TRANSIONOSPHERIC PROPAGATION STUDIES: NOVEL DATA SOURCES

R. Leitinger*

Abstract

For most locations the major data source for transionospheric propagation studies used to be the VHF beacon signal of geostationary communication satellites. The Faraday effect allowed to evaluate for the ionospheric electron content along a fixed ray path from the receiving station to the satellite. Since geostationary communication satellites have no VHF beacons any more the ionospheric research community has to use other possibilities.

Two are discussed in this contribution: the Global Navigation Satellite Systems (GNSS – presently GPS, the US Global Positioning System and its Russian equivalent GLONASS) as a data source which provides electron content with a regionally dense coverage and Ionospheric Tomography which provides two dimensional electron density profiles.

Space geodesy and related research areas (e.g., geodynamics) have established nets of GNSS receivers in fixed locations (compare e.g., [Zumberge et al., 1996]). In the near future new navigation applications like airport approach or land transport guidance systems also will make use of stationary GNSS receivers. The ionospheric correction signal gained by these fixed stations contains the electron content information along the ray paths from the receivers to the satellites. Careful consideration of transmitter and receiver influences is necessary to gain fully calibrated vertical electron content by means of the Group Delay effect. An other possibility is the use of Differential Doppler (carrier phase difference) to derive relative electron content.

There are two types of ionospheric tomography: Ground based tomography makes use of the Differential Doppler effect on the signals of the polar orbiting NNSS satellites. The satellite provides a "scan" in one direction, a meridional chain of receivers on the ground gives the intersecting ray bundles on which tomographic reconstruction is based. Space tomography uses occultation of GNSS signals received on board of a low orbiting (LEO) satellite. Classical inversion of the occultation signals provides an average height profile of electron density. With several GNSS observing satellites occultations can lead to intersection of ray bundles which would allow tomographic reconstruction. Otherwise the combination of single LEO satellite data with ground based data can be used for ionospheric tomography. This possibility is realistic already now: presently one GPS receiver is in orbit (GPS–MET) and the probability is very high that more receivers will be launched in the near future.

**Institut für Meteorologie und Geophysik, Universität Graz, AUSTRIA*

1 Introduction

For most locations the major data source for transionospheric propagation studies used to be the VHF beacon signal of geostationary communication satellites. The Faraday effect allowed to evaluate for the ionospheric electron content along a fixed ray path from the receiving station to the satellite. An other important data source is provided by the (US) Navy Navigation Satellite System (NNSS) which allows to derive latitudinal profiles of electron content by means of the Differential Doppler effect on the coherently transmitted 150/400 MHz signals. The PRIME (COST 238) TEC model is based on NNSS data (see [Bradley, 1995]). (No long term NNSS studies have been made outside Europe.)

Since geostationary communication satellites have no VHF beacons any more the ionospheric research community has to use other possibilities.

Two are discussed in this contribution: the US Global Positioning System (GPS) and its Russian equivalent (GLONASS) as a data source which provides electron content with a regionally dense coverage and ionospheric tomography which provides two dimensional electron density profiles.

Space geodesy and related research areas (e.g., geodynamics) have established nets of GPS receivers in fixed locations. In the near future new navigation applications like airport approach or land transport guidance systems also will make use of stationary GNSS (Global Navigation Satellite Systems – presently GPS and GLONASS) receivers. The ionospheric correction signal gained by these fixed stations contains the electron content information along the ray paths from the receivers to the satellites. Careful consideration of transmitter and receiver influences is necessary to gain fully calibrated vertical electron content by means of the Group Delay effect. An other possibility is the use of Differential Doppler (carrier phase difference) to derive relative electron content. The best quality is obtained by using temporally smoothed (filtered) group delay data for calibration combined with Differential Doppler data for resolution.

