A Status Report on Applying Discrete Inverse Theory to Ionospheric Tomography

E. J. Fremouw, J. A. Secan and R. M. Bussey
Northwest Research Associates, Inc., 300 120th Avenue NE, Bellevue, Washington 98005

B. M. Howe
Applied Physics Lab, University of Washington, Seattle, Washington 98105

ABSTRACT
Discrete inverse theory (DIT) provides an orderly framework in which to combine measurements of total electron content (TEC) with a priori information to image the ionosphere tomographically. We have developed a DIT-based tomographic processor for use with relative-TEC data. The processor's a priori information comprises the global mean of over 17,000 profiles generated from an ionospheric model, for use as a “generic background”; empirical orthogonal functions (EOFs) spanning the same model profiles, for use as vertical basis functions; and a red power-law horizontal spectrum. Relative TEC data are used to evaluate coefficients multiplying the EOFs and harmonics, thus quantifying a perturbation electron-density field. The perturbation field, which need not be small, is added to the a priori background to produce the image. We present here several images produced by employing the processor with simulated TEC data based on in situ ionospheric measurements and incoherent-scatter radar observations. ©1994 John Wiley & Sons, Inc.

I. INTRODUCTION
Among the techniques for tomographically inverting path-integral measurements of plasma density (i.e., TEC data) to image the ionosphere, those employing DIT have yet to be widely employed with field data. In a previous paper [1], we described such a linear algebraic approach and presented two sample inversions employing a weighted, damped least-squares procedure and simulated absolute TEC data.

We have since modified the procedure so that it can now directly employ relative TEC data, without need for prior ambiguity resolution, and we have committed the modified theory to a multiprogram inversion processor. In this article, we briefly review the theoretical basis of the approach, including the modification, and present simulated results from the processor. We address the case of multiple receivers observing a single source in sequential locations.

II. UNDERLYING CONSIDERATIONS
A. Theory. The absolute-TEC forward problem may be discretized by representing the dispersive phase, $\phi_h$, observed on a path, $h$, plus its $2n_h \pi$ ambiguity by any unknown dispersion offset, $\Delta \phi_h$, arising in the receiver and/or transmitter as proportional (via a constant, $C$) to the sum of TEC increments accumulated in cells, $i$, as follows:

$$\phi_h + 2n_h \pi - \Delta \phi_h = C \sum_i N_{ei} ds_{hi},$$

where $N_{ei}$ is the local mean electron density in the cell and $ds_{hi}$ is the elemental length of the path in the cell.

The inverse problem is to estimate the $N_{ei}$ from a multi-path set of $\phi_h$. To begin, we postulate a representative background profile (height-dependent mean values of density in a gaussian ensemble of ionospheric profiles) and suppose it to be subtracted from the $N_{ei}$ to obtain a perturbation field, $\delta N_{ei}$. We also compute the dispersive-phase contribution of the background and subtract it from the measured $\phi_h$ to obtain the following set of perturbation data, $\delta$:

$$\delta = C \sum_i \delta N_{ei} ds_{hi} - 2n_h \pi + \Delta \phi_h.$$  

We now note that the ambiguity is the same for all paths terminating at the same receiver so long as the receiver obtains a continuous record, and we assume that both receiver and transmitter offsets remain constant (although they may be of any magnitude) over the duration of the record. This being the case, subtracting the same datum, $\delta_0$, (e.g., the first one) from all values obtained by a given receiver (along with their identical ambiguities and offsets) results in

$$\delta - \delta_0 = C \sum_i \delta N_{ei}(ds_{hi} - ds_{ui}),$$

where $ds_{ui}$ is the pathlength of the reference ray through the ith cell (zero for most cells). That is, after dividing through by $C$ to obtain $\delta_h = (\delta_h - \delta_0)/C$, we have the following relative-TEC forward problem:

$$\delta_h = \sum_i \delta N_{ei}(ds_{hi} - ds_{ui}).$$

Our objective is to estimate $\delta N_{ei}$, to which we shall add the background profile to obtain estimates of $N_{ei}$.

