Ionospheric Data Assimilation by Use of the Kalman Filter

Hajj, George A.[1,2], Brian D. Wilson[1], Chunming Wang[2], and Xiaoqing Pi[1,2]

[1]Jet Propulsion Laboratory, California Institute of Technology,
[2]University of Southern California, Department of Mathematics

Abstract. A fully 3-dimensional Global Assimilative Ionospheric Model (GAIM) is currently being developed by a joint University of Southern California and JPL team. To estimate the electron density on a global grid, GAIM uses a first-principles ionospheric physics model ("forward" model) and one of two estimation techniques: the Kalman filter and 4DVAR (4-Dimensional VARiational). Because of the large dimension of the state (i.e., electron density on a global 3-D grid), implementation of a full Kalman filter is not computationally feasible. Therefore, we have implemented a band-limited Kalman in which a full time propagation of the covariance is performed but only a portion of the covariance matrix is retained. The retained elements are determined based on assumed physical correlation lengths in the ionosphere. The effectiveness of ground GPS data for specifying the ionosphere is assessed by assimilating slant TEC data from 98 sites into the GAIM Kalman filter and validating the retrieved electron density field against independent measurements. A series of GAIM retrievals are presented and validated by comparisons to: JPL's GIM map of vertical TEC, vertical TEC measurements from the TOPEX altimeter, and slant TEC data from GPS sites that were not included in the assimilation runs.

1. Introduction

The increasing reliance of our civilization on space technologies has made it clear that creating a "space weather" monitoring capability that provides timely and accurate space environment observations, specifications, monitoring and forecasting, is essential for the safe operation of various defense and commercial systems. The degree of success in creating such a "space weather" system depends mostly on (1) the ability to obtain *global* and *continuous* measurements related to the space environment and (2) the ability to incorporate these various measurements into a physical model in a self-consistent manner.

The state of monitoring and forecasting space weather today can be compared to that of conventional weather monitoring and forecasting almost half a century ago, when observations were fragmentary in space and time, and means of interpreting them were rudimentary. The global and continuous observations obtained in the lower atmosphere (e.g. from weather satellites and radiosondes), the ability to obtain these observations in a timely manner, and the advances made in global weather modeling and in data assimilation algorithms are major factors that have brought numerical weather prediction (NWP) models to their current level of success.

On the space environment front, we are witnessing a new era. Significant efforts are being planned to collect further information on solar activity and disturbances in the solar wind and the magnetosphere, and data on the upper atmosphere and ionosphere/plasmasphere are becoming truly global and continuous. A case in point is the GPS and the global network of over 200 ground receivers (the number is still rapidly increasing) that has created the unprecedented possibility of producing global maps of vertical total electron content (TEC) and ionospheric irregularities in near real-time updated sub-hourly [Pi et al., 1997a; Mannucci et al., 1998]. Moreover, within the next few years the number of flight receivers tracking GPS in "occultation" geometries will increase to more than a dozen in 3-4 years, providing an extremely dense global set of horizontal cuts through the ionosphere and allowing for accurate 4D global mapping of electron density. This data set, along with other data such as airglow radiances from the High Resolution Airglow/Aurora Spectroscopy Experiment (HIRAAS) on the ARGOS satellite as well as the Special Sensor Ultraviolet Limb Imager (SSULI) that will fly on future DMSP satellites, provide a truly unprecedented global 3D coverage of the upper atmosphere and ionosphere.

Our objective in this paper is to describe the development of a Global Assimilative Ionospheric Model (GAIM)[1] capable of assimilating a variety of data types including: (1) line-of-sight TEC measurements from GPS ground receivers, (2) change in TEC measurements taken from a low-Earth orbiter (LEO) tracking GPS satellites at positive and negative elevations (i.e., during GPS-LEO occultations), (3) in situ measurements of electron density, and (4) UV airglow radiances. Similar to neutral atmospheric weather models — which assimilate, solve for, and predict 4D fields (3 spatial and 1 temporal) of the atmospheric state parameters such as temperature, specific humidity and wind — GAIM assimilates, solves for, and predicts the electron density in the ionosphere and some of the

[1] GAIM also stands for Global Assimilation of Ionospheric Measurements. The latter acronym is used by the Utah State University (USU) led consortium. Both the USC/JPL and USU GAIM teams are funded by a Department of Defense (DoD) Multi-disciplinary University Research Initiative (MURI).

underlying driving forces ("drivers") such as production rates, dynamo electric fields, thermospheric neutral densities, temperatures and winds. In doing this GAIM applies two different techniques: (1) the Kalman filter or some approximation thereof, and (2) a 4-dimensional variational (4DVAR) technique. The former technique is used for solving the electron density in space and time while assuming the "drivers" to be known. The 4DVAR technique solves for the "drivers" from which the electron density is obtained by solving the ionospheric model equations.

