A three-dimensional time-dependent algorithm for ionospheric imaging using GPS

Cathryn N. Mitchell and Paul S.J. Spencer

Department of Electronic and Electrical Engineering, University of Bath, U.K.

Abstract

Global Positioning System (GPS) satellite receivers provide a world-wide network of phase and group delay measurements. The combination of two-frequency measurements can be used to derive the integral of the electron concentration along each satellite-to-receiver path, a parameter known as the Total Electron Content (TEC). At this stage these slant TEC data are difficult to interpret as they originate from a combination of a temporally changing ionosphere and spatially changing observation geometry. In this paper TEC data are inverted to evaluate the underlying distribution and time evolution of electron concentration. Accordingly, a new three-dimensional, time-dependent algorithm is presented here for imaging ionospheric electron concentration using GPS signals. The inversion results in a three-dimensional movie rather than a static image of the electron-concentration distribution. The technique is demonstrated using simulated ground-based GPS data from actual measurement geometry over Europe.

Key words *ionosphere – imaging – GPS – inversion – tomography*

1. Introduction

The ionosphere results from the interaction of solar emissions with the Earth's atmosphere. Its evolution is influenced by changes in solar emissions, atmospheric dynamics and the interplanetary- and geo-magnetic fields. Ionospheric morphology results from the interaction of these drivers and isan interesting topic for scientific study. Special instrumentation devoted to such studies has been developed and used extensively for many years.

The ionosphere affects electro-magnetic waves by refraction and the magnitude of this is frequency dependent. The result of this is that signals are bent and delayed in a manner that may be either advantageous or disadvantageous to radio systems. For example, HF communication beyond the horizon is possible only because oblique signals can be refracted and returned to the ground.In this case the user can determine the path of a signal by ray-tracing through a map of the electron concentration. *L*-band signals used in satellite navigation systems based on the time-of-flight of the signal are put in error by the unknown ionospheric delay. For the user needing to compensate for this error, the requirement is to be able to know the delay introduced by the Total Electron Content (TEC), along specific satellite-to-receiver ray paths. Again, this requires accurate knowledge of the distribution of electron concentration.

Recently, dual-frequency satellite signals have provided a new source of ionospheric information. The Global Positioning System (GPS) satellites are monitored by a network of dual-frequency receivers, recording the phase and time delay of each signal. These measure-

Mailing address: Dr. Cathryn N. Mitchell, Department of Electronic and Electrical Engineering, University of Bath, Claverton Down, Bath, BA2 7AY, U.K.; e-mail: C.N.Mitchell@bath.ac.uk

ments provide a valuable source of information about the ionosphere in the form of ray-path integrations of electron concentration. These measurements are rather uneven in distribution and coverage but have some similarities to polar-orbiting low-Earth orbit satellite TECs derived from the Navy Ionospheric Monitoring System (NIMS) satellites. In contrast to GPS TECs, NIMS satellite TECs map easily into a plane and thus can be inverted into images of electron concentration using tomographic methods.

Tomography is the inversion of networks of line-integral measurements to yield the two-dimensional spatial distribution of the integrated parameter. This is very important for ionospheric data combination since many measurements of the line integral of the electron concentration (*i.e.* the TEC) are easily obtained, but the parameter of interest is often the electron concentration itself. In the case of conventional ionospheric tomography, TEC measurements are made from dual-frequency trans-ionospheric signals along many intersecting raypaths between a LEO satellite and a chain of ground-based receivers. These TEC data are subsequently analysed and inverted using a suitable mathematical algorithm to produce two-dimensional images of electron concentration. A review of the experimental results from ionospheric tomography has been published by Bernhart *et al.* (1998). The application of the tomographic technique to ionospheric imaging was first proposed in 1986 by Austen *et al.* LEO satellite-to-Earth measurement geometry only allows observations over a limited number of viewing angles and Yeh and Raymund (1991) investigated some of the theoretical limitations this would impose on ionospheric tomography. In fact the orientations of the satellite-to-receiver ray paths are biased in a vertical sense with no ray paths running horizontally through the ionosphere. Consequently the vertical electron-concentration gradients are poorly defined by the TEC measurements alone. This geometrical constraint means that the reconstruction of the tomographic images is not straightforward and the reconstruction cannot be implemented directly using conventional inversion algorithms alone. There have been several solutions to the problem and the one described by Fremouw *et al.* (1992) has been widely used. This algorithm, adapted