The one-dimensional array of $d_h$ values (the multireceiver TEC data set with the reference value for each receiver subtracted from its subset) constitutes a "data vector," $d$. For
reasons described in Fremouw et al. [1], we choose to repre-
sent the two-dimensional (altitude and geomagnetic latitude,
say) field of $\delta N_e$ by means of two-dimensional basis functions and corresponding coefficients. Specifically, we expand it horizontally into a Fourier series of sines and cosines and vertically into a short series of empirical orthonormal functions (EOFs). This transforms the problem into estimating a "model vector," $\mathbf{m}$, the elements, $m_k$, of which are coefficients to be applied to multiplied pairs of horizontal and vertical basis functions. (For specifics, see Section 2.2 of Fremouw et al. [1]).

In matrix notation, the transformed forward problem is:

$$\mathbf{d} = \mathbf{Gm} + \mathbf{n},$$

where $\mathbf{G}$ is the geometry matrix describing the dependence of the data on the model parameters, with elements

$$G_{hk} = \partial d_k / \partial m_h,$$

and $\mathbf{n}$ is a "noise vector" whose elements, $n_k$, quantify the rms uncertainty due to dispersive-phase noise and model deficiency. The modeling procedure suffers from deficiencies such as the use of finite Fourier and EOF series, and these can be estimated. The uncertainty estimates can be different on different paths because of such variables as range-length loss and the path-lengths through structures not accounted for in the finite basis-function series.

With the forward problem formulated as in Eq. (5), the inverse problem becomes that of solving for an estimated model vector, $\hat{\mathbf{m}}$, employing a generalized inverse [2] matrix, $G^{-\#}$, as follows:

$$\hat{\mathbf{m}} = G^{-\#} \mathbf{d}.$$  \hspace{1cm} (7)

The central aspect of the procedure is to find $G^{-\#}$. Once that is accomplished and data are in hand, the matrix multiplication described by Eq. (7) is straightforward, as is transforming the resulting "model" (i.e., the estimated coefficients and their corresponding basis functions) into configuration space and adding the background profile removed in going from Eq. (1) to Eq. (2).

As described in Section 2.3 of Fremouw et al. [1], we employ the following generalized inverse for the weighted, damped, least-squares problem:

$$G^{-\#} = \nu G^T(G\nu G^T + \mathbf{e})^{-1},$$

where $^T$ and $^{-1}$ stand for transpose and standard inverse, respectively, and where $\mathbf{v}$ and $\mathbf{e}$ are the a priori covariance matrices of $\mathbf{m}$ and $\mathbf{n}$, respectively. In practice, we take the covariance matrices to be diagonal, so they become variance vectors (and, hence, we represent them by lower-case letters in our present notation).

Although Eq. (5) and following are formally identical to their counterparts in Fremouw et al. [1], there is an important difference between $\mathbf{G}$ here and that in the previous work. If we denote the geometry matrix for the absolute-TEC problem dealt with in Fremouw et al. by $\mathbf{G}_{ab}$, then its counterpart in the relative-TEC problem introduced here may be written as

$$\mathbf{G} = \mathbf{G}_{ab} - \mathbf{G}_o,$$

where $\mathbf{G}_o$ describes the dependence of the reference-ray measurements (one for each receiver) on the model parameters. This difference accounts for the subtraction on the right-hand side of Eqs. (3) and (4). In implementation, it merely means subtracting the elements of $\mathbf{G}_{ab}$ for each receiver's reference ray from all elements for that receiver (with the reference-ray elements of $\mathbf{G}$ becoming zero).