There are two types of ionospheric tomography: Ground based tomography makes use of the Differential Doppler effect on the signals of the polar orbiting NNSS satellites. The satellite provides a "scan" in one direction, a meridional chain of receivers on the ground gives the intersecting ray bundles on which tomographic reconstruction is based. Space tomography uses occultation of GNSS signals received on board of a low orbiting (LEO) satellite. Classical inversion of the occultation signals provides an average height profile of electron density. With several GNSS observing satellites occultations can lead to intersection of ray bundles which would allow tomographic reconstruction. Otherwise the combination of single LEO satellite data with ground based data can be used for ionospheric tomography. This possibility is realistic already now: presently one GPS receiver is in orbit (GPS/MET on MicroLab 1, see Ware et al., [1996]) and the probability is very high that more receivers will be launched in the near future.

2 GPS (GLONASS): calibration, TEC intercomparison, adaptive modeling

The plasma information on the signals of GNSS satellites can be extracted in two ways: using Group Delay (delay of modulation phase – GD) or using Differential Doppler (carrier phase difference – DD). Most users think in terms of GD. With adequate knowledge of the delays in the satellite transmitter and in the ground receiver this gives absolute ("calibrated") values of electron content from the transmitter (T) to the receiver (R), $I_G(t) = \int_{\mathrm{T}}^{\mathrm{R}} N_e\ ds + b + \nu_G$ (t: observation time, N_e: electron density, ds: ray path element, b "bias" (residual calibration error), ν_G: random fluctuation). GD is less sensitive than DD and therefore accuracy and resolution are limited. Low elevation has to be avoided (multipath problems). Furthermore, the GNSS signals are deliberately degraded and disturbed most of the time which leads to very large fluctuations of individual I–values. Long term averages (integration time $\geq$ 15 minutes) only are usable. DD is much more sensitive but because phase differences can be measured modulo 2π only ("$2n\,\pi$ ambiguity") the observations give "uncalibrated" electron content and a calibration constant has to be found: $I_D(t) = \int_{\mathrm{T}}^{\mathrm{R}} N_e\ ds + C + \nu_D$ (C: calibration constant, ν_D: random fluctuation (phase noise)). Sufficiently good values for C can be found by combining data from several receiving stations and several GNSS satellites but the best way is to use a suitable average of GD ($\overline{I_G}$) to "calibrate" DD.

There seems to be agreement now among the specialists that the "bias" and "calibration" problems which have bothered GPS users for years are solved with sufficient accuracy and reliability.

Compared with the "old" data sources (Faraday effect on the signals of geostationary satellites and Differential Doppler effect on the signals of the polar orbiting NNSS satellites) electron content from GNSS poses two problems: temporal and spatial dependencies are mixed (12 hour orbits) and the observed values are total content from the satellite in a height of about 20000 km to the ground (it is "ionospheric" plus "plasmaspheric" content). The best way to separate temporal and spatial components is the use of regional networks of receivers and to observe all available satellites from each receiver. "Now casting" procedures lead to the regional distribution of (vertical) electron content (TEC) for given time. An other approach is updating of a regional model with observed electron content. DLR Neustrelitz provides hourly maps of electron content with a delay of one or two days [Jakowski et al., 1996]. This product is available via the internet: `www.nz.dlr.de/gps/gps-ion.html` (contact: Dr. Norbert Jakowski; example: Figure 1).

Comparison of GPS–TEC with electron content from other sources is an important task, especially in view of the uncertain future the other sources. Figure 2 shows an example for the comparison of GPS with NNSS results made in Florence [Ciraolo and Spalla, 1997]. This example proves the reliability of the data interval between the quartiles if individual GPS results are used for long term studies.

Intercomparison studies have been made at other institutions, too, both in Europe and in the US.