from another geophysical application, uses a set of vertical orthonormal functions, created from ionospheric models, to image the vertical profile and a power law spectrum to select the horizontal structures from a Fourier basis. The algorithm presented in this paper is an extension into three-dimensions of that presented by Fremouw *et al.* (1992) for two-dimensional tomographic imaging. In addition, a time-dependence has been incorporated into the inversion, allowing for changes in electron concentration throughout the imaging period. This allows the algorithm to invert multi-directional slant TEC data derived from GPS signals over a period of typically one hour into movies of electron concentration.

2. Theory

In this section the theory relating directly to the inversion problem is outlined. A detailed description of the propagation of electromagnetic waves through an ionised medium and the effects of the ionosphere on radio systems can be found in Davies (1990).

Total electron content is defined as the line integral of the electron concentration along a path from a satellite, S, to a receiver, R. The TEC (b_s) may be expressed as

$$b_s = \int_S^R N(r, \theta, \phi)\, ds \qquad (2.1)$$

in which N is the electron concentration, r the radial distance from the centre of the Earth, θ is the latitude, ϕ the longitude and s the distance along the satellite-to-receiver ray path. TEC can be measured using Faraday rotation or differential Doppler techniques. In the latter case, dual-frequency radio signals that propagate through the ionosphere are subject to a differential phase change due to the dispersive nature of the plasma. As a first order approximation the change in the differential phase shift is directly proportional to the change in Total Electron Content (TEC) between the transmitter and receiver. In addition to the phase technique, GPS signals also contain information about the signal delays. The measurement of differential delay is subject

to systematic errors known as interfrequency biases associated with the satellite and receiver hardware and to random and systematic errors due to multipath. Nevertheless, it is still useful to utilise differential delay measurements for TEC calibration as has been done by others, for example see Ciraolo and Spalla (1997).

The first stage of the inversion is to set up a three-dimensional grid of voxels (*i.e.* volume pixels), each bounded in latitude, longitude and altitude, and to compute the length of each element of a satellite-to-receiver signal propagation path though each intersected voxel. For i TEC measurements and j voxels the unknown electron concentration is defined to be constant within each voxel and contained in the column vector x of j components. The problem may now be expressed as

$$Ax = b \qquad (2.2)$$

where A is an $i \times j$ matrix of the path lengths within each voxel and b are the i observed TECs. This cannot be solved directly as the matrix A is rectangular, highly singular and incorporates no prior information as to the likely solution. To overcome this difficulty a mapping matrix, X, is used to transform the problem to one for which the unknowns are n coefficients of orthonormal basis functions, the combination of which will give the final image of electron concentration.

The choice of orthonormal basis functions is critical in the determination of the final solution to this underdetermined inversion problem. Here the basis functions (X) were generated using a spherical harmonic expansion to represent the horizontal variation and Empirical Orthonormal Functions (EOFs) for the radial variation in electron concentration. The spherical harmonics provide a flexible basis to determine the horizontal distribution of ionisation which should be well defined by the measurements. The EOFs form a constraint to the vertical profile, only allowing a certain range of possible solutions.

This is now expressed mathematically as

$$AXW = b \qquad (2.3)$$

where the matrix X contains the basis functions, such that AX defines a basis set of line integra-

tions of electron concentration. Since the ray-path integrals derived from differential phase along the same satellite-to-receiver paths are subject to an unknown cycle offset adjacent rows of the matrices AX and b can be differenced in a manner to negate the effect of this on the solution. The n unknowns, W, now represent the relative contribution of the basis functions where

$$W = (AX)^{-1}b. \qquad (2.4)$$

Here $(AX)^{-1}$ is a generalized inverse matrix such that W is the maximum likelihood solution. In this case Singular Value Decomposition (SVD) has been chosen to solve the inversion. Applying singular value decomposition to the matrix AX returns two orthogonal matrices U and V and a diagonal matrix of singular values, w, thus

