

Data assimilation of ground GPS total electron content into a physics-based ionospheric model by use of the Kalman filter

G. A. Hajj,^{1,2} B. D. Wilson,¹ C. Wang,³ X. Pi,^{1,2} and I. G. Rosen³

Received 12 December 2002; revised 27 June 2003; accepted 23 July 2003; published 24 February 2004.

[1] A three-dimensional (3-D) Global Assimilative Ionospheric Model (GAIM) is currently being developed by a joint University of Southern California and Jet Propulsion Laboratory (JPL) team. To estimate the electron density on a global grid, GAIM uses a first-principles ionospheric physics model and the Kalman filter as one of its possible estimation techniques. 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. Of the possible suboptimal implementations of the Kalman filter, we have chosen a band-limited Kalman filter where a full time propagation of the state error covariance is performed, but it is always kept sparse and banded. The effectiveness of ground GPS data for specifying the ionosphere is assessed by assimilating slant total electron content (TEC) data from 98 sites into the GAIM Kalman filter and validating the electron density field against independent measurements. A series of GAIM analyses are presented and validated by comparisons to JPL's global ionospheric maps (GIM) of vertical TEC (VTEC) and measurements from TOPEX. A statistical evaluation of GAIM and GIM against TOPEX VTEC indicates that GAIM accuracy is comparable or superior to GIM. *INDEX TERMS*: 2447 Ionosphere: Modeling and forecasting; 3337 Meteorology and Atmospheric Dynamics: Numerical modeling and data assimilation; 2494 Ionosphere: Instruments and techniques; 3210 Mathematical Geophysics: Modeling; 3260 Mathematical Geophysics: Inverse theory; *KEYWORDS*: ionosphere, forecast, data assimilation, Kalman filter

Citation: Hajj, G. A., B. D. Wilson, C. Wang, X. Pi, and I. G. Rosen (2004), Data assimilation of ground GPS total electron content into a physics-based ionospheric model by use of the Kalman filter, *Radio Sci.*, 39, RS1S05, doi:10.1029/2002RS002859.

1. Introduction

[2] 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. Space weather can effect power grids at middle and high-latitudes, disrupt communication systems, and degrade the performance of navigation and reconnaissance 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.

[3] 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 the main factors that have brought numerical weather prediction (NWP) models to their current level of success.

[4] On the space environment front, we are witnessing a new era. Significant efforts are being planned to collect further information on solar activities and disturbances in the magnetosphere, and data on the upper

¹Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, USA.

²Now at Department of Mathematics, University of Southern California, Los Angeles, California, USA.

³Department of Mathematics, University of Southern California, Los Angeles, California, USA.

atmosphere and ionosphere/plasmasphere are becoming truly global and continuous. A case in point is the Global Positioning System (GPS), in which a global network of over 100 ground receivers and regional networks of hundreds to over 1000 receivers created the unprecedented possibility of producing global maps of vertical total electron content (TEC) and ionospheric irregularities in near-real time updated subhourly [Pi *et al.*, 1997; Mannucci *et al.*, 1998]. Moreover, within the next few years the number of flight receivers tracking GPS in a limb-viewing geometry for ionospheric occultations [Hajj *et al.*, 1994; Leitingner *et al.*, 1997; Hajj and Romans, 1998; Schreiner *et al.*, 1999] will increase to nearly a dozen, providing an extremely dense global set of horizontal cuts through the ionosphere and allowing for accurate 4-D global mapping of electron density [Hajj *et al.*, 2000]. This data set, along with other data such as UV airglow radiances [e.g., Dymond *et al.*, 2001] from current and future missions, provide a truly unprecedented global coverage of the upper atmosphere and ionosphere.

[5] A long-term objective of our research is to develop, validate, and use in operational and research modes a Global Assimilative Ionospheric Model (GAIM) capable of assimilating a variety of data types including: (1) slant TEC (the integral of electron density along the transmitter-receiver line-of-sight) 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 which are related to the state in a nonlinear manner. Similar to neutral atmospheric weather models (which assimilate, solve for, and predict 4-D 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 underlying forcing functions (“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-D variational (4DVAR) technique. The former technique is used to solve for the electron density in space and time without attempting to solve for or adjust the “drivers.” 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 where the Kalman filter is used to estimate the initial electron densities, while the 4DVAR is used to estimate the drivers and to produce a prediction for the next data assimilation cycle. The

4DVAR technique is described elsewhere [Rosen *et al.*, 2001; Pi *et al.*, 2003]. Background description of GAIM can also be found in the work of Hajj *et al.* [2000] and Wang *et al.* [2004]. Other background information on the use of stochastic inverse theory and the Kalman filter to ionospheric mapping can be found in the work of Fremouw *et al.* [1992] and Howe *et al.* [1998, and references therein].

[6] 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, 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 or negative ways. Therefore optimal assimilation of any data type requires careful tuning of its error covariance, proper evaluation of the data representation 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.

[7] The paper is organized as follows. In section 2 we review the formulation of the Kalman filter. In section 3 we discuss some practical considerations related to the full Kalman filter, such as memory requirements and number of operations, and introduce the band-limited Kalman filter. In section 4 we describe the University of Southern California (USC)/Jet Propulsion Laboratory (JPL) GAIM physics model and its solution grid. In section 5 we present examples of ionospheric specifications from GAIM analyses for 22–24 May 2002 and validation results against global ionospheric maps (GIM) and TOPEX. A conclusion is given in section 6.

2. Kalman Filter

[8] We introduce the following definitions (commonly used in NWP) [e.g., Ghil and Malanotte-Rizzoli, 1991; Daley, 1991]:

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

[9] The observations m_k^o are assumed to be related linearly to the true state x_k^f through an observation operator H_k via the equations

$$m_k^o = H_k x_k^f + \varepsilon_k^o \quad (1)$$

$$\varepsilon_k^o = \varepsilon_k^m + \varepsilon_k^r, \quad (2)$$

where ε_k^o is the observational error, which is composed of the measurement error, ε_k^m , and a representation error, ε_k^r . The latter is due to the discretization in time and space of the solution for the ionospheric state (for a description of TEC representation error, see *Hajj et al.* [2000]). For TEC measurements the relation between the observations and the state is already linear. A linearization procedure might be required to relate the true state at time $k+1$ to the true state at time k , which can then be written in the form

$$x_{k+1}^f = \Psi_k x_k^f + \varepsilon_k^q, \quad (3)$$

where Ψ_k is a forward model, which can be represented in a matrix form and ε_k^q is a process noise which reflects our uncertainty in the forward model. A linearization procedure is not required in our case since our dynamical model is already linear, as we shall explain later.

[10] If M_k , R_k , and Q_k are used to denote the measurement, representation, and process noise covariances, respectively, then the Kalman filter can be summarized by the following set of equations:

$$x_k^a = x_k^f + K_k (m_k^o - H_k x_k^f) \quad (4)$$

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

$$P_k^a = P_k^f - K_k H_k P_k^f \quad (6)$$

$$x_{k+1}^f = \Psi_k x_k^a \quad (7)$$

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

K is known as the Kalman gain and P^a and P^f are the analysis and forecast covariances, respectively. The vector $(m_k^o - H_k x_k^f)$ is known as the innovation vector, and it represents the observation vector minus the predicted observations based on the forecast. The Kalman filter was first introduced by *Kalman* [1960] and *Kalman and Bucy* [1961] for linear systems of ordinary differential equations. An overview of the use of the Kalman filter for meteorology can be found in the work of *Ghil and Malanotte-Rizzoli* [1991] and *Daley* [1991].

[11] In the data assimilation process, during a given time step (indexed by k in equations (1)–(8)) the state is

assumed to be constant (time steps are taken to be 12 min in our analysis below). According to the Kalman formalism, at time t_0 (the center of the first time interval), given a forecast (initial) state, x_0^f , a forecast state error covariance, P_0^f , and a set of observations, m_0^o (collected in the interval $t_0 - \Delta t$, $t_0 + \Delta t$; $\Delta t = 6$ min in our case) with covariances R_0 and M_0 , 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 (equation (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 right-hand side (RHS) of equation (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) by use of the forward model, Ψ_k (equation (7)). Similarly, we can propagate the analysis state covariance to the next time step by use of equation (8). The propagated state and covariance serve as the forecast for the next time step (12 min), and the process repeats recursively.

[12] The process noise, Q , in equation (8) reflects our uncertainty in the forward model and forms one of the most crucial input to the Kalman filter. If Q is too small, then P_0^a and P_0^f will become too small as more data points are assimilated. This would eventually result in giving unjustifiably larger weight to the model causing the data to have no impact. Inversely, if Q is too large, the information in P_0^a , which carries in it our knowledge of the state based on the physics and previous data, would be lost. Choosing the correct Q requires much care in understanding the limitations of the model under different physical conditions (e.g., during magnetically quiet or disturbed periods); furthermore, for Q to be optimal, it must have the right information regarding the correlation between the state elements. As a first step, for the analysis shown in the subsequent sections, we choose an ad hoc Q_k with diagonal elements $\sigma_i^2 = (10^{10} + 0.2 \times N_i)^2$, where N_i is the electron density in voxel i in units of e/m^3 . The additive term ensures that voxels with small electron densities have error bars that are not too small. The multiplicative term serves the same purpose but for voxels with large electron densities. At this point we offer no justification for this choice of Q_k other than to say that after testing several additive and multiplicative values, the ones listed here seemed to give better postfit residuals and better agreement of the analysis with independent data.