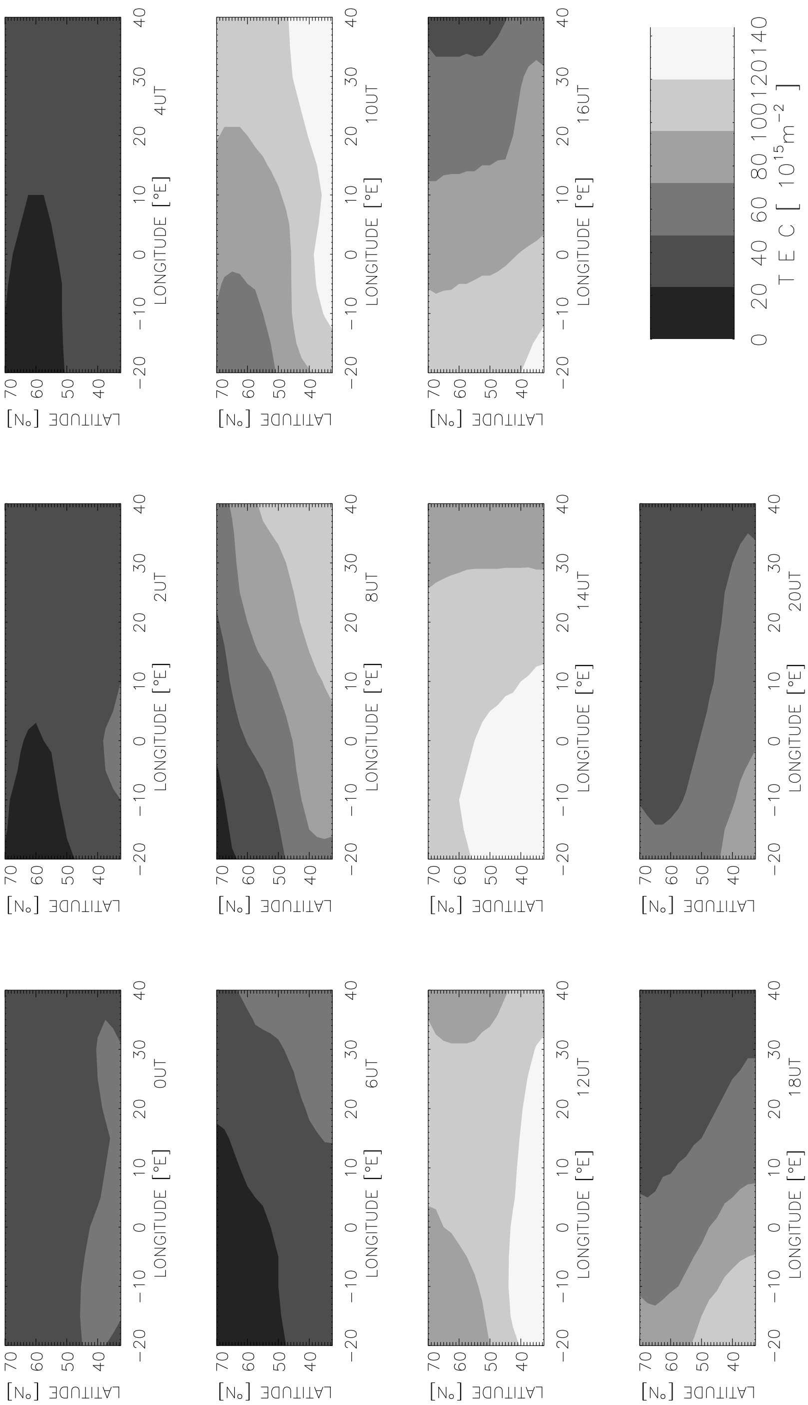

Figure 1: Example for regional TEC model, updated with GPS data. February 2, 1995, bi–hourly maps. Product of DLR Neustrelitz (`www.nz.dlr.de/gps/gps-ion.html`). (Black and white version; the original product is colour coded.)