$$(AX)^{-1} = V(\text{diag}(1/w))U^T. \qquad (2.5)$$

The solution to the inverse problem is then given by

$$W = \left(V(\text{diag}(1/w)U^T)\right)b \qquad (2.6)$$

where the reciprocal of the terms in w that are sufficiently small, typically 10^{-7} of the dominant singular weight, are zeroed to account for degeneracy in the AX matrix. A description of the use of SVD can be found in Golub and Van Loan (1989). Finally, the electron densities within each voxel, j, are recovered using

$$x = XW. \qquad (2.7)$$

GPS satellites have an orbital period of just under 12 h and are inclined at 55° to the equatorial plane. From a fixed location on the Earth they appear to move slowly across the sky and for high-elevation passes take several hours from rise to set time. A series of TEC measurements from any single GPS-to-ground link consist of both spatial and temporal changes in the ionosphere. The *time-independent* algorithm described is based on the assumption of stationarity of the medium and for the ionosphere this generally requires a short measurement time of a few minutes. This short measurement period would result in limiting the data quantity and spatial

coverage when using GPS. Extending this algorithm into a *time-dependent* inversion allows a great increase in the quantity and angular coverage of measurements that can be used in each inversion. In addition, since there is no implicit stationarity on the ionospheric electron concentration distribution it allows moving features to be imaged in a consistent manner rather than being averaged out. The algorithm can be extended into a time-dependent inversion by incorporating *a priori* information about the evolution of the electron concentration during a specified period of time. Assuming that the change of the electron concentration within a voxel with time is linear, which makes sense if short enough time periods are dealt with, then it is possible to write the same system of equations to solve for the change in the relative contributions of each basis function.

A relatively short period is chosen for the time-dependent inversion, for example one hour, and data collected at typically 30 s intervals are considered. The change in the ray path geometry, defined in the D matrix, multiplied by the unknown change in electron concentration (y) is equal to the change in TEC (c)

$$Dy = c. \qquad (2.8)$$

Thus the matrix, D, is formed from the difference in ray-path geometry at successive time intervals, y is the unknown change in electron concentration between each time interval and c are the measured changes in TEC. The mapping matrix, X, is again used to transform the problem to one for which the unknowns are the linear changes in coefficients (G) of a set of n appropriately selected ortho-normal basis functions

$$D(XG) = c \ . \qquad (2.9)$$

The matrices can be re-written such that the ray path geometry is multiplied directly by the mapping matrix to create the basis set

$$DX(G) = c \qquad (2.10)$$

and the change in the unknown contributions of each of these line integrations of electron concentration is solved for

$$G = (DX)^{-1} c \qquad (2.11)$$

The time-dependent solution to the inverse problem is then given by

$$y = XG \qquad (2.12)$$

and this can be combined with the static solution of (2.2)-(2.7).

Critical factors in the solution are the choice of the vertical profile set to form the basis functions and the total time over which linear changes can be assumed. Linear changes are generally enforced over time periods of one-hour.

3. Results

In this section simulations of two- and three-dimensional static and time evolving inversions are demonstrated. For the initial simulation a two-dimensional latitude *versus* altitude model of the electron concentration was created. The model was not intended to represent any particular ionosphere but to contain features that might be difficult to image using a static solution. This is to demonstrate that a time-evolving ionosphere can be imaged with the algorithm presented here.