Although the two approaches are currently disjoint, they can be combined in an operational scenario. The 4DVAR technique is described elsewhere [*Pi et al.*, 2001]. In this paper, our focus is on the use of the Kalman filter for estimating the ionospheric electron density state and its implementation. Even though the current GAIM is capable of assimilating a number of data sources as listed above [Hajj *et al*, 2000], we limit the scope of this study to assimilating ground TEC measurements from a network of 98 globally distributed stations. In doing so, we are following the general tradition and "wisdom" of the NWP community, which introduces new measurements into numerical weather models only after very careful examination and much evaluation. The reason is that each data set has its own nuances and characteristics, and it could influence the data assimilation output in both positive and negative ways. Therefore, optimal assimilation of any data type requires careful tuning of its error covariance, proper evaluation of the data representativeness errors, examination of the effect of the data on the analysis and its covariance, and examination of the consistency of the assumptions used in the Kalman filter and its solution.

In section 2 we review the formulation of the Kalman filter, discuss some practical considerations related to the full Kalman filter such as memory requirements and number of operations, and introduce some approximate forms of the Kalman filter. We describe the USC/JPL GAIM physical (forward) model and its solution grid in section 3. Section 4 presents examples of ionospheric specifications from GAIM retrievals on June 13, 2001 and validation results using comparisons to other data. Section 5 presents our conclusions and plans for future work.

2. The Kalman filter

We introduce the following definitions (commonly used in NWP):

x_k^t	The *true state*: a discrete representation of the true ionospheric state (density) at time *k*.
$x_k^a = \left\langle x_k^t / m_k^o, x_k^f \right\rangle$	The *analysis*: an estimate of x_k^t given measurements at time *k*, and a forecast x_k^f
$x_k^f = \left\langle x_k^t / m_{k-1}^o \right\rangle$	The *forecast*: an estimate of x_k^t given measurements up to time $k-1$

The observations m_k^o are related linearly to the true state x_k^t through an *observation operator* H_k via the equation

$$m_k^o = H_k x_k^t + \varepsilon_k^o \tag{1}$$

$$\varepsilon_k^o = \varepsilon_k^m + \varepsilon_k^r \tag{2}$$

where ε_k^0 is the observational error which is composed of the *measurement error*, ε_k^m, and a *representativeness error*, ε_k^r. The latter is due to the discretization in time and space of the solution for the ionospheric state. We relate the true state at time *k*+1 to the true state at time k via the following equation

$$x_{k+1}^t = \Psi_k x_k^t + \varepsilon_k^q \tag{3}$$

where Ψ_k is a *forward model* and ε_k^q is a *process noise* which reflects our uncertainty in the forward model.

Let M_k, R_k *and* Q_k be the measurement, representativeness and process noise covariances, respectively. Then the Kalman filter can be summarized by the following set of Equations [*Bierman 1977*]:

$$x_k^a = x_k^f + K_k\left(m_k^o - H_k x_k^f\right) \tag{4}$$

$$K_k = P_k^f H_k^T \left(H_k P_k^f H_k^T + R_k + M_k\right)^{-1} \tag{5}$$

$$P_k^a = P_k^f - K_k H_k P_k^f \tag{6}$$

$$x_{k+1}^f = \Psi_k x_k^a \tag{7}$$

$$P_{k+1}^{f} = \Psi_k P_k^a \Psi_k^T + Q_k \qquad (8)$$

K is known as the Kalman gain, P^a and P^f are the analysis and forecast covariances, respectively.
The vector $\left(m_k^o - H_k x_k^f\right)$ is known as the *innovation* vector, and it represents the observation vector minus the predicted observations based on the forecast state.

In the data assimilation process, during a given time step (indexed by k in Eqs. 1-8) the state is assumed constant (time steps are taken to be 12 minutes in our analysis below). According to the Kalman formalism, at time $k = 0$, given a forecast (a priori) state, x_0^f, a forecast state covariance, P_0^f, and a set of observations, m_0^O (collected between time 0-12 min.) with covariances R_k and M_k, then an improved estimate of the state (x_0^a, *the analysis*) at time 0 can be obtained by adding the innovation vector operated upon by the Kalman gain to the forecast state (Eq. 4). Moreover, because of the inclusion of the data during this time step, the forecast state covariance is reduced by the second term on the RHS of Eq. (6) to give the analysis state covariance at time 0. Using a dynamical model of the ionosphere, we can then propagate the state from the first time step (0 min.) to the next one (12 min.) via the *forward model*, Ψ_k. (Eq. 8). Similarly, we can propagate the analysis state covariance to the next time step via Eq. 8. The process noise, Q_k, in Eq. 8 reflects our uncertainty in the forward model. The propagated state and covariance serve as the forecast for the next time step (12 min.) and the process repeats recursively.