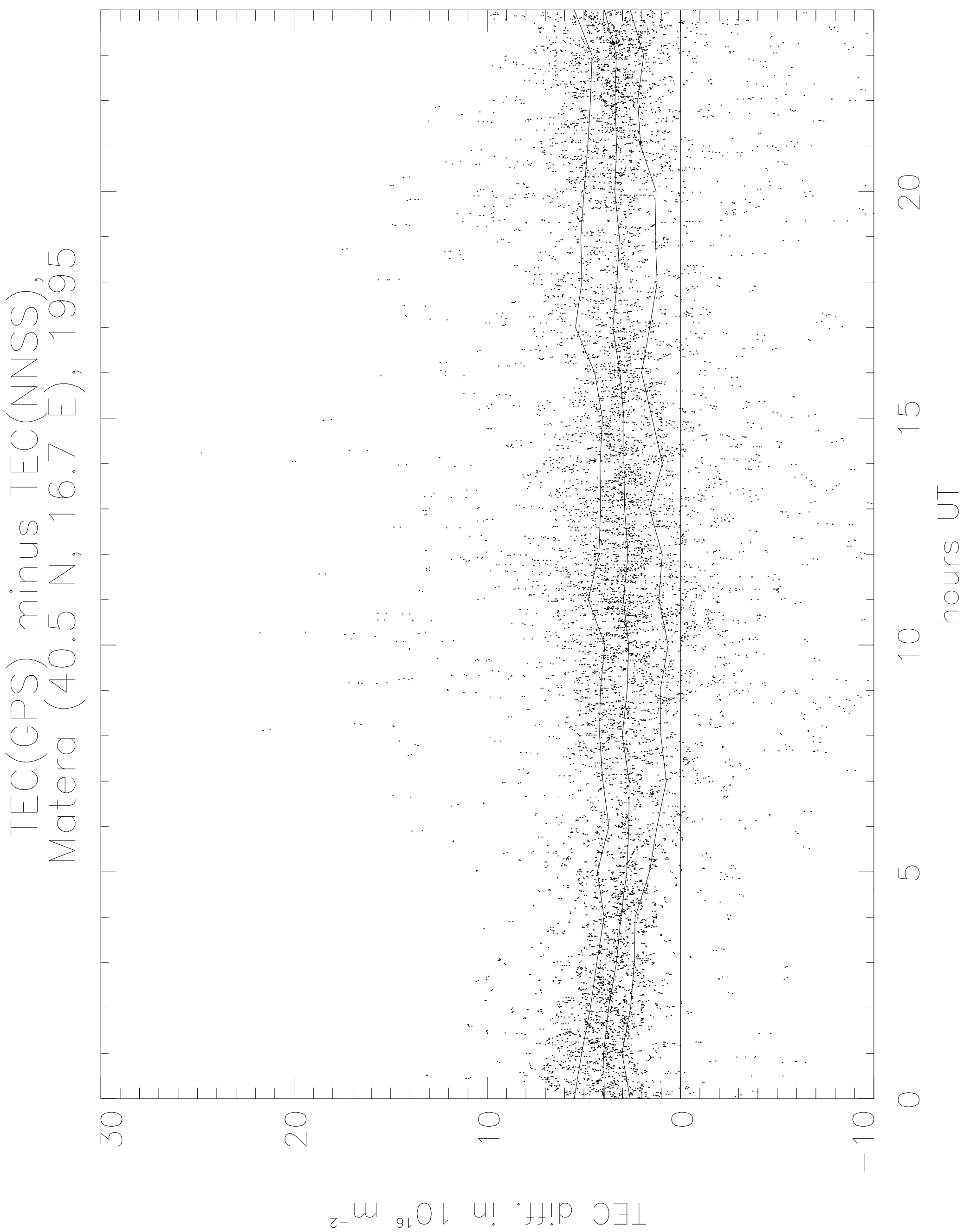

Figure 2: Comparison of TEC data from GPS and NNSS. Difference of TEC from GPS minus TEC from NNSS for 1995 in units of 10^{16} m^{-2}. Dots: individual data. The lines mark the upper quartile, the median, and the lower quartile. Note that calibration errors can influence both GPS and NNSS results. No effort was made to eliminate "bad" data. Data prepared at IROE from NNSS observations made at L'Aquila and Firenze and GPS observations made at Matera.