The vertical profile of the ionisation in the model was formed from Chapman functions (Chapman, 1931). In the initial model the ionosphere had a constant peak height at 285 km and scale height of 110 km, with a peak electron concentration of 4×10^{11} m^{-3}. In the final model the peak height to the north of the image was increased to 370 km, the scale height to 140 km and the peak electron concentration was increased to 20×10^{11} m^{-3}. A large depletion in ionisation was superimposed onto the images to represent the night-time mid-latitude structure known as the main trough. The centre latitude of the trough was 48º in the first frame and 45ºN in the final frame. Thus the changes to the ionosphere included a 3º southward movement of the trough, an increase in electron concentration and an increase in the peak and scale height to the north of the trough. The
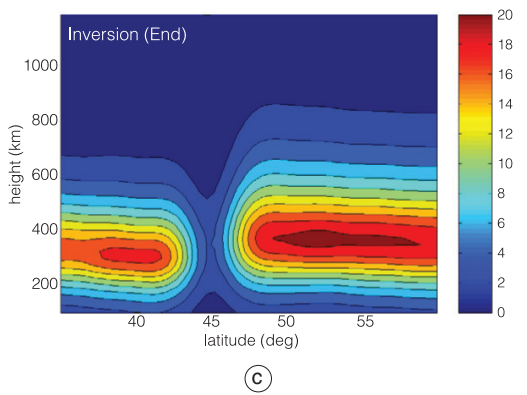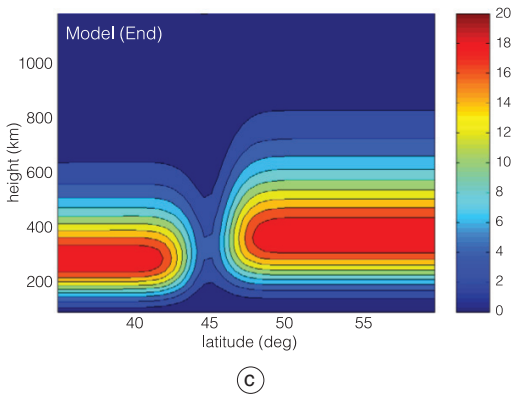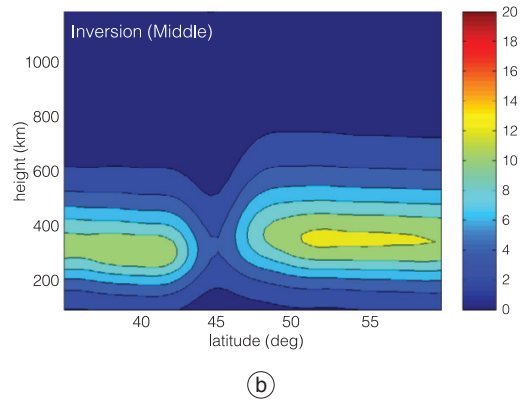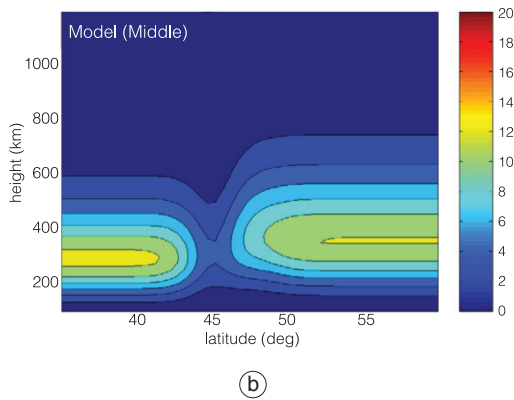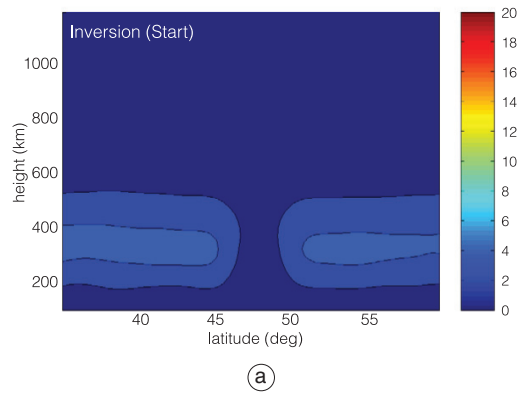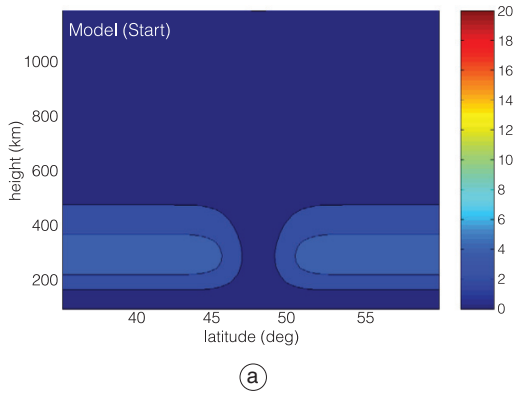
Fig. 1a-c. Two-dimensional maps of electron concentration. a) Initial frame of the model ionosphere. b) Middle frame of the model ionosphere. c) Final frame of the model ionosphere. The contours are shown every 2.0×10^{11} electrons m^{-3}.