B. Application. In applying the foregoing theory, two important considerations arise: the choice of basis functions and the weights to assign to them in defining the a priori variance vectors, $\mathbf{v}$ and $\mathbf{e}$. The relative magnitudes of these two vectors determine the degree to which the processor treats the inversion as an overdetermined least-squares problem versus that to which it produces the minimum-length solution, which would be a latitudinally uniform set of background profiles. The solution actually achieved is a weighted combination of the two. Substituting Eq. (7) into Eq. (5) produces

$$\hat{\mathbf{m}} = G^{-\#} \mathbf{Gm} + G^{-\#} \mathbf{n},$$

the second term of which indicates that the amount by which the estimated model vector, $\hat{\mathbf{m}}$, differs from the true one, $\mathbf{m}$, includes an inverted combination of rms phase noise and model deficiency. The weight that the processor gives to each datum in solving the least-squares problem is dictated by the a priori values assigned to the squares of these combinations, which are the elements of $\mathbf{e}$. Similarly, the weight given to estimating each model parameter is dictated by the a priori values assigned to the elements of $\mathbf{v}$. (Because the a priori expectation values of the $\mathbf{m}$ elements are zero, we also have the a priori condition that $v_k = \langle m_k^2 \rangle$. We have incorporated into our processor a capability for accepting interactively prompted user estimates of the relative magnitudes of $\mathbf{e}$ and $\mathbf{v}$ (the "damping parameter" described in Menke [2]). In most instances, the default weights described below should be appropriate.

Horizontal structures in ionospheric plasma density are dominated by large scales and are characterized appropriately by a power-law spatial spectrum having an outside-cutoff, so we choose a Fourier representation for them. The set of harmonics employed is dictated by the image span and horizontal resolution available and by computational considerations regarding practicable matrix size. Our processor employs 60 harmonics. The estimated contribution of model deficiency to the a priori estimates of $\mathbf{e}$ account mainly for ionospheric variance outside this finite spectrum.

For assigning a priori weights to the elements of $\mathbf{v}$, we assume a $k^{-3}$ (one-dimensional) spatial spectrum with an outside-scale wavelength of $5^\circ$ of latitude, along with eigenvalues emerging from a vertical EOF analysis of an ionospheric model. The model employed is that described by Secan [3]. In the present work, one objective was to identify a background profile and associated EOFs that would be broadly applicable. Accordingly, we performed the EOF analysis for the full range of latitudes (north and south) and longitudes, four GMTs, both solstitial and equinoctial conditions, and three sunspot numbers. The background profile incorpo-
rated into our processor is the mean of 17,472 profiles spanning these conditions, our intent being that it represent an “all-time mean, generic ionosphere” to be perturbed by means of TEC data.

Initially, we expected to employ the EOFs spanning the foregoing broad conditions as vertical basis functions and their eigenvalues directly in evaluating the a priori elements of \( \nu \), but tests conducted with such functions and a priori coefficients were not fully satisfactory. Becoming concerned about undue sensitivity to the basis functions used, we employed EOFs generated from more limited ranges of conditions (especially geomagnetic latitude) to perform direct best fits to several test-case fields of electron density (without attempting tomographic inversion of integrals through the fields). Although there were differences in the best-fit renderings, we found them all to be reasonably satisfactory. The problem lay not in the choice of basis functions, but rather in the limited vertical resolution (resulting from finite grazing angle) achievable with surface-to-orbit geometry.

Specifically, TEC variations produced by horizontal structures were reflected in low-order vertical modes, but the limitation on vertical resolution prevented higher-order modes from restoring the relatively unchanged vertical profiles. Our solution was to impose a “zero-order” basis function proportional to the background profile. We then assigned a portion of the low-order variance contained in the eigenvalues obtained from the EOF analysis to this additional basis function when establishing the a priori variance vector, \( \nu \). This departure from the mathematical purity of strictly orthogonal basis functions permits accommodation of horizontal structures without distorting the vertical profile.