[13] The process of assimilating data continuously and propagating the model at each time step in the manner described above is formally known as continuous data assimilation or 4-D data assimilation (not to be confused with 4DVAR) [*Daley*, 1991]. In continuous data assimilation the philosophy is that even if the initial condition and/or the model are imperfect, the accumulation of data will gradually force the model integration to the true

ionospheric state. In continuous data assimilation the analysis at time t_k depends on all observations taken at $t < t_k$. However, it is also possible to include measurements taken at $t > t_k$ when estimating the state at time t_k by use of 4DVAR and/or Kalman smoothing [see, e.g., Ghil *et al.*, 1997].

[14] For the Kalman filter to be an unbiased, maximum likelihood, minimum variance 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^o) also minimizes the cost functional [Bierman, 1977]

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

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 error covariances.

3. Approximations to the Kalman Filter

[15] Since one of the main purposes of ionospheric data assimilation is to produce an ionospheric specification or prediction that is useful for space weather applications, timeliness, where the analysis can keep pace with the data, is a key factor for a practical implementation of the Kalman filter. Because of the large dimension of the state (i.e., the number of volume elements or voxels used to represent the ionospheric state which is of order $N = 10^5$ to 10^6), the full Kalman filter may not be computationally feasible in a timely manner. This is true because of memory storage limitations and the number of computations required. Saving the state covariance in memory requires saving N^2 double precision numbers or 80–8000 Gb. However, updating $P_k^f \rightarrow P_k^a$ (equation (6)) when assimilating M TEC measurements requires of order $M \times N^2$ operations, where each operation is defined as the time needed to extract three double precision numbers, B , C , and D , from high-speed storage, the evaluation of $A = B + C \times D$, and the transfer of A to high-speed storage. (The same covariance update requires $M \times N$ operations when assimilating *in situ* measurements.) Furthermore, updating $P_k^a \rightarrow P_{k+1}^f$ (equation (8)) requires of order N^3 operations. However, the latter transformation (equation (8)) can be made of order $c_1 N^2$ operations, where c_1 is roughly constant, by taking advantage of the fact that diffusion takes place along magnetic flux tubes and using a common grid to solve for the dynamical equations of the ionosphere and to solve for the ionospheric state in the Kalman filter as described later. In our implementation of GAIM and for a time step of 12 min, c_1 is of order 1000.

[16] To appreciate the level of computations needed to perform the full Kalman estimation, consider the following example. A subset of 98 GPS stations from the continuously operational global network operated by the International GPS Service (IGS) collects nearly 700 five-min-averaged line-of-sight TEC measurements every 12 min. Assimilating these measurements and updating the state covariance every 12 min requires of order $(700 + 1000) \times N^2$ operations. An Intel chip with 2 GHz speed at best performs only 2×10^9 operations (as defined above) per second. For $N = 10^5$ – 10^6 and for a whole day run, this translates into 2×10^{15} – 2×10^{17} operations or 12–1200 days. High-resolution operational numerical weather prediction models solve for $\sim 10^7$ – 10^8 variables (temperature, water vapor, and zonal and meridional wind components on a $1^\circ \times 1^\circ$ grid at 30–60 pressure levels). This makes it clear why the full Kalman filter is prohibitive, even on the fastest parallel computers available to date. This is also why the meteorological community has devised numerous approaches/approximations to the Kalman filter including optimal interpolation [e.g., Lorenc, 1981], partitioned Kalman, various reduced Kalman filters, and band-limited Kalman. Upon evaluation of these various options, we opted for the band-limited Kalman for reasons given below.

[17] The different options in implementing a Kalman filter present various types of difficulties. For instance, in a partitioned Kalman filter, one splits the region being estimated (the ionosphere) into numerous smaller manageable regions where the Kalman filter is applied separately, and then information is exchanged across boundaries. However, implementation of this approach presents serious problems for assimilating data integrated across large regions such as line-of-site TEC taken at low elevations and limb TEC or UV data. In a partitioned Kalman, data going across different partitions have to be discarded. An examination of the coverage of data (from ground and space) will make it clear that one will most likely end up throwing away the data that could be most valuable (e.g., GPS occultations or UV limb sounders) for sensing the vertical structure of the ionosphere.

[18] In a reduced Kalman filter one solves the forward model on a high-resolution grid but estimates the error covariance and a correction to the forecast on a coarse-resolution or “reduced” grid. The suitability of this approach depends on the amount and type of data available. If the data are sparse and localized (i.e., not integrated over the region being modeled), the reduced Kalman presents an attractive option since the assimilation of data will affect the regions being observed while having the full resolution of the forward model. However, if data are fairly dense and integrated over the region (e.g., TEC and UV), the reduced Kalman may present a limitation since, by construction, the observations will only yield structures at the coarse-resolution level.

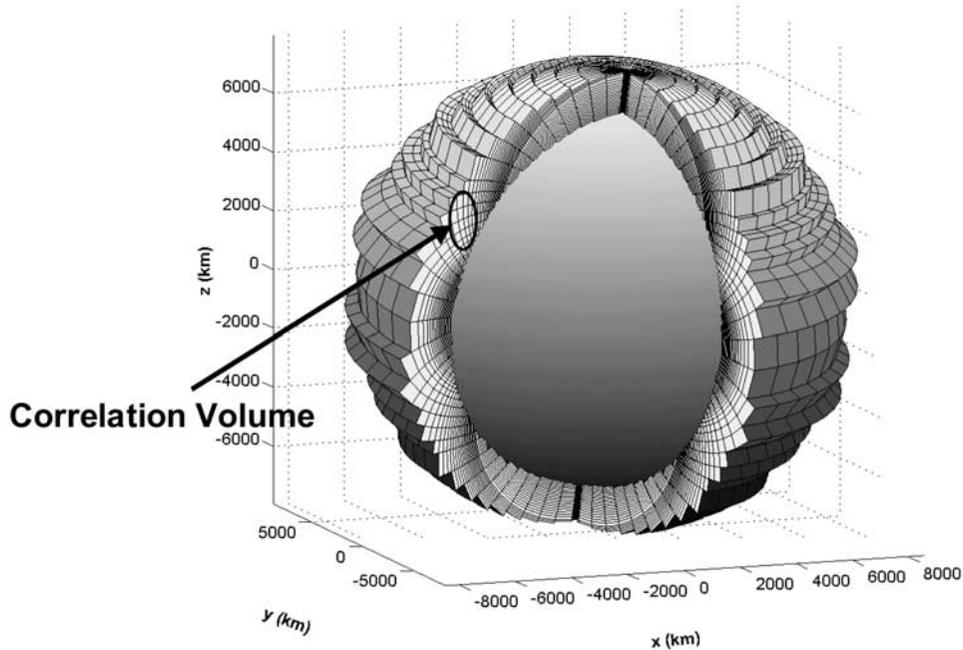


Figure 1. An example of a GAIM grid used in modeling the ionosphere, representing an Eulerian frame divided along constant geomagnetic field lines, constant geomagnetic potential lines, and constant geomagnetic longitudes. The ellipsoid represents an example of the “correlation volume” used to set the correlation between neighboring elements for the band-limited Kalman. An element centered at the ellipsoid will have zero covariance with elements outside the volume.

[19] Our choice for a band-limited implementation of the Kalman filter is driven by our desire for maximum flexibility. In the band-limited approximation, all the Kalman steps (equations (4)–(8)) are performed as usual. However, the state error covariance is truncated such that for a given voxel i , only a subset of the entire set of voxels will have nonzero correlation (i.e., $P_{ij} \neq 0$ for some preselected $j \in [1, N]$). In the simplest example of a band-limited Kalman, a given voxel i will have nonzero covariance only with voxels within a specified “correlation volume” as depicted in Figure 1. In a more complicated example the nonzero correlation terms associated with a voxel i could be all the voxels that are along the same or neighboring magnetic field lines as voxel i , therefore accounting for the strong coupling normally observed along magnetic field tubes. Our implementation is very general where the user can arbitrarily specify the “volume of correlation” associated with any voxel i . In the limit when the volume of correlation for each element covers the entire region being modeled, the band limited becomes the same as the full Kalman filter.

[20] It is worth noting that the term “band-limited” is strictly valid only for a 1-D system, where the elements can be indexed such that the state covariance matrix is zero everywhere outside a finite band around the diag-

onal. However, for higher-dimensional systems, the covariance matrix will always have nonzero elements away from the diagonal terms (Figure 2). Therefore, when referring to a band-limited Kalman, the term is used only in some abstract sense where in fact the actual state covariance is full, albeit very sparse.