The above process assumes that the measurements are linearly related to the data via the observation operator (Eq. 1) and that the forward model is a linear operator (Eq. 7). As discussed below, both of these assumptions are satisfied for the type of data and physical model we are dealing with. For the Kalman filter to be an unbiased, maximum likelihood, minimum variance (and therefore optimal) estimator, the measurements and state errors need to follow Gaussian statistics and be unbiased. In that case, it is possible to show that the Kalman filter estimator (x_k^a) also minimizes the cost functional [*Bierman*, 1977]

$$J_k = \sum \left(m_k^o - H_k x_k^t\right)\left(R_k + M_k\right)^{-1}\left(m_k^o - H_k x_k^t\right)^T + \left(x_k^t - x_k^f\right)P_k^{f\,-1}\left(x_k^t - x_k^f\right)^T \qquad (9)$$

where the sum is over all the measurements during step k. This equality can be used to check the consistency of our assumptions on the magnitude of the state and measurement covariances.

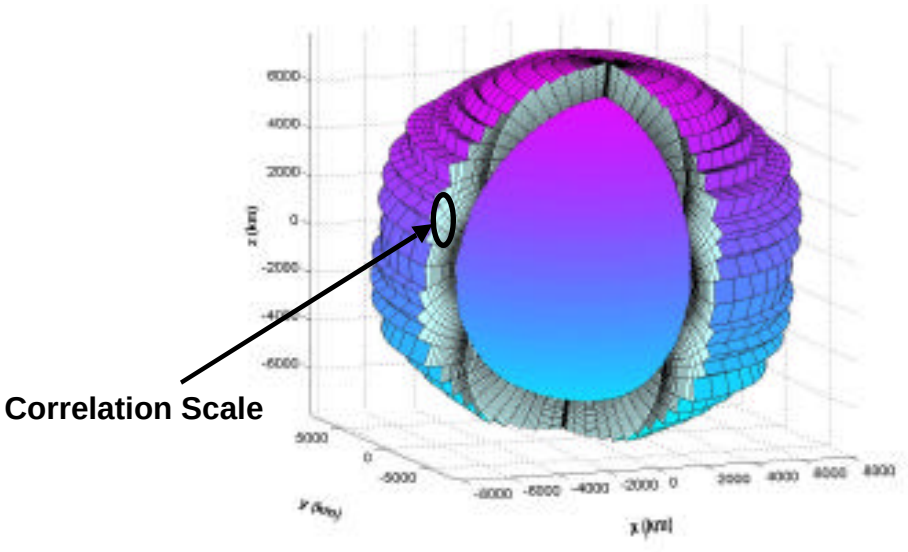

Figure 1. The grid used in modeling the ionosphere representing a Eulerian frame divided along constant geomagnetic field lines, constant geomagnetic potential lines, and constant longitudes. The ellipsoid represents the "correlation volume" used in setting the correlation between neighboring elements for the band-limited Kalman. Elements inside the volume have zero correlation with elements outside the volume.

Approximations to the Kalman filter

Because of the large dimension of the state (i.e., the number of volume elements or voxels), the full Kalman filter is not computationally feasible. Several approximations to the Kalman filter have been implemented including an "approximate Kalman", optimal interpolation, and a band-limited Kalman. In the approximate Kalman, the operation in Eq. (6) is skipped. In optimal interpolation, the operations in Eqs. (6) and (8) are skipped, therefore the state covariance is taken to be constant as a function of time. In the band-limited approximation, all steps from (Eqs.

4-8) are performed, but the state covariance is truncated such that only neighboring voxels within a specified correlation volume have non-zero correlations (Fig. 1). The advantage of the band-limited Kalman over optimal interpolation is that it maintains the time evolution of the state covariance. A realistic representation of the state covariance is paramount for obtaining accurate estimates of the state, especially when the data are sparse relative to the size of the state. The band-limited Kalman filter maintains a sensible covariance while at the same time it reduces the number of computational steps substantially, thereby making it usable for global, medium-resolution, ionospheric runs (Fig. 2).

The number of operations required by these various approximations is indicated in Figure 3. It is obvious from the figure that for a global ionospheric model with a medium grid resolution in latitude, longitude and height (a total of ~13,000 elements), both the full Kalman and the approximate Kalman cannot be computed in a reasonable amount of time. On the other hand, the band-limited Kalman computation is very manageable.