We have also introduced an additional ad hoc vertical basis function. Our broadly based EOF analysis, the functions being empirical by definition, fails to isolate the E layer from the F layer. Because the layers do undergo somewhat distinct dynamic morphologies, and because the foregoing zero-order function is heavily weighted in the F layer, it seemed prudent to introduce a separate function weighted mainly in the E layer. Such a function was, in fact, isolated in EOFs spanning a limited (high-latitude, winter) range of conditions. We added it to our “generic” set, and again redistributed the a priori variances, keeping the total variance (magnitude of \( \nu \)) equal to that obtained from the generic EOF analysis.

The vertical basis functions in our processor are depicted in Figure 1. The solid curves represent EOFs 1–4 from the broadly based analysis. The dashed curve is the “zero-order” function derived from (proportional to) the generic background profile, and the dotted one is that added to represent the E layer. In our processor, the latter is employed as the last of six vertical basis functions identified as of order 0–5. They are taken to account for 78.3, 11.0, 7.3, 2.3, 0.6, and 0.5% respectively, of the a priori variance about our generic background profile. For each vertical function, its share of the a priori variance is taken to reside half in case-to-case (“temporal”) variations and the other half in the \( k \) spatial spectrum described above. The spatially dc term, as it is applied independently to each vertical function, permits latitudinally uniform variations in the profile, just as the zero-order vertical function accommodates vertically uniform horizontal structures.

![Vertical Basis Functions](image)

**Figure 1.** Vertical basis functions employed. Solid curves depict four EOFs derived from 17,472 disparate ionospheric profiles. Dashed curve denotes a “zero-order” basis function that is proportional to the mean of those profiles. Dotted curve demarks a basis function heavily weighted in the E layer.

### III. SIMULATED RESULTS

We have tested our processor primarily by means of actual data from the in situ plasma-density sensors flown in the Defense Meteorological Satellite Program (DMSP). To obtain such test data, we first extended horizontal (essentially latitudinal) records taken near 800 km altitude vertically by means of representative model profiles [4]. These “true” profiles were derived by means that were totally independent of the model used to generate our background profile and EOFs, stemming from a combination of other published models. Specifically, we used the Bent model [5] to describe the shape of the F layer near its peak, with its density and height initialized respectively by means of the CCIR model [6] and a prescription by Jones et al. [7]. Below the F-layer peak, we employed the USAF Air Weather Service RBTEC model [8]. The topside profile shape was described by means of a simple two-component (\( O^+ \) and \( H^+ \)) diffusive-equilibrium model. Once initialized, the thickness and peak density of the F layer were adjusted iteratively to match the in situ density measured at each profile latitude by means of DMSP. Similarities between the “true” profile and the a priori background profile stem solely from the fact that both are based on (independent) models intended to describe the actual F layer.

To obtain simulated TEC data, we integrated through the two-dimensional fields of “true” electron density, without adding numerical noise. Our DIT procedure incorporates an estimate of uncertainty due to measurement noise via the variance vector \( \nu \) in Eq. (8). In all instances, the magnitude of \( \nu \) is vastly smaller than that of \( \nu \), the a priori model variance vector. We believe that a posteriori uncertainty from model variance that is unresolved because of “model deficiency” is substantially larger than that which would arise from measure-
ment noise in field data, although we have not tested this expectation empirically. The limited number of Fourier harmonics and small number of EOFs employed constitute the primary model deficiencies. Simulated records derived from in situ data taken by means of the Special Sensor for Ions, Electrons, and Scintillation (SSIES) on board DMSP Satellite F8 on 24 January 1990 appear in Figure 2 for five hypothetical TEC receiving stations located between 58 and 68° north latitude. (Our processor operates in geomagnetic coordinates, treating plasma density as constant in geomagnetic longitude, but we used geocentric and geographic coordinates interchangeably in our tests.) The effective sample period in all of our simulated data approximated 3 sec.