Fig. 2a-c. Two-dimensional maps of electron concentration resulting from the time-dependent inversion. a) Initial frame of the inversion. b) Middle frame of the inversion. c) Final frame of the inversion. The contours are shown every 2.0×10^{11} electrons m^{-3}.

modelled electron concentration varied linearly between the extremes of the two images shown in fig. 1a,c. Six ground-based receivers, evenly spaced every 5° between 35° and 60°N, were used in the simulation. Two counter-orbiting NIMS type satellites at 1200 km altitude were modelled. NIMS were chosen since their polar-orbits are applicable to two-dimensional imaging. A minimum of two satellites is required to separate temporal and spatial changes in this type of imaging. TECs were simulated between each satellite and receiver for 1° steps in satellite latitude. Thus a single set of twelve (two satellites and six receivers) simulated TEC measurements were found by integration through a single frame of the ionosphere. Subsequent sets of TECs were found by integration through subsequent frames of the ionosphere according to changes described above, relating to the time-evolution.

The time-dependent inversion results are shown in fig. 2a-c. The three images are taken from the first, middle and final frames of the inversion. Comparing fig. 2a with fig. 1a it can be seen that the image represents the peak height, scale height and electron density quite well. The latitudinal location of the trough and its depth are in agreement with the model. Comparing fig. 2b with fig. 1b it can be seen that the electron concentration is slightly low to the south of the trough and high to the north. The contour to the far north of the image is closed, indicating that the inversion has produced a slightly low peak density to the northern edge of the image. Nevertheless, the scale heights are well reproduced by the inversion and the step in peak height across the trough is clearly seen. The latitudinal location of the trough has also been reconstructed well. The final image (fig. 2c compared to fig. 1c) shows that the features are well replicated but the electron concentration is slightly higher than that of the model to the north of the trough. There is a slight gradient in the peak height in the reconstructed image that is not in the model. Nevertheless the overall features and the changes in the ionosphere have been very well accommodated in this time-dependent inversion.

In order to make a comparison, the inversion was also performed using the time-independent algorithm. The result is shown in fig. 3. Sur-
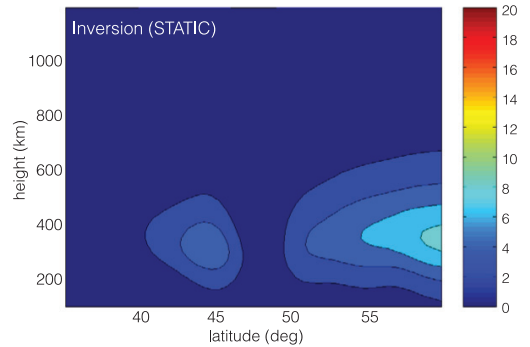

Fig. 3. Electron concentration determined using the time-independent (static) inversion. The contours are shown every 2.0×10^{11} electrons m^{-3}.

prisingly, the resulting image has a lower peak concentration than that found in the middle frame of the model. Some artificial structuring has appeared in the imaging, caused by the inconsistencies between the measurements. This shows that it is not possible to rely on the static imaging if extreme density gradients and moving features could be present in the ionosphere.

The time-dependent algorithm in three-dimensions is now demonstrated using a one-hour simulation. The ionosphere over Europe was modelled using the IRI model (Bilitza, 1990) at 2 min intervals from 06 to 07 UT on 20 April 2002. The IRI model was chosen because of its ability to represent the ionosphere globally in three-dimensions without any discontinuities, thus allowing smooth integration with which to simulate TEC. Actual GPS satellite orbits and 15 receiver locations from the International GPS Service network were used. The slant TECs were simulated by integration along the satellite-receiver path every two minutes throughout the hour through high-resolution grids of IRI electron densities. The high-resolution grid, ensuring no discontinuities in the simulated TECs, was 1° latitude by 2° longitude and 25 km altitude extending from 35ºN to 65ºN in latitude, 10ºW to 40ºE in longitude and 100 km to 1200 km in altitude. The inversion was based on the same grid.