[21] The band-limited Kalman reduces the number of operations required to update the state error covariance matrix from N^2 to $A \times N$, where A is the number of voxels within the correlation volume, V_{corr} . Since, for a fixed V_{corr} , A grows linearly with N , the number of operations required to update the covariance is given by $(V_{\text{corr}}/V) \times N^2$, where V is the volume of the entire region being modeled. For instance, if the correlation volume is chosen to extend over 10 degrees in longitude, 10 degrees in latitude, and 100 km in altitude, and the total volume modeled covers the globe with a height span between 100 and 1600 km, then V_{corr}/V would equal $(20/360) \times (20/180) \times (100/1500) \approx 1/10^4$, making the band-limited 4 orders of magnitude faster than the full Kalman and reducing it to a manageable size. 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

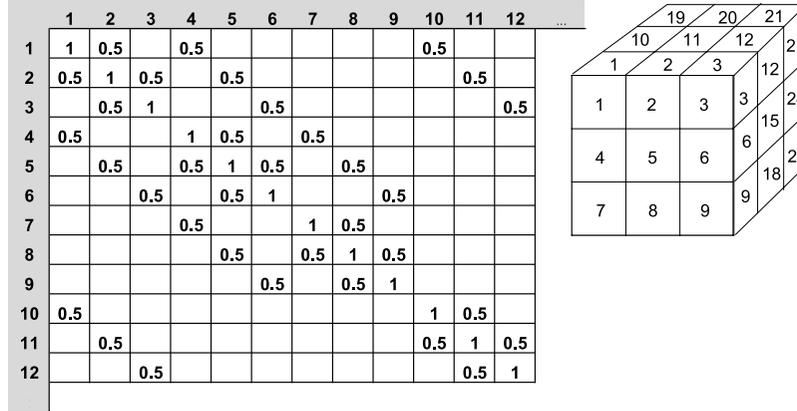


Figure 2. An example of a covariance matrix for a 3-D structure with nonzero correlation between immediate neighbors.

computational steps substantially, thereby making it usable for global, medium-resolution, ionospheric runs.

4. Forward Model

[22] A detailed description of the GAIM physical model is given in the work of *Pi et al.* [2003]. In summary, we solve the conservation of mass and momentum equations for a plasma, which account for production, loss, and transport of the major ionization species in the F region (O^+). These equations can be written as

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

$$-\nabla(nk_B T) + nM\mathbf{g} + cn(\mathbf{E} + \mathbf{V} \times \mathbf{B}) - nM\nu(\mathbf{V} - \mathbf{U}) = 0, \quad (11)$$

where n is the ion number density; \mathbf{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; \mathbf{g} is gravitational acceleration; c is speed of light; \mathbf{E} and \mathbf{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 \mathbf{U} is neutral wind. An equation similar to equation (11) can be obtained for the electrons, and after ignoring terms that are multiplied by the electron's mass, we obtain

$$-\nabla(n_e k_B T_e) - en_e(\mathbf{E} + V_e \times \mathbf{B}) = 0. \quad (12)$$

In addition, we also have

$$n_e = n, \quad n_e V_e = nV. \quad (13)$$

[23] The ion and electron densities are obtained by solving the above equations and making use of the

empirical or parameterized models of the thermosphere (mass spectrometer and incoherent scatter radar (MSIS)) [*Hedin*, 1991], thermospheric winds (Hedin wind model (HWM)) [*Hedin et al.*, 1996], solar EUV (SERF2) [*Tobiska*, 1991], and electric fields [e.g., *Heppner and Maynard*, 1987; *Fejer et al.*, 1991]. Given all the driving forces, it should be clear that equations (10)–(13) are linear in the ion and electron densities. This linearity is broken once more ions are introduced or the conservation of energy equation is added. Using a single-ion model and not solving for the energy balance equation is not uncommon and is used by many ionospheric models. The simplifications and speed offered by the linearity of such a model may well justify its use for data assimilation purposes. One should keep in mind that, in the presence of data, the estimation process will still yield accurate solution of electron densities, even in regions where the background given by the model maybe in error.

[24] Traditionally, these dynamical equations are rewritten in a moving Lagrangian coordinate frame [e.g., *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 a 3-D space into a family of time-dependent ordinary differential equations in a 1-D space following the moving flux tubes. However, this approach also introduces significant complications for data assimilation since the measurements are taken in 3-D space across different field lines (e.g., TEC from ground-to-satellite or satellite-to-satellite links), making the mapping between data and the model parameter space (equation (1)) very difficult to construct. This, in principle, can be overcome by using two frames: a Lagrangian frame used to solve the dynamical equa-

tions (10)–(13) and an Eulerian frame where a set of voxels, fixed in space and time, are used to solve for the Kalman filter equations (4)–(8). Ion and electron densities in the two frames can be related to each other by means of interpolation. We refer to this approach as the dual-frame approach.

[25] A more elegant and efficient approach is to solve both the Kalman filter equations and the dynamical equations in the same Eulerian frame. In this case the volume elements used to discretize the dynamical equations and to perform the Kalman filter are the same and they are defined by the intersection of constant magnetic field lines, constant magnetic potential lines, and constant magnetic longitudes. We refer to this approach as the single-frame approach.

[26] The USC/JPL GAIM uses the single-frame approach, where Earth's magnetic field is modeled by an eccentric tilted dipole (Figure 3). There are two main advantages to the single-frame over the dual-frame approach. (1) The dual-frame approach requires interpolation of the densities back and forth between the two frames at each time update in the Kalman filter. (2) The time update matrix $[\Psi_k]_{ij}$ (used in equations (7) and (8)) is by definition equal to the partial derivatives $\partial n_i(k+1)/\partial n_j(k)$, where $n_i(k+1)$ and $n_j(k)$ are the densities in voxel i at time $k+1$ and voxel j at time k , respectively. In the single-frame approach it is possible to analytically construct this matrix of partial derivatives and directly compute it. In the dual-frame approach, additional complications arise in trying to relate the partials of the densities in the Lagrangian frame to those in the Eulerian frame. In the face of these complications, one might have to construct $[\Psi_k]_{ij}$ by perturbing $n_j(k)$, solving the dynamical forward equations to propagate the state to time $k+1$, and then computing the change in $n_i(k+1)$. This has to be done for one voxel at a time, therefore requiring as many forward runs as the dimension of the state. Since the single-frame approach uses the same voxels to solve the dynamical and Kalman equations, it is possible to explicitly form the matrix Ψ_k without having to run the forward model. This represents substantial time saving and makes the implementation of the Kalman filter more feasible.

5. Results and Validation

[27] The USC/JPL GAIM model is now able to assimilate four major data types: absolute slant TEC measurements from ground GPS receivers, change in TEC data from GPS occultations, in situ electron density measurements, and UV airglow radiances. Here we present results from GAIM runs when assimilating only ground GPS-based TEC measurements using the band-limited Kalman filter for the period 22–24 May 2002. This period was chosen to assess the performance of GAIM on quiet days (22 and 24) and a disturbed day (23). For each day, nearly

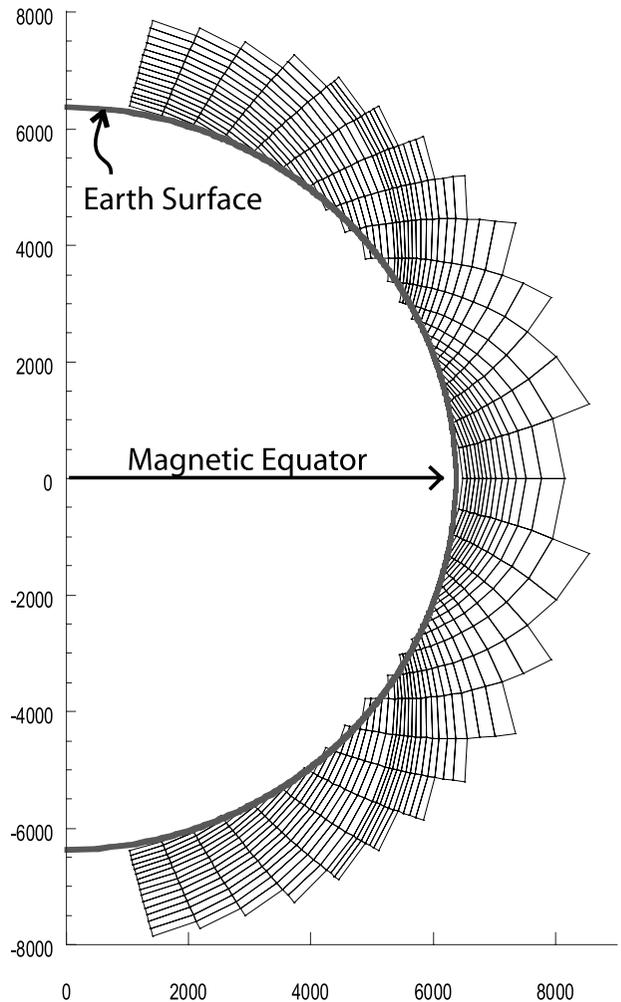


Figure 3. A cross section at one magnetic longitude of the volume elements used in the GAIM runs presented in section 5. These are defined by intersecting constant magnetic field lines, constant magnetic geopotential lines, and constant magnetic longitudes. The vertical axis is aligned with the magnetic dipole used to model Earth's magnetic field.