Presently, our processor is configured for data from high-inclination sources in low earth orbit, such as on board the “Transit” satellites of the U.S. Navy Navigation Satellite System. It employs all available rays above 5° north and south of the northernmost and southernmost receiving stations. Figure 3 presents results for a simulated pass in which we assumed that observations were available only when the satellite was between 53° and 73° latitude (so that the processor used the shorter curve in Fig. 1). The upper panel shows the “true” electron-density field derived from the DMSP data, and the lower one shows the tomographic reconstruction obtained. The processor generated the geometry matrix for this case in 8 min 35 s on a Sun SPARCstation 10 Model 30 with 64 Mbytes of memory and inverted the matrix in 9 min 27 s. (Neither of these operations requires the data to be in hand; they may be performed at any convenient time after the appropriate satellite ephemeris is available.) Multiplying the inverted matrix by the data and transforming the result to configuration space to produce the image shown in the lower panel took 11 s on the same machine. (All times include interaction with a UNIX script.)

In Figure 3, horizontal structures on scales down to somewhat less than 1° of latitude (about 100 km in extent) are reproduced rather well. (The wavelength of the highest harmonic employed for all inversions illustrated in this article was 0.67°, unless otherwise noted.) Smaller structures are not fully resolved. For instance, the reconstructed $1.0 \times 10^{11} \text{ el/m}^2$ contour is barely broken near 61° latitude and is not broken at all near 64° latitude, whereas the “true” contour is completely pinched off in both places. Vertical rendering is less satisfactory; the entire reconstructed layer is somewhat too low and the bottomside gradient is too shallow. Note, for instance, that the first contour lies near 200 km in the lower panel, whereas it is consistently above 200 km in the upper one.

To identify the effect of improved vertical resolution, we

![Figure 2](image)

**Figure 2.** Simulated TEC records for five hypothetical receiving stations, derived from SSIES records of plasma density recorded on DMSP F8 on 24 January 1990. Short curves are employed in first test discussed in text; long ones are employed in second test.
reran this case with lower-elevation (and therefore lower-grazing-angle) data (employing all rays that the processor is designed to accept, as described above). This result, obtained using the longer TEC curves in Figure 2, is shown in the lower panel of Figure 4. There is modest vertical improvement, with the first contour on the bottomside now being slightly above 200 km and the vertical placement of most other contours being improved as well.

It is interesting that the additional rays have also improved the horizontal resolution. The $1.0 \times 10^{11}$ contour now breaks essentially where it should. There have also been minor losses, however, such as the foreshortening of the 2.0 contour equatorward of the main ion trough and disappearance of the two 1.5 contours in the resolved feature near the center of the plot. Moreover, the additional rays substantially increased run time. Now generating the matrix took 31 min 21 s, and inverting it took 49 min 21 s. Generating the image from the data took 27 s. In the next two cases, we will use only rays emanating from between $53^\circ$ and $73^\circ$. For them, the run time was similar to (slightly shorter than) that quoted for the case in Figure 3.

In Figure 5, we present results from a higher-density case than the foregoing, having somewhat less prominent horizontal structures in the trough region and auroral F layer. Again, the "true" electron-density field is displayed in the upper panel and the tomographic result in the lower one. The gray scale is the same in the present case as in the preceding (and succeeding) one, to facilitate comparison of absolute densities. Here, the midlatitude F layer reaches a peak density between $6.5 \times 10^{11}$ and $7.0 \times 10^{11}$ el/m$^3$, a fact reproduced in the tomographic reconstruction. The reconstruction places the equatorward wall of the main trough (closing of the $4.0 \times 10^{11}$
contour) accurately near 62.4° latitude. All major horizontal structures are reproduced, although their peak densities are slightly underestimated. That the peak density is estimated within $0.5 \times 10^{11}$ el/m$^3$ (somewhat more closely if finer contours are plotted) does indicate a reasonable level of success in dealing with the $2\pi \sigma$ ambiguity. Vertical resolution again is less satisfactory, with the layer being too low and the bottomside gradient too shallow.

Figure 6 contains results from a case combining the greater density range of Figure 5 with the more prominent horizontal structure of Figures 3 and 4. They are consistent with the foregoing cases, with horizontal resolution being superior to vertical rendering and with peak density being underestimated by about $5 \times 10^{10}$ el/m$^3$ (8%).