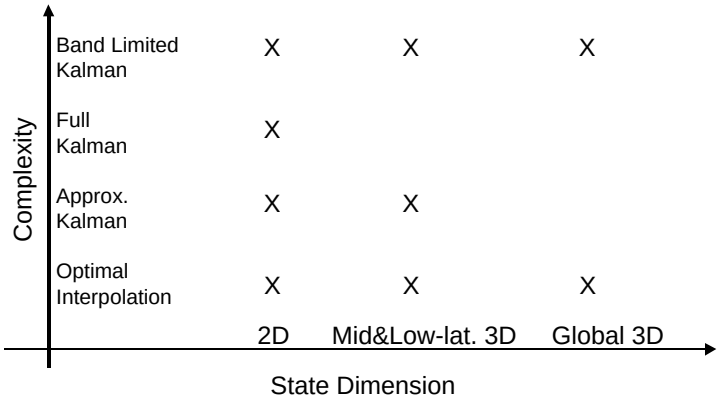

Figure 2. Degree of complexity as a function of the state dimensions. The figure indicates that the full Kalman is practical only for 2-dimensional models of the ionosphere, the approximate Kalman is practical for 2-D and mid- and low-latitude 3-D models, and the optimal interpolation and band-limited Kalman can be used for the 2-D and 3-D global ionospheric model.

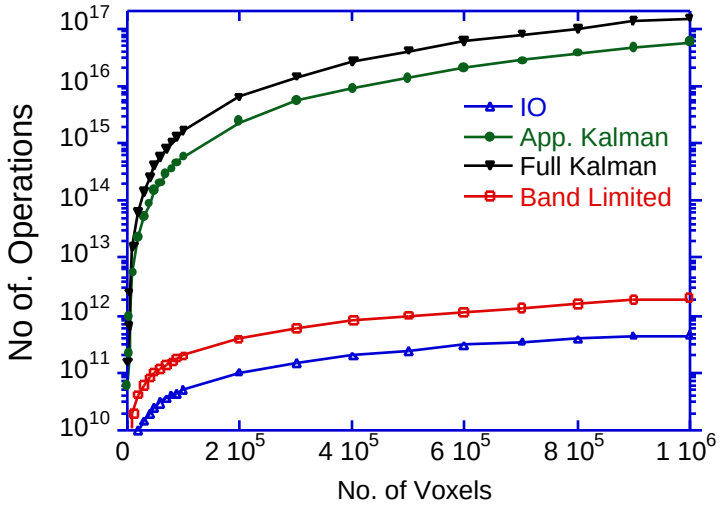

Figure 3. Number of operations required for a 24-hour assimilation run using 100,000 TEC measurements plotted for the different Kalman approximations as a function of the number of volume elements in the ionospheric model grid. On a DEC alpha workstation, it takes nearly 6 hours of real time to perform ~10^{12} operations.

3. Forward Model

Our physical model is based on a simplified version of the Sheffield University Plasmasphere/Ionosphere Model (SUPIM) [Bailey *et al.*, 1993] for the middle and low-latitude regions. We solve the conservation of mass and momentum equations for a plasma, which account for production, loss, transport of the major ionization specie in the F-region (O^+), neutral wind, and electric forcing. These equations can be written as

$$\frac{\partial n}{\partial t} + \quad (n\mathbf{V}) = (P - L) \tag{10}$$

$$- \left(nk_B T\right) + nM\mathbf{g} + cn\left(\mathbf{E} + V \times \mathbf{B}\right) - nM\nu\left(V - \mathbf{U}\right) = 0 \qquad (11)$$

where n is the ion number density; **V** is velocity; P and L are production and loss rates respectively; k_B is Boltzmann's constant; T is temperature, M is molecular mass, **g** is gravitational acceleration, **E** and **B** are electric and magnetic fields respectively; ν is the collision frequency for momentum transfer between the atomic oxygen ion and the neutral particles; and **U** is neutral wind. An equation similar to Eq. (11) can be obtained for the electrons, and after ignoring terms that are multiplied by the electron's mass, we obtain

$$- \left(n_e k_B T_e\right) - en_e\left(\mathbf{E} + V_e \times \mathbf{B}\right) = 0 \qquad (12)$$

In addition we also have

$$n_e = n, \qquad n_e V_e = nV \qquad (13)$$

We solve for the ion and electron densities by solving the above equations and making use of the empirical or parameterized models of the thermosphere (MSIS), thermospheric winds (HWM), solar EUV (SERF2, [Tobiska, 1991]), and electric fields [e.g., Heppner and Maynard, 1987; Fejer and Scherliess, 1997].