Figure 4a,b shows the vertical TEC from the IRI model at 06 UT (fig. 4a) and 07 UT (fig. 4b). A smooth gradient in TEC is seen rising
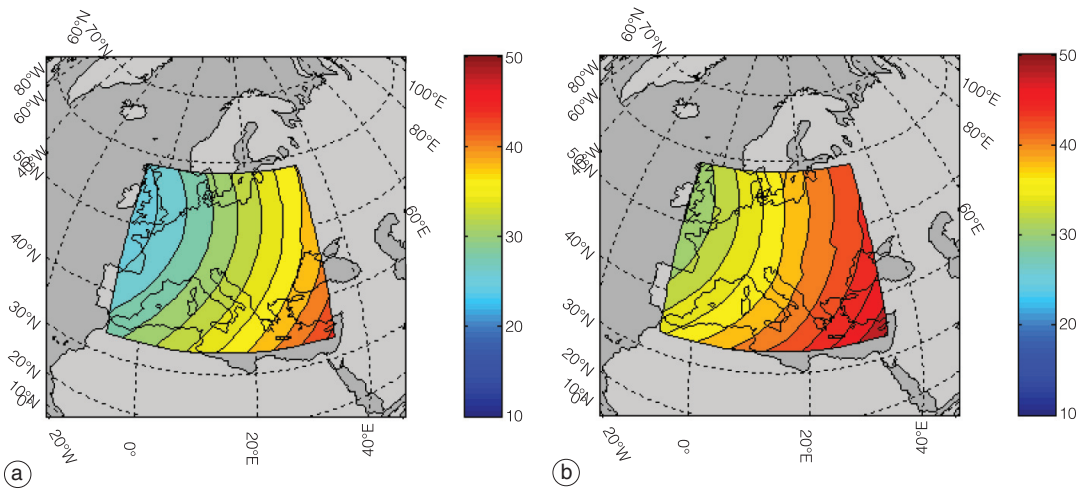
Fig. 4a,b. Vertical TEC from the IRI-95 model for (a) 06 UT and (b) 07 UT on 20 April 2002. Contours are shown every 2 TECU.
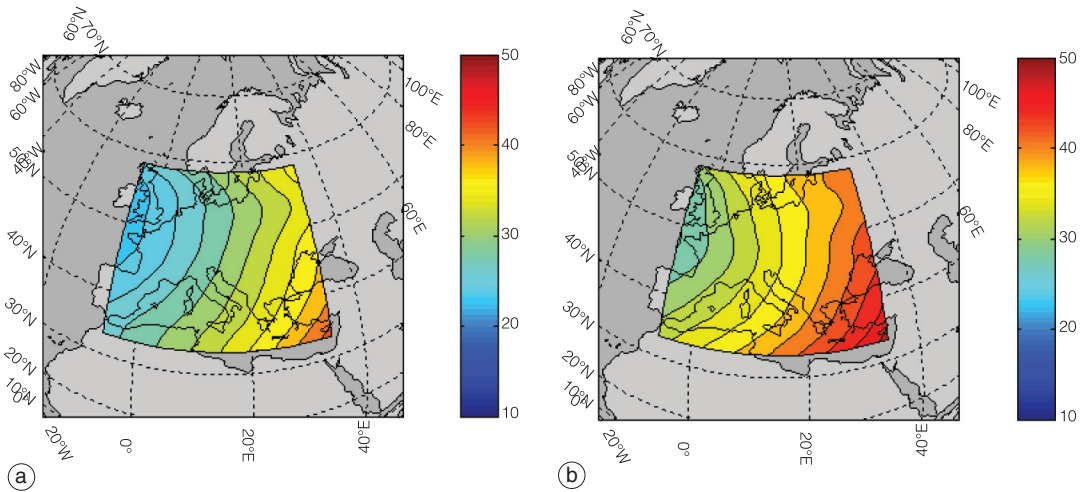

Fig. 5a,b. Vertical TEC resulting from integration through the time-dependent inversion of model phase data from ground-based receivers and GPS satellites for (a) 06 UT and (b) 07 UT. Contours are shown every 2 TECU.

in magnitude toward the south-east. Over this one-hour period the model shows an overall increase in TEC of around 15%. The lower two panels of the figure show the vertical TEC obtained by integration through the first and last frames of the three-dimensional time-dependent reconstruction. Comparing fig. 5a with fig. 4a, it can be seem that the inversion has produced a very accurate representation of the zonal and meridional gradients in TEC. The reconstruction shows a slight underestimation of TEC towards the south of the image and overestimation in

the north. There is also some evidence of an artificial change in the zonal TEC gradient to the north of the image, resulting from the use of spherical harmonics and a limited number of discrete voxels in the image reconstruction. The last frame of the inversion (fig. 5b) shows the vertical TEC at 07 UT. Again the image shows an accurate representation of the zonal and meridional gradients and a slight underestimation of TEC. This underestimation may be the result of the limited number of EOFs used in the inversion.