200,000 GPS TEC measurements, sampled at 1 measurement every 5 min, were available from 96–98 GPS receiver sites using an elevation cutoff of 10 degrees (Figure 4). These TEC measurements are based on dual-frequency phase measurements leveled to the pseudorange with the GPS instrumental biases determined using the JPL GIM technique [Mannucci *et al.*, 1998].

[28] A cross section at a constant geomagnetic longitude of the grid used in the GAIM run is shown in Figure 3. Table 1 summarizes the boundary of the region used in the GAIM run, the vertical and horizontal

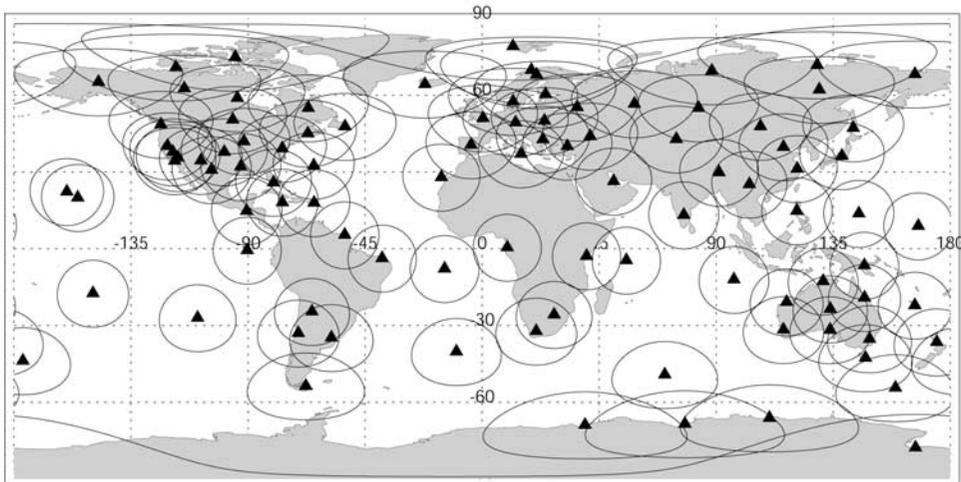


Figure 4. Coverage from 97 ground GPS stations used in GAIM for 23 May 2002. Because of data outage, the exact subset of sites and their number varies from day to day. This number was 96 on 22 May and 98 on 24 May 2002. The contour around each station corresponds to the visible region of the ionosphere at 450 km height for a 10 degree elevation mask.

resolutions of the grid, and the correlation lengths used in each of the radial, longitudinal, and latitudinal directions. The resolution specified is only approximate since the $p - q$ grid (Figure 3) does not map into a uniform grid in geomagnetic latitude, longitude, and height. The correlation volume is intentionally kept small to reduce the number of nonzero off-diagonal terms in the state error covariance matrix to speed up the assimilation run.

[29] Table 2 summarizes the geomagnetic conditions for each day. All 3 days assumed the same $E \times B$ drift climatology (that of June solar maximum conditions), MSIS for the neutral densities and temperatures, and HWM for neutral wind. In presenting our results below, we distinguish between two different GAIM runs: (1) GAIM climatology, which refers to the GAIM 3-D densities obtained by running the GAIM model without

assimilating any data; and (2) GAIM analysis, which refers to the GAIM 3-D densities obtained by assimilating the ground TEC data described above. In both cases, GAIM yielded a 3-D specification of electron density every 12 min for the entire 3 days considered. These can be integrated vertically to create global 2-D maps of VTEC. The GAIM analyses of electron density were validated in two ways: (1) comparison of the GAIM VTEC maps to the GIMs computed from the same ground GPS TEC data, and (2) comparison of GAIM VTEC values to independent TOPEX measurements.

5.1. Validation Against GIM

[30] GIM is a mapping technique which assumes a thin shell ionospheric model at 450 km. The details of the technique are described in the work of *Mannucci et al.* [1998]. In a nutshell, GIM solves for VTEC using a basis set of bicubic splines with local support on a spherical shell. The 2-D spherical grid is fixed in solar-magnetic coordinates (magnetic local time). By mapping line-of-sight TEC measurements to VTEC at the ionospheric shell piercing point, GIM solves for VTEC on the grid

Table 1. Specifications of the Grid and Correlation Used in the Data Assimilation Run

Parameter	Value
Modeled region longitude range	0–360°
Modeled region latitude range	–85–85°N
Modeled region altitude range	100–1500 km altitude
Latitude resolution	5°
Longitude resolution	15°
Altitude resolution	80 km
Total number of volume elements (voxels)	13,107
Correlation length in latitude	5°
Correlation length in longitude	15°
Correlation length in height	80 km

Table 2. Specifications of the Physics Input for the Three Days in May 2002

Date	$F = 10.7$	ap
22 May 2002	185.6	8
23 May 2002	184.8	78
24 May 2002	193.9	2

using a square root information filter (SRIF) [Bierman, 1977]. SRIF is equivalent to the Kalman filter but uses the square root of the inverse covariance in order to improve the condition number (reflected in the ratio of its largest to smallest eigenvalues) of matrices and therefore help numerical stability. GIM does not make use of any dynamical model, and therefore it is entirely data driven. In regions where there is no data (e.g., gaps in Figure 4), GIM relies on persistence in time to obtain a solution for VTEC. (More specifically, the a priori VTEC value at a given vertex is set equal to its value from the previous time step with a covariance that grows according to a first-order Gauss-Markov process.) Since the stations are rotating underneath the solar-magnetic reference frame in which the grid is defined, nearly all vertices will have some links going through them during the span of a few hours.

[31] Since GIM is a straightforward interpolation of the GPS TEC data using a 2-D shell model, it serves as a proxy for the information content of the GPS data set. GIM matches the TEC data (mapped to vertical) quite well near the GPS sites. However, GIM interpolation is less accurate at distances greater than 1000 km from the nearest site. Moreover, because of the thin shell model used by GIM, horizontal structures in the ionosphere can potentially create artifacts in the VTEC maps, therefore reducing their accuracy near strong gradient regions. Both of these limitations need to be remembered as we compare GAIM to GIM.

[32] To perform the comparison, a GIM global map is updated every 15 min for the 22–24 May 2002 period and interpolated to the 12-min GAIM runs for the same period. Owing to space limitation, we only show the comparison between GAIM and GIM at one time frame, about the middle of the period considered (23 May 1100 UTC), which exhibits features that are representative of all the other time frames.

[33] Figure 5 shows snapshots of VTEC from GAIM and GIM, along with maps of absolute and relative differences. Both GAIM climatology (Figure 5a, obtained by running the forward model and without assimilating any data) and GAIM analysis (Figure 5b, obtained by assimilating GPS TEC data) are shown. Figure 5a illustrates that GAIM climatology differs by more than 50% from reality (or at least the GIM proxy) in certain regions. By contrast, Figure 5b shows that GAIM analysis gives VTEC data that are very close to those of GIM, indicating that the TEC data are being used effectively by GAIM. Further examination of the GAIM analysis shows that GAIM reveals the equatorial anomaly more clearly and with higher resolution than GIM (compare left two panels of Figure 5b). This is an indication that the limitation induced by the GIM's thin shell model is reduced or eliminated by the 3-D GAIM grid. In addition, since GAIM climatology and GAIM

analysis appear to be quite different, even in regions where data is sparse (e.g., in equatorial regions over the Atlantic Ocean and Africa), we conclude that the dynamics introduced by the physical modeling of GAIM plays a significant role in the data assimilation.

[34] To appreciate the significance of the role of the physical model used in GAIM, consider the following. Currently, the GAIM Kalman filter only solves for the O^+ ion and electron densities and does not adjust any of the drivers. Therefore it is expected that if data stop flowing into GAIM, the GAIM analysis will revert to the GAIM climatology on timescales ranging from minutes to several hours. The wide range of timescales is due to the fact that the production, loss, convection, and diffusion of ions respond to the various driving forces on different timescales. For example, the fast recombination rate of molecular ions causes the F1 region to disappear quickly at night, while the slower recombination rate of atomic ions causes the F2 region to last long after dusk when the radiation from the Sun stops. Even when the driving forces are only approximately correct, the dynamical model plays an important role in assigning the correct timescale at different regions and local times in the ionosphere. Effectively, when the timescale is long, the initial condition of the ion densities will have a stronger effect on the evolution of the ionosphere, therefore extending the influence of data over longer periods and larger regions. When the timescale is short, the initial condition of the ion densities will have little effect on the evolution of the ionosphere, therefore limiting the influence of data to shorter periods and smaller regions. Thus the model plays a crucial role in assigning the proper time correlation length at different local times, heights, and latitudes. This information is completely lost if one uses a constant timescale everywhere, as would be the case if no dynamical model is used to map the state or the state error covariance (equations (7) and (8)).