Traditionally, the dynamical equations are rewritten in a moving Lagrangian coordinate frame [Bailey *et al*., 1993]. The motion of this coordinate frame is dictated by the plasma drift perpendicular to the geomagnetic field lines. This approach introduces significant computational efficiency by transforming a time-dependent partial differential equation in 3-dimensional space into a family of time-dependent ordinary differential equations in a 1-dimensional space following the magnetic field lines. However, this 1-D approach introduces a significant difficulty for data assimilation, since the measurements are taken in 3-D space from ground-to-satellite and satellite-to-satellite links, making the mapping between data and the model parameter space very difficult to construct. We will elaborate on this point further when we discuss the construction of the observation operator, *H*, the time propagation of the state, and the time propagation of the state covariance matrix.

Because of these considerations, we solve Eqs. (10-13) in a fixed Eulerian frame. The solution grid uses a *p-q-l* coordinate system [*Kendall*, 1962; *Bailey et al*., 1993] in which the voxels are bounded by constant geomagnetic field lines, constant geomagnetic potential lines, and longitudes, where the **B** field is defined by an eccentric tilted dipole. A 3-D view of the solution grid appears in Figure 1, and the details of the grid resolution are given in Table 1 below.

4. Results and Validation

The USC/JPL GAIM model is now able to assimilate four major data types: unbiased slant TEC measurements from ground GPS receivers, change in TEC data from GPS occultations, in situ electron density measurements, and UV airglow radiances. In this paper we only present results from a GAIM assimilation of ground GPS-based TEC measurements using the band-limited Kalman filter on an example day, June 13, 2001. For this day, 183,000 unbiased TEC measurements were computed from 98 GPS receiver sites using an elevation cutoff of 10 degrees. The GPS instrumental biases were determined using the JPL GIM technique [Mannucci, *et al*., 1998]. Table 1 summarizes the specifications for the geomagnetic conditions, GAIM grid, physics drivers, and input dataset. The GAIM assimilation run yielded a 3-D specification of electron density every 12 minutes during the day, which can be integrated through to create a global 2-D map of vertical TEC or to compute slant TEC predictions for arbitrary ray paths.

Date	June, 13, 2001
Modeled region longitude range	0-360 deg.
Modeled region latitude range	-85 to 85 N.
Modeled region altitude range	100-1500 km altitude
Latitude resolution	5 deg.
Longitude resolution	15 deg.
Altitude resolution	80 km
Total no. of volume elements (voxels)	13,107
Vertical drift pattern	Climatology for June, Solar Max
F10.7	186
Average F10.7	165
Kp index	8
No. of TEC data points	183,000 over 24 hours

Table 1. Detailed specifications for the grid, physics input and data used in the data assimilation run.

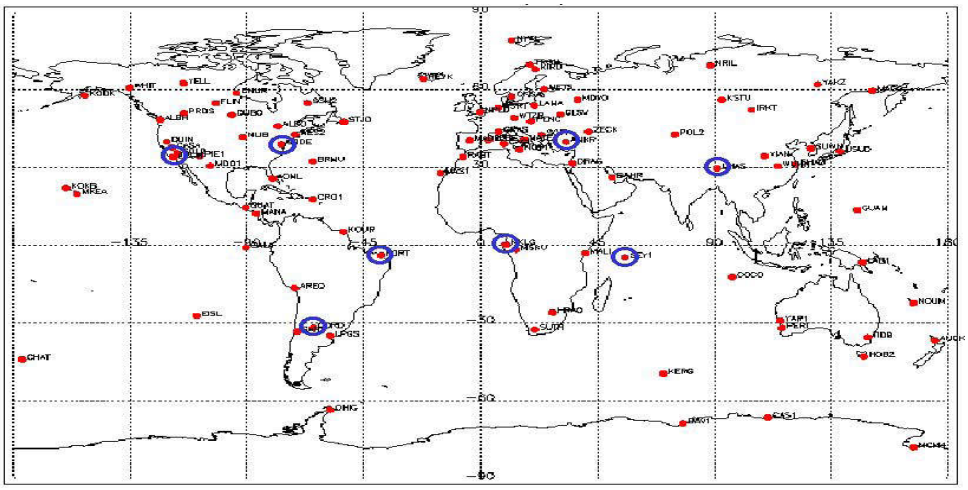

Figure 4. A map of the 98 ground GPS stations used in the assimilation. The eight "circled" sites were withheld from a second GAIM run and used to validate the accuracy of slant TEC prediction.