Twenty-four simulations were run to test the inversion method over an entire day. One three-dimensional time-dependent inversion was done for each one-hour period of the 20 April 2002. Each reconstruction was evaluated by comparison of vertical TEC between the reconstruction and the IRI model at the start of the hour, to determine the mean and the mean absolute error in both TEC units and as percentages of the model TEC.

The mean errors in TEC were found using

$$\frac{1}{Nb}\sum_{i}^{Nb}\left(binv_i - biri_i\right) \qquad (3.1)$$

where $biri$ are the 750 (Nb) vertical TECs found by integration through the IRI model represented on the 30 latitude $\times$ 12 longitude grid used for the inversion and $binv$ are the corresponding values found by integration through the electron concentrations in the inversion.

The mean absolute errors in TEC were found using the following equation:

$$\frac{1}{Nb}\sum_{i}^{Nb}\left|binv_i - biri_i\right| . \qquad (3.2)$$

The mean errors in TEC as percentages were found using

$$\frac{100}{Nb}\sum_{i}^{Nb}\frac{\left(binv_i - biri_i\right)}{biri_i} \qquad (3.3)$$

and the mean absolute error in TEC, as a percentage, has been found using

$$\frac{100}{Nb}\sum_{i}^{Nb}\frac{\left|binv_i - biri_i\right|}{biri_i} . \qquad (3.4)$$
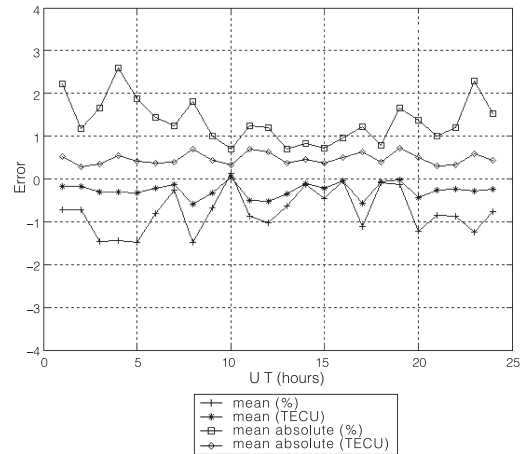

Fig. 6. Errors found between the inversion and the model over twenty-four, one-hour runs. The errors were calculated at the start of each hour using eqs. (3.1)-(3.4).

Figure 6 shows the errors, *i.e.* the differences between the vertical TEC in the image and the vertical TEC in the model, over the twenty-four hour period. It can be seen from the graphs that there is a slight underestimation of vertical TEC of the order of 1%. Slight variations occur for each reconstruction, but the accuracy is no worse than 1 TECU. The absolute values indicate that the TEC may be underestimated in some areas of the image and over in others, since at 4 UT, for example, the error is 2.5% in absolute but only – 1.5% in mean error. It is important to note that in this test a general set of basis functions were generated from spherical harmonics and orthonormal vertical functions. These basis functions happen to fit the vertical profiles of certain images better than others and the choice of them remains a critical factor in the accuracy of the images.

4. Conclusions

A new algorithm has been presented which could use carrier-phase observations to image the electron concentration in the ionosphere.

The algorithm can be applied to two- or three-dimensional spatial imaging either with or without a time evolving solution. The advantage of using the time-evolving solution has been demonstrated by means of a two-dimensional tomographic simulation. It was shown to be appropriate to use the time-dependent approach to image an ionosphere with a moving trough and increasing electron concentration. In particular, the time-independent algorithm can prevent imaging artefacts caused by moving structures and changing layer heights. Watermann *et al.* (2002) have already recognised this limitation in ionospheric imaging and suggest that ionospheric imaging in the polar cap can be relied upon when the magnetometers suggest a slower plasma convection. The time-dependent algorithm shown here could be useful for imaging the polar ionosphere, provided that sufficient satellite data are available to separate the spatial and temporal variations.