5.2. Time Evolution During a Magnetic Storm

[35] The May period chosen for our analysis was centered around a magnetic storm. Figure 6 shows the hourly Dst for 3 days starting at 0000 UT, 22 May 2002, and indicates the onset of a magnetic storm at 1200 UT, 23 May. The 3-hour ap and Kp indices for 23 May are given in Table 3. Figure 7 shows hourly GAIM analysis of VTEC for the 23 May at UT = 1200, 1300, . . . , 1700, a period which corresponds to the main phase of the storm. For comparison, we also show the corresponding VTEC maps at the same local time for 22 May as a proxy for the expected ionospheric features during a quiet period. Comparing the 2 days, a clear enhancement of VTEC at the southern geomagnetic equatorial region is seen at 1400 UT during the storm. Furthermore, at 1500 UT, an enhancement of the equatorial anomaly extending from 0800 to 2000 LT, can be seen during the

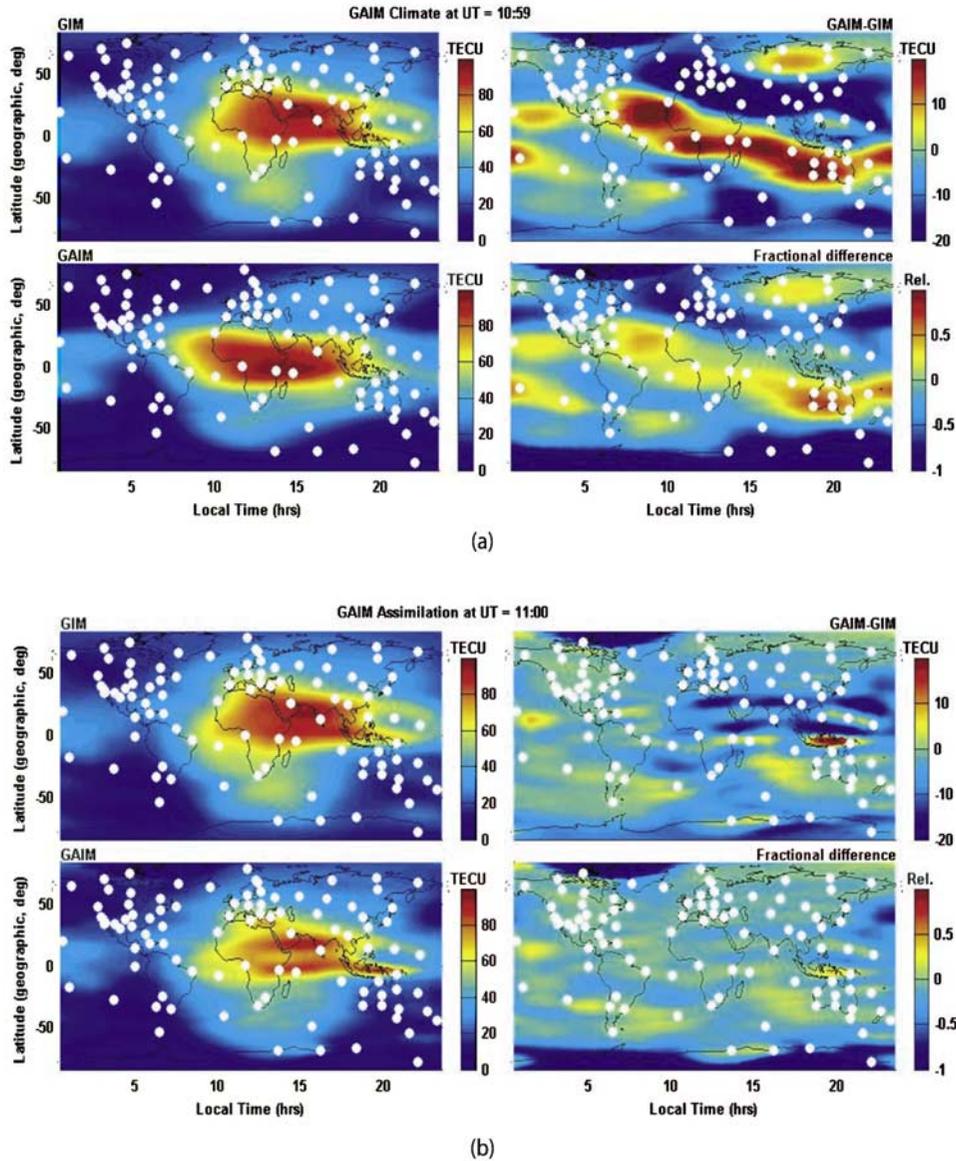


Figure 5. (a) Comparison of GIM and GAIM climatology at about 1100 UT on 23 May 2002. Top left: Global ionospheric map (GIM) of vertical TEC. Bottom left: global vertical TEC obtained from vertically integrating GAIM climatology runs. Top right: the difference between GAIM climatology and GIM. Bottom right: fractional difference of GAIM climatology and GIM (defined as $2 [GAIM - GIM] / [GAIM + GIM]$). The climatology is obtained by running GAIM without assimilating any data and with input indices given in Table 2. (b) Same as Figure 5a but showing the GAIM analysis obtained by assimilating ground TEC data from 23 May 2002. The dots indicate the GPS ground receivers. The GAIM analysis shows the equatorial anomaly more distinctly than GIM.

storm day relative to the quiet day. This is presumably caused by a storm-induced enhancement of the eastward electric field. Figure 8 shows cross sections of GAIM analysis of electron densities at the same universal times

as in Figure 7 but only for 23 May. The cross sections are taken at 7.5 geographic longitude, and therefore the local time is given by the UT + 30 min. The features seen in the VTEC maps, such as the enhancement of

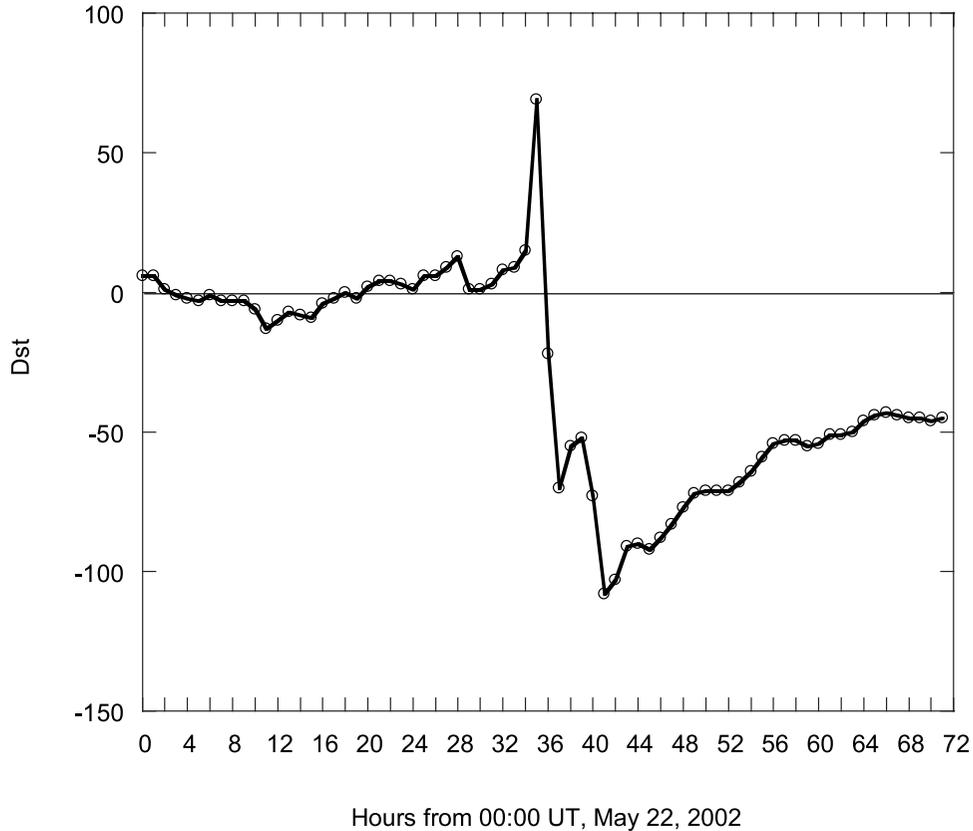


Figure 6. *Dst* values for the period 22–24 May 2002 showing the main phase of a storm between 1300 and 1700 UT on 23 May.

VTEC at 1400 UT, are clearly seen in the densities as well.

5.3. Validation Against TOPEX

[36] VTEC below the TOPEX track derived from the dual-frequency altimeter has been used extensively as an independent data source for validation [Ho *et al.*, 1997; Codrescu *et al.*, 2001]. Given the precision, latitudinal coverage, and long time series of the TOPEX data, it offers a unique and powerful means of validation. When compared to VTEC derived from GPS ground data, TOPEX VTEC are especially challenging given that they are measurements taken exclusively over the ocean where few GPS stations exist. Also, validation against TOPEX will mainly tell us how well GAIM can estimate VTEC but says nothing about how well it can estimate densities. Validation of densities will require a different data source and will be the subject of a future investigation.