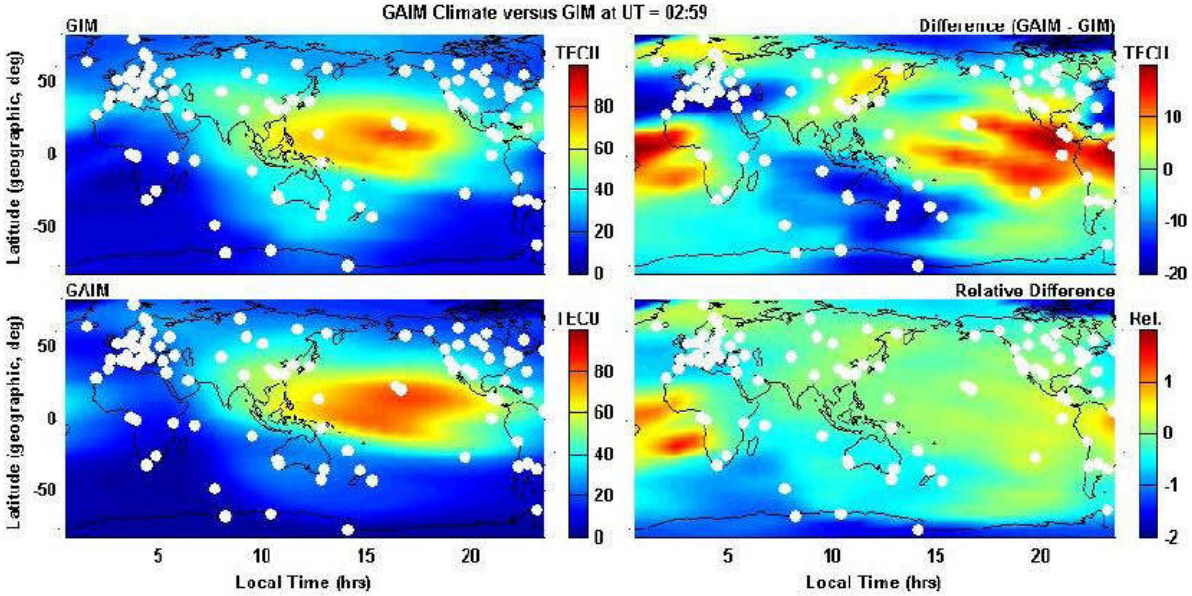

Figure 5. GAIM "climate" versus GIM vertical TEC at 3 UT on 06/13/2001.

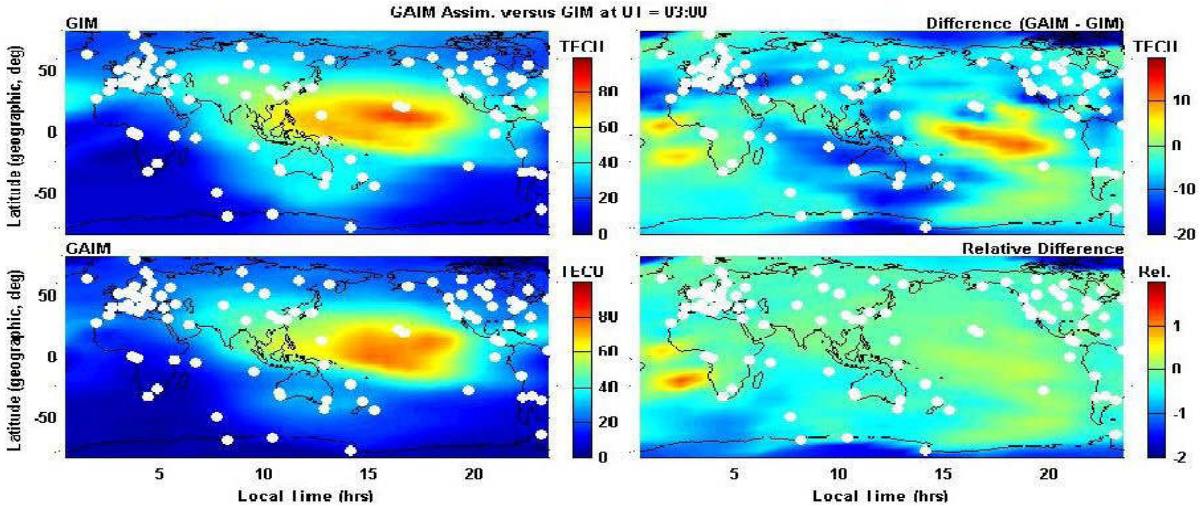

Figure 6. GAIM "assimilation" versus GIM vertical TEC at 3 UT on 06/13/2001.

The GAIM specifications of electron density were validated in four ways: (1) comparison of the GAIM vertical TEC maps (as an animated movie) to GIM maps computed from the same ground GPS TEC data, (2) comparison of GAIM densities and peak heights to independent NmF2 and HmF2 measurements from the global network of ionosondes, (3) comparison of GAIM vertical TEC values to independent TOPEX measurements, and (4) comparison of GAIM slant TEC predictions to measurements from independent GPS receivers. Figure 4 shows a map of the 98 GPS receiver sites with circles around the 8 independent validation sites. Due to space constraints, only a few examples of the validation results can be presented here. Results for the ionosonde comparisons appear in a companion paper [Wang *et al*., 2002].