It has already been established that GPS data can be used in ionospheric imaging by using a short measurement period and combining with other measurements or models. Bust *et al.* (2001) have shown such results from their combined ionospheric campaign in the Caribbean region. The work presented here opens up the possibility to produce images of the ionosphere using GPS data on its own. The key to this is the incorporation of the time-dependence in the imaging process and this has been demonstrated with an IRI-based simulation with realistic satellite and receiver geometry taken from an actual case study. The work demonstrates the potential of the technique to use GPS data for applications requiring ionospheric mapping and for geophysical research.

In the work presented here only the differential phase technique is utilised to provide relative TEC from ground-based receivers. A future development of the work will involve the incorporation of many different data sources into the image. It is already feasible to extend the input matrix to contain any approximately linear ionospheric measurement such as information derived from ionograms. Hajj *et al.* (1994) suggested using the satellite-to-satellite transmission of GPS to LEO satellite measurements in a tomographic framework to provide the so-called «missing horizontal rays» and improve the vertical resolution. In addition, LEOs provide measurements over the oceans and into remote polar caps, thus enabling the ionosphere to be studied continuously in both space and time. Thus this technique to create global image of the ionosphere will soon be used to solve practical radio propagation problems and to study the Earth's ionosphere on a global scale.

Acknowledgements

REFERENCES

AUSTEN, J.R., S.J. FRANKE and C.H. LIU (1988): Ionospheric imaging using computerised tomography, *Radio Sci.*, **23**, 299-307.

BERNHARDT, P.A., R.P. McCOY, K.F. DYMOND, J.M. PICONE, R.R. MEIER, F. KAMALABADI, D.M. COTTON, S. CHARKRABARTI, T.A. COOK, J.S. VICKERS, A.W. STEPHAN, L. KERSLEY, S.E. PRYSE, I.K. WALKER, C.N. MITCHELL, P.R. STRAUS, H. NA, C. BISWAS, G.S. BUST, G.R. KRONSCHNABL and T.D. RAYMUND (1998): Two dimensional mapping of the plasma density in the upper atmosphere with Computerized Ionospheric Tomography (CIT), *Phys. Plasmas*, **5**, 2010-2021.

BILITZA, D. (Editor) (1990): International Reference Ionosphere, National Space Science Data Centre/World Data Centre for Rockets and Satellites, Lanham, MD, *Report n. NSSDC/WDC-A-R&S,* 90-22

BUST, G.S., D.S. COCO and T.L. GAUSSIRAN (2001): Computerized ionospheric tomography analysis of the combined ionospheric campaign, *Radio Sci*., **36** (6), 1599-1606.

CHAPMAN, M.A. (1931): The absorption and dissociation ionisation effect of monochromatic radiation in a rotating Earth, *Proc. Phys. Soc.,* **43** (1), 26.

CIRAOLO, L. and P. SPALLA (1997): Comparison of ionospheric total electron content from the NNSS and GPS, *Radio Sci.*, **32** (3), 1071-1080.

DAVIES, K. (1990): *Ionospheric Radio* (Peter Peregrinus Ltd., London, U.K.), pp. 580.

FREMOUW, E.J., J.A. SECAN and B.M. HOWE (1992): Application of stochastic inverse theory to ionospheric tomography, *Radio Sci.*, **17**, 721-732.

GOLUB, G.H. and C.F. VAN LOAN (1989): *Matrix Computations* (Johns Hopkins University Press, Baltimore), pp. 642.

HAJJ, G.A., R. IBAÑEZ-MEIER, E.R. KURSINSKI and L.J. ROMANS (1994): Imaging the ionosphere with the Global Positioning System, *Int. J. Imag. Syst. Technol.*, **5**, 174-184.

WATERMANN, J., G.S. BUST, J.P. THAYER, T. NEUBERT and C. COKER (2002): Mapping plasma structures in the high-latitude ionosphere using beacon satellite, incoherent scatter radar and ground-based magnetometer observations, *Ann. Geophysics*, **45** (1), 177-189.

YEH, K.C. and T.D. RAYMUND (1991): Limitations of ionospheric imaging by tomography, *Radio Sci.*, **26**, 1361.