[37] We start by comparing VTEC from TOPEX, GAIM climatology, GAIM analysis, and GIM. Figure 9 shows VTEC for 8 out of the 27 TOPEX tracks on 23 May 2002. Also shown in Figure 9 are the tracks of the TOPEX footprint and neighboring GPS stations used

in the assimilation. We note the following features in the comparisons:

[38] 1. GAIM climatology matches TOPEX very well in some tracks (e.g., track 10) while it differs significantly in others (e.g., 13, 14, 15, 16, and 19).

[39] 2. The GAIM analysis is significantly different from the GAIM climatology and compares much better with TOPEX (visible in all tracks), indicating that GPS TEC data are being assimilated effectively.

[40] 3. The agreement between the GAIM analysis and TOPEX is quite good in many cases (e.g., 8, 14, 15, 16, and 19).

Table 3. Three-Hourly *ap* Index on 23 May 2002

Time, UT	<i>ap</i>	<i>Kp</i>
0130	12	3–
0430	12	3–
0730	7	2
1030	111	7–
1330	179	8–
1630	236	8+
1930	48	5
2230	18	3+

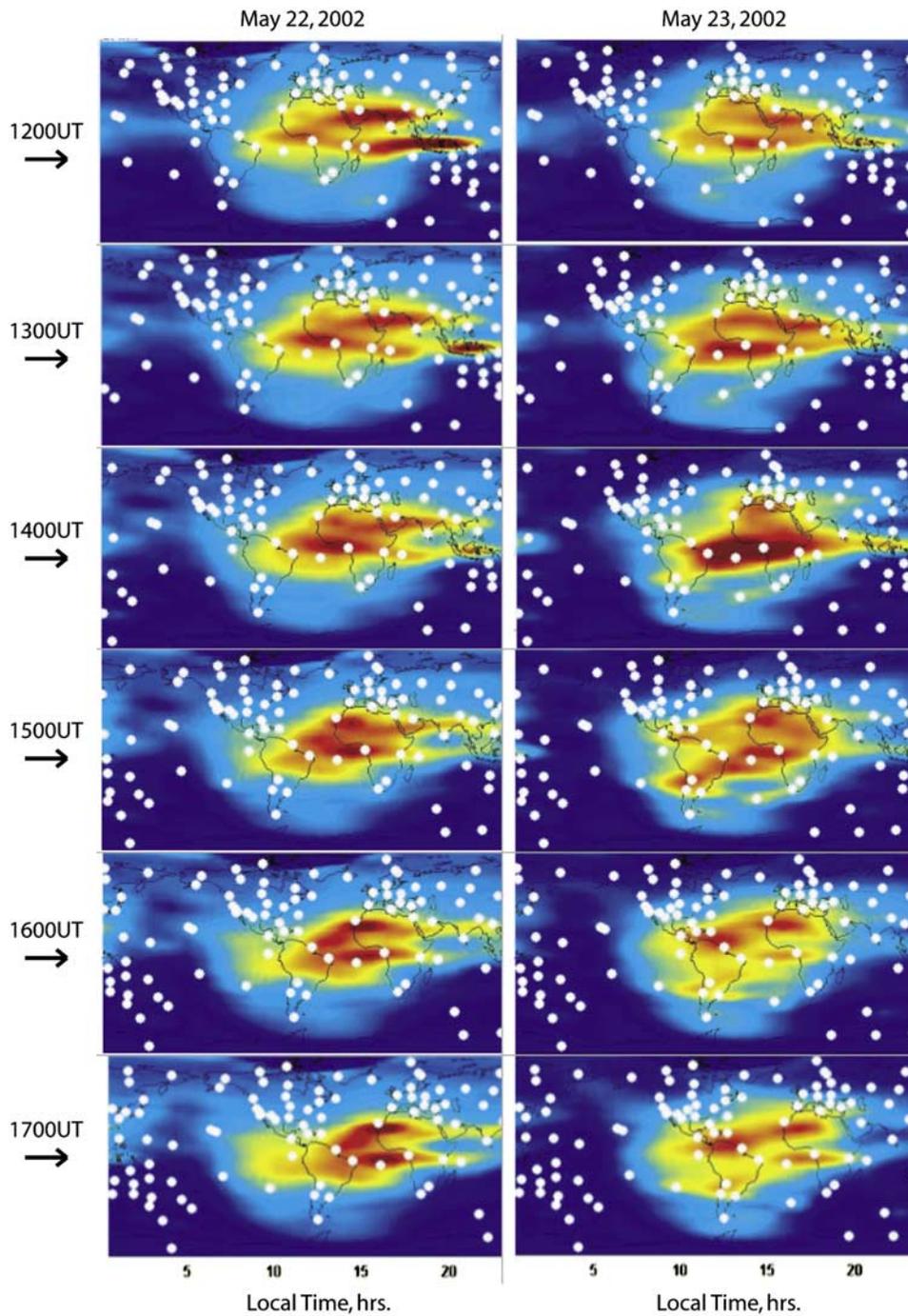


Figure 7. Hourly global snapshots of vertical TEC obtained by vertically integrating hourly 3-D GAIM analyses for the days indicated on the first row and the UT indicated on the first column. A disturbed day (23 May 2002) is shown next to a quiet day (22 May 2002) for comparison.

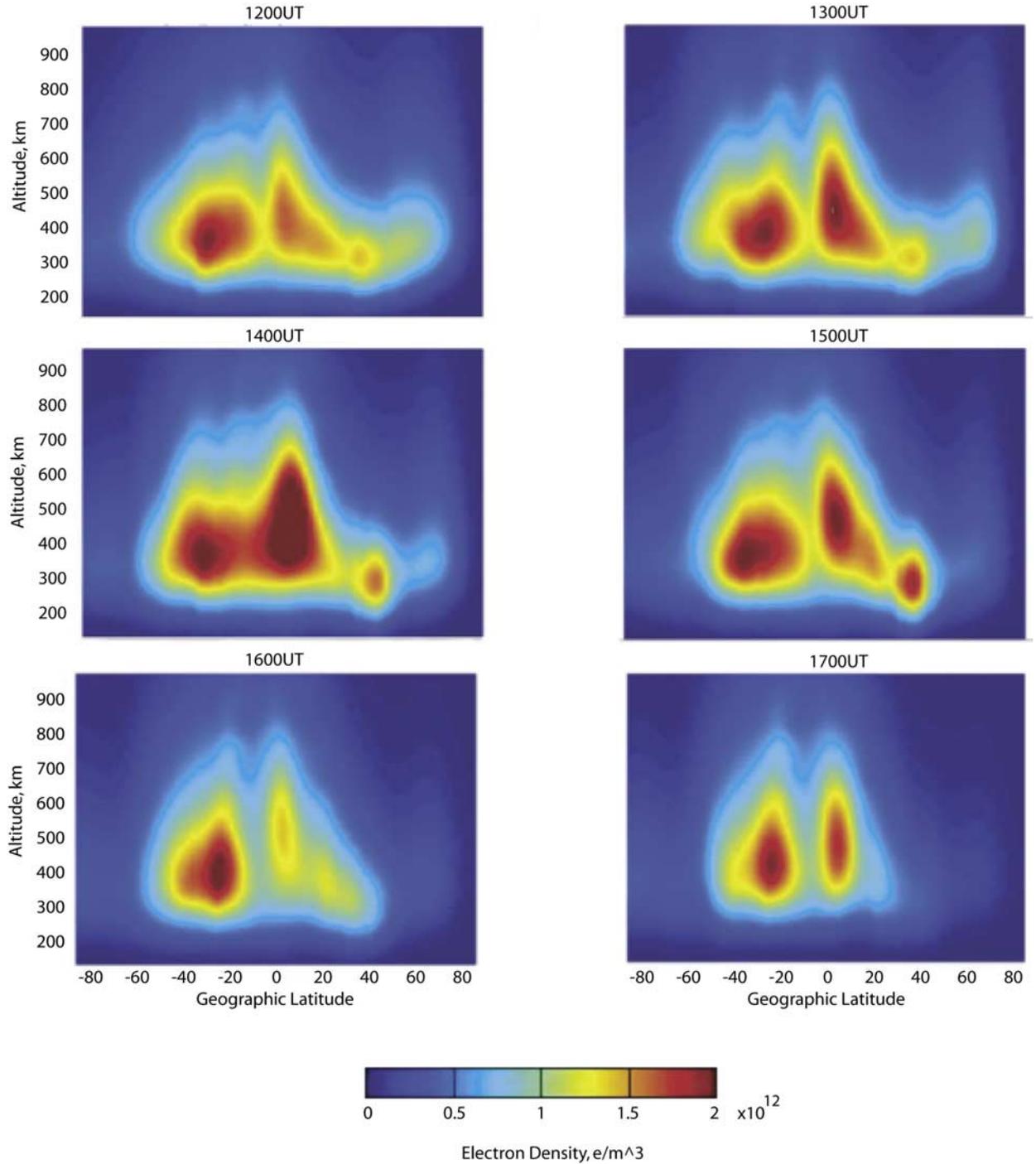


Figure 8. Hourly electron density snapshots (in units of 10^{12} electrons/ m^3) at 7.5 geographic degrees longitudinal planes obtained from the same hourly GAIM analyses shown in Figure 7 for 23 May 2002. The UT is indicated above each panel; local time is given by UT + 30 min.