Figure 5 shows snapshots of vertical TEC from the GAIM forward model ("climate" run with no data assimilated) and GIM, along with maps of the absolute difference and relative difference. Since the GIM map of vertical TEC is a straightforward interpolation of the GPS TEC data using a 2D shell model, it can serve as an approximate "proxy" for the information content of the GPS dataset. GIM matches the TEC data (mapped to vertical) quite well near the GPS sites, but its interpolation is less accurate at distances greater than 1000 km from the nearest site. Figure 5 illustrates that the GAIM vertical TEC level before data assimilation significantly differs from reality (or at least the GIM proxy). Figure 6 shows the analogous vertical TEC maps for GAIM versus GIM after the GPS data were assimilated into GAIM. The absolute and relative differences between GAIM and GIM are significantly reduced near the GPS sites, indicating that the TEC data are being used effectively by the GAIM Kalman filter.

The validations using TOPEX measurements also show that assimilating the TEC data can dramatically improve the GAIM density specification and resulting vertical TEC. Figure 7 shows a comparison of vertical TEC from the GAIM climate run, the GAIM assimilation run, GIM, and IRI95 versus TOPEX measurements. For the seventh TOPEX pass on this day, the northern anomaly peak happens to fall near the GPS site at Hawaii and one sees that the GAIM assimilation result (vertical TEC of 61 TECU at the peak) matches the TOPEX measurements (and GIM) much better than the GAIM climate prediction (too high, 73 TECU at the peak). Using the TEC data, the GAIM assimilation improves on the GAIM climate in matching the vertical TEC at the northern anomaly peak and following the gradient toward mid-latitudes. There are no GPS sites near the southern anomaly peak and as a result one sees that the GAIM climate and assimilation results are almost identical in this latitude region, and neither the GAIM TEC nor the GIM TEC values follow the structure of the southern anomaly peak. In terms of overall RMS difference over the TOPEX track, the GAIM climate prediction is worse than the IRI95 model (10.1 versus 9.2 TECU) and the GAIM assimilation (7.1 TECU) is better than IRI95 but worse than GIM (4.9 TECU).

A figure of merit for any ionosphere model is its accuracy in predicting slant TEC for arbitrary ray paths. To assess the accuracy of GAIM's slant TEC predictions we performed two assimilation runs: one using 98 GPS sites and a second using 90 sites, withholding 8 as independent validation sites. One can then predict the slant TEC in three ways — from the GAIM climate, from the 98-site assimilation (corresponds to postfit residuals), and from the 90-site assimilation (evaluating prediction accuracy) — and compare it to the independent slant TEC measurements. Figure 8 shows the differences from the observed slant TEC at the JPLM site for the three cases. The GAIM "climate" prediction of slant TEC for the JPLM ray paths exhibits maximum errors of 120 TECU with a slant RMS error of 26.3 TECU over the day. The GAIM assimilation prediction accuracy (RMS error = 4.5 TECU) is not significantly worse than the postfit residuals (RMS = 3.2 TECU), which is reasonable since there are data sites near JPLM and the mid-latitude ionosphere can be smooth. Prediction accuracy is not as good in the equatorial region. Overall, the 8-site analysis of slant TEC prediction accuracy indicates that GAIM errors in the current data-sparse regime are a function of the distance to the nearest GPS site and the regional correlation length of the ionosphere.

5. Conclusions

The USC/JPL GAIM model uses a first-principles physics model of the ionosphere and a band-limited Kalman filter to assimilate multiple types of ionospheric measurements. Although one could use 2-D and 3-D tomographic inversion techniques to "image" the ionosphere using a single data type, such as GPS occultations or UV airglow radiances, formal data assimilation techniques are required to properly handle and combine multiple data types with each other and with model information from the physics-based forward model. Future GAIM work will include the simultaneous input of multiple data types into assimilation runs and the development of an operational scheme in which 4DVAR and the band-limited Kalman filter are used together to assimilate all available data and solve for both the density state and key physical drivers (e.g., **E** x **B** vertical drifts and neutral winds).

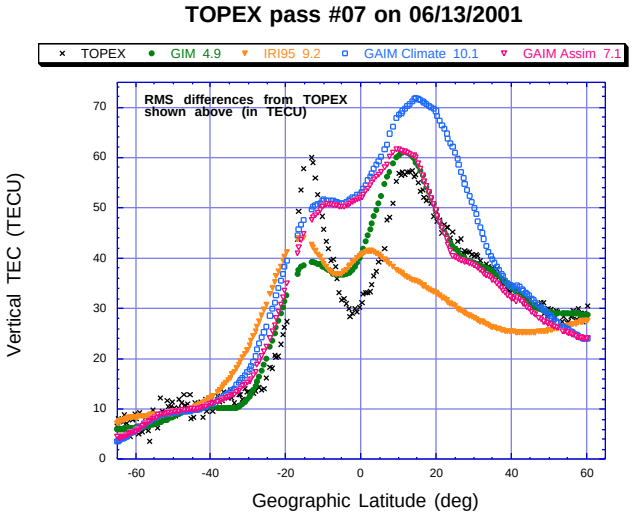

Figure 7. Comparison of vertical TEC from GAIM, GIM, and IRI to TOPEX measurements.

