space that defines the some of the model assumptions such as noise levels and correlation-length values that are built into the Kalman filter algorithms.

Clearly the process of validation and intercomparison of results could be expanded indefinitely. Additional intercomparisons should include both internal consistency checks with GPS observations and intercomparisons with other data types such as TOPEX/JASON altimetry.

Thus it appears that we are entering a very exciting time in the history of ionosphere research when a combination of novel data and analysis techniques plus extraordinary advances in computer resources promise to enormously expand our understanding of the details of the behavior of the ionosphere.

REFERENCES

- [1] Austen, J.R., S.J. Franke, and C.H. Liu, Ionospheric imaging using computerised tomography, *Radio Science*, 23 (3), 1988, pp. 299-307.
- [2] Yeh, K.C., and T.D. Raymund, Limitations of ionospheric imaging by tomography, *Radio Science*, 26 (6), 1991, pp. 1361-1380.
- [3] Fremouw, E.J., J.A. Secan and B.M. Howe, Application of stochastic inverse theory to ionospheric tomography, *Radio Science*, 27 (5), 1992, pp. 721-732.
- [4] Bernhardt, P.A., R.P. McCoy, K.F. Dymond, J.M. Picone, R.R. Meier, F. Kamalabadi, D.M. Cotton, S. Chakrabarti, T.A. Cook, J.S. Vickers, A.W. Stephan, L. Kersley, S.E. Pryse, I.K. Walker, C.N. Mitchell, P.R. Straus, H. Na, C. Biswas, G.S. Bust, G.R. Kronschnabl, and T.D. Raymund, Two dimensional mapping of the plasma density in the upper atmosphere with computerized ionospheric tomography (CIT), *Physics of Plasmas*, 5 (5), 1998, 2010-2021.
- [5] Bust, G.S., D.S. Coco, and T.L. Gaussiran, Computerized ionospheric tomography analysis of the combined ionospheric campaign, *Radio Science*, 36 (6), 2001, pp. 1599-1605.
- [6] Scherliess, L., R.W. Schunk, J.J. Sojka, and D.C. Thompson, Development of a physics-based reduced state Kalman filter for the Ionosphere, in *2002 Ionospheric Effects Symposium*, Alexandria, VA, J.M. Goodman, ed., May 7-9, 2002
- [7] Colombo, O.L., M. Hernandez-Pajares, J. Miguel Juan, J. Sanz, "Ionospheric Tomography helps resolve GPS ambiguities on the fly at distances of hundreds of kilometers during increased geomagnetic activity", *Proceedings IEEE "PLANS 2000" Meeting*, San Diego, March 2000.
- [8] Fuller-Rowell, T.J., M.V. Codrescu, R.G. Roble, and A.D. Richmond, How does the thermosphere and ionosphere react to a geomagnetic storm?, in *Magnetic Storms*, B.T. Tsurutani, W.D. Gonzalez, Y. Kamide, and J.K. Arballo, eds., American Geophysical Union Technical Monograph 98, Washington DC, 1997.
- [9] Prolss, G.W., Magnetic storm associated perturbations of the upper atmosphere, in *Magnetic Storms*, B.T. Tsurutani, W.D. Gonzalez, Y. Kamide, and J.K. Arballo, eds., American Geophysical Union Technical Monograph 98, Washington DC, 1997.
- [10] Buonsanto, M.J., Ionospheric storms—a review, *Space Science Review*: 88, p563, 1999.
- [11] Pongratz, M.B., D.M. Suszcynsky, T.J. Fitzgerald, and A.R. Jackson, Satellite-Based Global TEC monitoring, *Proceedings URSI National Radio Science Meeting*, in press, Boulder, CO 2004.