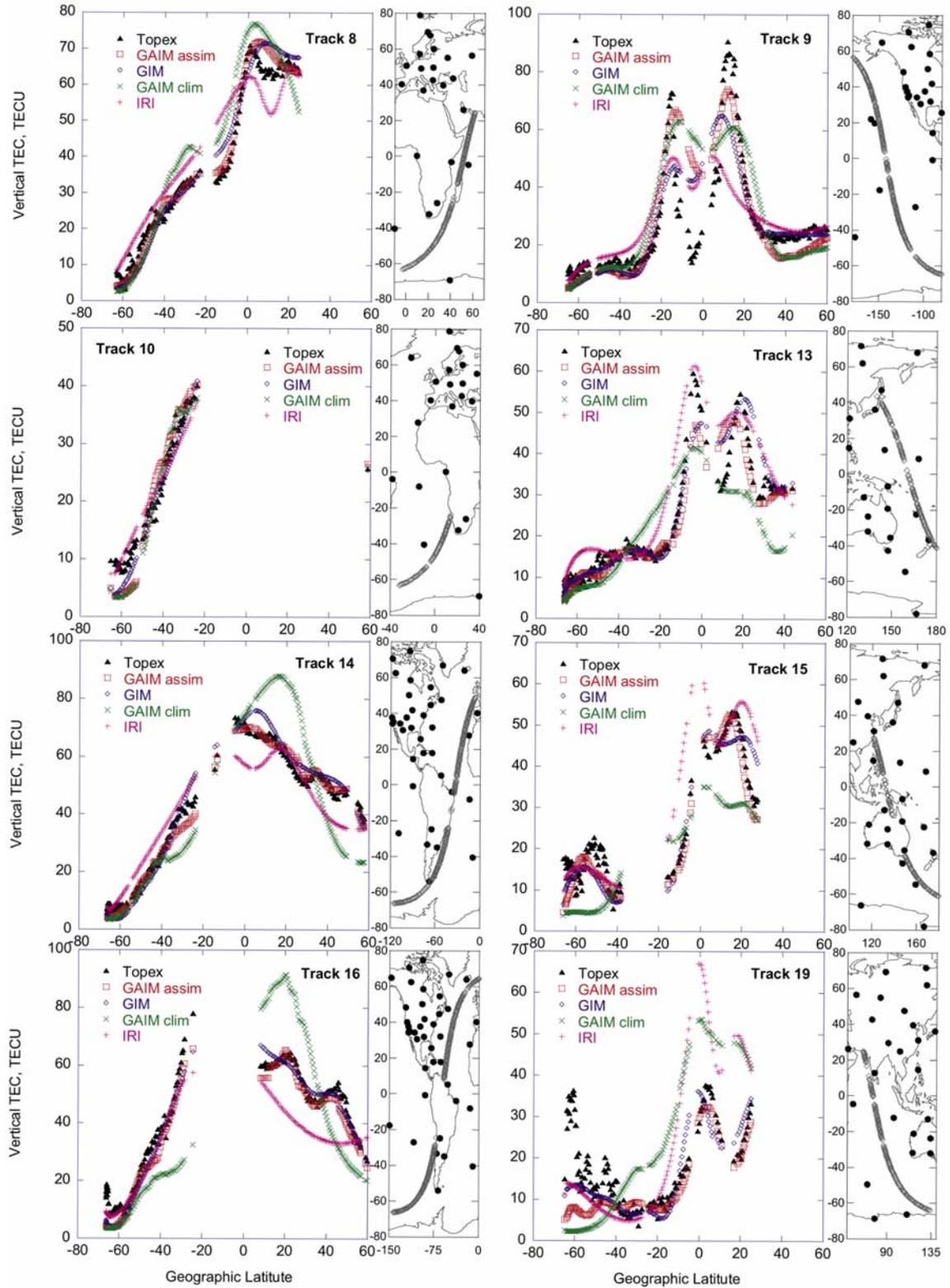


Figure 9.

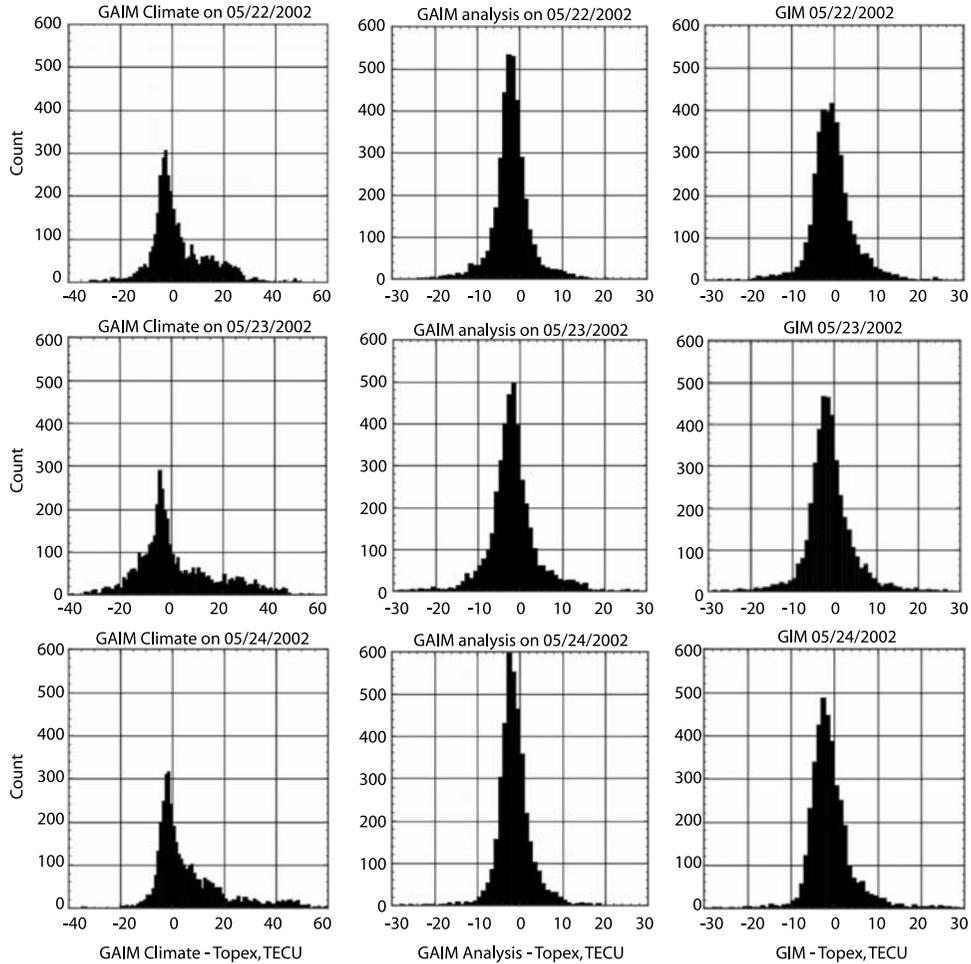


Figure 10. Histograms of vertical TEC differences between GAIM climatology (left panels), GAIM analysis (i.e., assimilation) (middle panels), or GIM (right panels) and those obtained from TOPEX for the three days 22–24 May 2002.

[41] 4. Whenever an equatorial anomaly appears in TOPEX, it also appears in the GAIM analysis, (e.g., 9, 13, and 19), in some cases with great fidelity (e.g., 19).

[42] 5. GAIM appears to be able to capture steep VTEC gradients associated with the equatorial anomaly better than GIM (e.g., 9, 15, 16, and 19). This is presumably due to the thin shell limitation of GIM.

5.4. Statistical Comparison to TOPEX

[43] To further assess the performance of GAIM, we examine histograms of VTEC difference between GAIM

climatology and TOPEX (left panels of Figure 10), GAIM analysis and TOPEX (middle panels), and GIM and TOPEX (right panels) for all TOPEX tracks during 22–24 May 2002. Statistical summaries of the histograms are given in Table 4. We emphasize that the VTEC differences between GAIM analysis and TOPEX have a standard deviation of $\sigma_{\text{GAIM/A}} = 5.2$ TECU (1 TECU = 10^{16} e/m²) over the 3 days, which is almost 3 times better than the standard deviation for GAIM climatology ($\sigma_{\text{GAIM/C}} = 13.8$ TECU), twice better than IRI ($\sigma_{\text{IRI}} = 9.6$ TECU), and slightly superior to GIM

Figure 9. Comparison of vertical TEC below the TOPEX track for different tracks on 23 May 2002. To the right of each figure is the TOPEX ground marked by UT and neighboring ground GPS receivers. The left panels correspond to ascending TOPEX tracks with an ascending node at ~ 1000 LT. The right panels correspond to descending tracks at ~ 2200 LT.