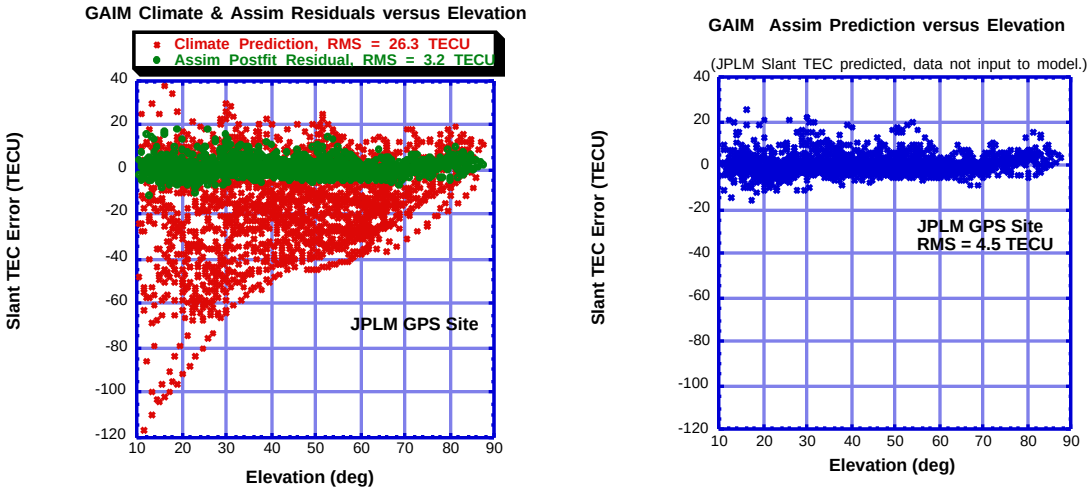

Figure 8. Differences between slant TEC observations from the site JPLM and GAIM predictions plotted versus elevation for three cases: climate (RMS error = 26.3 TECU), assimilation including JPLM data (postfit RMS residual = 3.2 TECU), and assimilation excluding JPLM data (prediction RMS error = 4.5 TECU).

Acknowledgements

We thank Graham Bailey and Arthur Richmond for their assistance in the effort of building GAIM. This work is supported by the Department of Defense through a Multiple-disciplinary University Research Initiative. The research conducted at the Jet Propulsion Laboratory is under a contract with NASA.

References

Bailey, G. J., R. Sellek, and Y. Rippeth, *Ann. Geophy.*, **11**, 263-272, 1993.

Bierman, G. J., *Factorization Methods for Discrete Sequential Estimation*, Academic Press, 1977.

Fejer, B. G., L. Scherliess, *J. Geophy. Res.*, **102**, 24047.

Hajj G. A., L. C. Lee, X. Pi, L. J. Romans, W. S. Schreiner, P. R. Straus, C. Wang, "COSMIC GPS ionospheric sensing and space weather", Terrestrial, atmospheric and oceanic sciences, **11**, 1, pp 235-272, Mar. 2000.

Heppner, J. R., and N. C. Maynard, *J. Geophys. Res., 92*, 4467-4489, 1987.

Kendall, P. C., *J. Atmos. Terr. Phys.*, *24*, 805-811, 1962.

Mannucci A., B. Wilson, D. Yuan, C. Ho, U. Lindqwister, and T. Runge, *Radio Science*, **33**, 3, 565-582, 1998.

Pi X., A. J. Mannucci, U. J. Lindqwister, and C. M. Ho, *Geophy. Res. Lett.*, **24**, 18, pp. 2283-2286, 1997.

Pi, X., C. Wang, G. A. Hajj, I. G. Rosen, and B. D. Wilson, "Observation system simulation experiments using a global assimilative ionospheric model", *Proc. of Beacon Symposium*, Boston, MA, June, 2001*b*.

Tobiska, W. K., *J. Atmos. Terr. Phys.*, *53*, 1005-1018, 1991.

Wang, C., G. Hajj, X. Pi, I. G. Rosen, and B. Wilson, "A Review of the Development of a Global Assimilative Ionospheric Model", Proceedings of Ionosphere Effects Symposium, Alexandria, VA, May 2002.