Finally, in Figure 7 we present a somewhat simpler case, containing only a single auroral F-layer “blob” poleward of the main trough, which was presented to us “blind.” The original contours appear in Figure 6 of the paper by Raymundo et al. [9], and renderings of it with our absolute-TEC processor are contained in Figures 4 and 6 of Fremouw et al. [1]. That processor employed (a different set of) eight EOFs and 30 harmonics. In our present work, we have found that investing computer time in generating and inverting matrix elements pertaining to more than about six EOFs is not cost-effective, because of the limited vertical resolution available inherently from surface-to-orbit viewing geometry. Investing in higher harmonics is productive, however, to control ringing when resolving sharp horizontal gradients such as that in the present case. In this case, the wavelength of the highest harmonic was 0.60° of latitude (about 60 km).
A comparison of Figure 7 with the original contours in Raymund et al. [9] reveals the following characteristics of this case. As with the other cases, horizontal rendition is quite good, with placement of major features (trough and "blob") being accurate within 0.2° (about 20 km), which is the size of a plot resolution cell. Compared with our earlier processor [1], vertical rendering and resolution are somewhat improved. The bottomside contours (starting with $4 \times 10^{10}$ el/m$^3$, the lowest one in the original field [5]) for the most part are placed appropriately within a resolution cell (20 km) or so. The topside gradient is improved, but still somewhat shallower than the original. The F-layer peak is about 20 km (one resolution cell) too high, but its tilt (increasing height with increasing latitude) has been rendered quite well with five stations spanning 8° of latitude. The estimate of peak density is about $5 \times 10^{10}$ el/m$^3$ too low, whereas that of our previous processor was about $4 \times 10^{10}$ too high. Thus, we have not lost much in terms of ambiguity resolution in going from an absolute-TEC processor to one that operates directly with relative TEC data.

The fact that the F-layer tilt present in the original contours [9] was reproduced in our inversion was encouraging and surprising. As discussed in Fremouw et al. [1] and by other authors, vertical resolution achievable with orbit-to-surface tomography inevitably is coarse, as a result of geometric limitations on grazing angle. We did not impose peak-height variations on our self-generated test data to investigate vertical resolving power, which should be done in the future.
Presently, we are taking a more formal approach to such investigation by computing the effective vertical (and horizontal) correlation function of our process.

IV. CONCLUSION
From this work, we conclude that weighted, damped, least-squares DIT holds substantial promise for application to ionospheric tomography using relative TEC obtainable by means of existing satellites in low earth orbit. Aside from applying the technique to such data, three kinds of productive work could be performed now. First, the possibility should be explored that a background profile and corresponding EOFs tailored to specific observing conditions would further improve tomographic renderings. For the processor developed in this work, an efficient approach would be to incorporate functions from our existing model "data base" for specific combinations of receiver-chain location, time of day, season, and sunspot number. Second, this approach then should be adapted to incorporate external field data such as a bottoms-side ionosonde profile. Third, a hybrid EOF-Fourier technique should be developed that would introduce into the problem the ionosphere's second natural coordinate system (with one axis aligned along the geomagnetic field).

ACKNOWLEDGMENTS
This article was prepared under IR&D funding from Northwest Research Associates, Inc. (NWRA). The tomographic processor employed for the work reported was developed by NWRA under Small Business Innovative Research Contract No. F19628-90-C-0116 from the Geophysics Directorate of the
Tomographic Reconstruction: $N_e$ (10$^7$ el/m$^2$)

Figure 7. Simulated results from a case presented to us "blind." For original contours, see Raymund et al. [9]; for renderings by our absolute-TEC processor, see Fremouw et al. [1].

U.S. Air Force Materiel Command's Phillips Laboratory. We thank J. A. Klobuchar of Phillips Lab for the simulated TEC data used in producing Figure 7 and for his continued interest in our work.

REFERENCES