Table 4. Statistics on Vertical TEC Differences Between GAIM Climatology (GAIM/C), GAIM Analysis (GAIM/A), GIM, or IRI and Those of TOPEX for the Three Days of 22–24 May 2002, Sorted by Increasing Standard Deviation^a

Day in May 2002	Quantity	N	Average	SD	Minimum	Maximum	RMS
24	GAIM/A–TOPEX	3901	−1.3	4.4	−28.6	36	4.6
22	GAIM/A–TOPEX	3763	−2	4.9	−35	41.7	5.3
22	GIM–TOPEX	3763	−0.6	5.2	−30.7	28	5.2
23	GIM–TOPEX	4102	−1.2	5.7	−35	29.3	5.8
24	GIM–TOPEX	3901	−1	5.9	−36.7	36	6
23	GAIM/A–TOPEX	4102	−1.7	6.1	−30.4	35.1	6.3
22	IRI–TOPEX	3763	0.8	8.3	−38.2	32.5	8.3
24	IRI–TOPEX	3901	3.3	8.8	−39.5	38.1	9.4
22	GAIM/C–TOPEX	3763	2.3	10.8	−32	49.2	11
23	IRI–TOPEX	4102	2.2	11.1	−48.2	56.4	11.3
24	GAIM/C–TOPEX	3901	6.4	14.2	−35	63.9	15.6
23	GAIM/C–TOPEX	4102	1.6	15.3	−45.5	55.8	15.4

^aThe statistics include the total number of points for each day (N), the average, the standard deviation (SD), the minimum and maximum differences and the RMS.

($\sigma_{\text{GIM}} = 5.6$ TECU). Both GAIM analysis and GIM VTEC are biased low by 1–2 TECU relative to TOPEX when it should be high given that TOPEX is at 1330 km and GPS is at 20,000 km altitude. One TECU translates into altimetric range delay of ~ 2 mm, which is well within the error budget of TOPEX; therefore the bias could very well be due to TOPEX.

6. Conclusions

[44] 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, formal data assimilation techniques are required to fully exploit the information in a physics-based model. Comparison of GAIM analysis and GIM shows very close agreement, which is a minimum requirement for a successful assimilation model given that the two approaches rely on the same data source (ground TEC). However, the comparison of GAIM and GIM to TOPEX VTEC indicates that GAIM is comparable or superior to GIM in that it is able to capture sharp changes in VTEC better than GIM, the latter being limited by the thin shell model and the lack of physics. Future GAIM studies will include the assimilation of other data types beside TEC, more extensive statistical comparisons and validation, and the simultaneous estimation of ion densities and some of the ionospheric drivers by means of exchanging information between the 4DVAR and the Kalman filter.

[45] **Acknowledgments.** We thank Graham Bailey for his assistance in the effort of building GAIM. This work is supported by the Department of Defense through a Multidisciplinary

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 (1993), A modeling study of the equatorial topside ionosphere, *Ann. Geophys.*, *11*, 263–272.
- Bierman, G. J. (1977), *Factorization Methods for Discrete Sequential Estimation*, Academic, San Diego, Calif.
- Codrescu, M. V., K. L. Beierle, T. J. Fuller-Rowell, S. E. Palo, and X. L. Zhang (2001), More total electron content climatology from TOPEX/Poseidon measurements, *Radio Sci.*, *36*(2), 325–333.
- Daley, R. (1991), *Atmospheric Data Analysis*, Cambridge Univ. Press, New York.
- Dymond, K. F., S. A. Budzien, A. C. Nicholas, S. E. Thonnard, R. P. McCoy, and R. J. Thomas (2001), Electron densities determined by inversion of ultraviolet limb profiles, *J. Geophys. Res.*, *106*(A12), 30,315–30,321.
- Fejer, B. G., E. R. de Puala, S. A. Gonzales, and R. F. Woodman (1991), Average vertical and zonal *F* region plasma drifts over Jicamarca, *J. Geophys. Res.*, *96*, 13,901–13,906.
- Fremouw, E., J. Secan, and B. Howe (1992), Application of stochastic inverse theory to ionospheric tomography, *Radio Sci.*, *27*(5), 721–732.
- Ghil, M., and P. Malanotte-Rizzoli (1991), Data assimilation in meteorology and oceanography, *Adv. Geophys.*, *33*, 141–266.
- Ghil, M., K. Ide, A. F. Bennett, P. Courtier, M. Kimoto, and N. Sato (Eds.) (1997), *Data Assimilation in Meteorology and Oceanography: Theory and Practice*, 496 pp., Meteorol. Soc. of Jpn., Tokyo.
- Hajj, G. A., and L. J. Romans (1998), Ionospheric electron density profiles obtained with the global positioning system: Results from the GPS/MET experiment, *Radio Sci.*, *33*(1), 175–190.

- Hajj, G. A., R. Ibañez-Meier, E. R. Kursinski, and L. J. Romans (1994), Imaging the ionosphere with the Global Positioning System, *Int. J. Imaging Syst. Technol.*, *5*, 174–184.
- Hajj, G. A., L. C. Lee, X. Pi, L. J. Romans, W. S. Schreiner, P. R. Straus, and C. Wang (2000), COSMIC GPS ionospheric sensing and space weather, *Terr. Atmos. Oceanic Sci.*, *11*(1), 235–272.
- Hedin, A. E. (1991), Extension of the MSIS thermospheric model into the middle and lower atmosphere, *J. Geophys. Res.*, *96*, 1159–1172.
- Hedin, A. E., et al. (1996), Empirical wind model for the upper, middle and lower atmosphere, *J. Atmos. Terr. Phys.*, *58*, 1421–1447.
- Heppner, J. R., and N. C. Maynard (1987), Empirical high-latitude electric-field models, *J. Geophys. Res.*, *92*, 4467–4489.
- Ho, C. M., B. D. Wilson, A. J. Mannucci, U. J. Lindqwister, and D. N. Yuan (1997), A comparative study of ionospheric total electron content measurements using global ionospheric maps of GPS, TOPEX radar, and the Bent model, *Radio Sci.*, *32*(4), 1499–1512.
- Howe, B. M., K. Runciman, and J. A. Secan (1998), Tomography of the ionosphere: Four-dimensional simulations, *Radio Sci.*, *33*(1), 109–128.
- Kalman, R. (1960), A new approach to linear filtering and prediction problems, *Trans. ASME J. Basic Eng.*, *82*, 35–45.
- Kalman, R., and R. Bucy (1961), New results in linear filtering and prediction theory, *Trans. ASME J. Basic Eng.*, *83*, 95–108.
- Leitinger, R., H. P. Ladreiter, and G. Kirchengast (1997), Ionosphere tomography with data from satellite reception of Global Navigation Satellite System signals and ground reception of Navy Navigation Satellite System signals, *Radio Sci.*, *32*(4), 1657–1669.
- Lorenc, A. (1981), A global three-dimensional multivariate statistical interpolation scheme, *Mon. Weather Rev.*, *109*, 701–721.
- Mannucci, A., B. Wilson, D. Yuan, C. Ho, U. Lindqwister, and T. Runge (1998), A global mapping technique for GPS-derived ionospheric total electron content measurements, *Radio Sci.*, *33*(3), 565–582.
- Pi, X., A. J. Mannucci, U. J. Lindqwister, and C. M. Ho (1997), Monitoring of global ionospheric irregularities using the worldwide GPS network, *Geophys. Res. Lett.*, *24*(18), 2283–2286.
- Pi, X., C. Wang, G. A. Hajj, G. Rosen, B. D. Wilson, and G. J. Bailey (2003), Estimation of $E \times B$ drift using a global assimilative ionospheric model: An observation system simulation experiment, *J. Geophys. Res.*, *108*(A2), 1075, doi:10.1029/2001JA009235.
- Rosen, I. G., C. Wang, G. Hajj, X. Pi, and B. Wilson (2001), An adjoint method based approach to data assimilation for a distributed parameter model for the ionosphere, paper presented at *40th IEEE Conference on Decision and Control*, Inst. of Electr. and Electron. Eng., Orlando, Fla., 4–7 Dec.
- Schreiner, W. S., S. V. Sokolovskiy, C. Rocken, and D. C. Hunt (1999), Analysis and validation of GPS/MET radio occultation data, *Radio Sci.*, *34*, 949–966.
- Tobiska, W. K. (1991), Revised solar extreme ultraviolet flux model, *J. Atmos. Terr. Phys.*, *53*, 1005–1018.
- Wang, C., G. A. Hajj, X. Pi, G. Rosen, and B. D. Wilson (2004), Development of a Global Assimilative Ionospheric Model, *Radio Sci.*, *39*, RS1S06, doi:10.1029/2002RS002854, in press.

X. Pi and B. D. Wilson, Jet Propulsion Laboratory, 4800 Oak Grove Drive, M/S: 138-308, Pasadena, CA 91109-8099, USA. (xiaoqing.pi@jpl.nasa.gov; brian.d.wilson@jpl.nasa.gov)

G. A. Hajj, I. G. Rosen, and C. Wang, University of Southern California, Department of Mathematics, Los Angeles, CA 90089-1113, USA. (george.hajj@jpl.nasa.gov; grosen@math.usc.edu; cwang@cams.usc.edu)