3 Ionospheric tomography

Ionospheric tomography makes use of slant electron content (STEC) from Differential Doppler. Presently satellite transmitters (beacons) and ground receivers are used. The orbital movement of a beacon satellite provides many transmitter locations but shifted in time — a transmitter "scans" the object. Stability of the object during the useful portion of a satellite pass is a necessary requirement. In the case of the ionosphere this requirement is fulfilled for "Low Earth Orbit" (LEO) satellites (orbital heights below about 2000 km) which scan from horizon to horizon within a fraction of an hour. The ideal geometry for tomography requires a scan around the object. Since the body of the earth is not transparent for radio waves we have to be content with partial scans and suffer from lack of data contrary to technical applications of tomography. Furthermore the number of receiver locations is very small, much smaller than in most technical applications.

We have receivers or transmitters on two opposite sides of the object but have no receivers and transmitters perpendicular to the surface of the Earth or to the orbit of the scanning satellite. The necessary requirement of intersecting "rays" is fulfilled but low elevation rays are missing and the object is not scanned in the vertical direction.

Ground based ionospheric tomography uses slant electron content I_D as the observed quantity ("projection value" in tomographic terminology). Since phases or phase differences are measured *modulo* 2π only, it is **not** a measured quantity. A constant has to be added to the measured values for "calibration". Each station needs its own constant for each satellite pass. Good estimates can be gained by means of "calibration methods" which make use of additional information or of "past experience" on ionospheric layering. Since tomographic reconstruction needs data from chains of receiving stations usually derivatives of the "two stations method" [Leitinger et al., 1975] are applied to estimate the constants. An other possibility is to leave the constants as additional unknowns for which solutions are found during the reconstruction process.

Because of the missing data problem ionospheric tomography cannot apply most of the reconstruction methods used for technical (medical) applications, e.g., filtered backprojection, overlay of sinograms, etc. In all practicable methods published so far ionospheric tomography deals with linear equation systems which are overdetermined but ill conditioned. No "unique" solutions exist but incorporation of additional information ("*a priori* knowledge") leads to "plausible" solutions. The amount of "*a priori* knowledge" used (or needed) depends on the reconstruction algorithm. Some use quite general properties and assumptions, others apply ionospheric models explicitly to fill in for missing data.

Extensive discussions and comparisons have shown what to expect and not to expect from the various evaluation methods published. It is quite clear that different ionospheric situations need different approaches. The criteria for the applicability of a specific reconstruction method are found in the properties of the primary data (slant electron contents) and in the latitude dependence of vertical electron content derived from STEC.

For the case of comparatively strong medium scale structures (e.g., main trough, auroral latitudes, medium-scale TIDs) ionospheric tomography has clearly reached the production stage (Figures 3 and 4). The reconstruction method used is based on the MART algorithm and very careful preparation of the "background ionosphere" which is used as the initial guess for the iterative algorithm [Kersley et al., 1993; Kersley et al., 1994; Pryse et al., 1995; Mitchell et al., 1995]. The incorporation of ionosonde measurements into the background ionosphere to improve the vertical profile in the tomographic image has been presented by Heaton et al. [1995]. A more recent publication by Kersley et al. [1996] shows a tomographic image with a significant transition in the height of the layer peak from the mid-latitude F-layer to an auroral E layer, verified by the EISCAT radar. Plausibility of these results is very high and we should not expect major improvements from other reconstruction methods.

One promising new approach for weaker structures (mid latitudes) is the "model less" approach used by G. Fehmers in his Ph.D. thesis [1996]. Presently his results do not give sufficient plausibility for layer height but improvements can be expected from the introduction of stricter constraints for the solutions.

4 The future of ionospheric tomography

Ground based tomography with use of the Navy Navigation Satellite System (NNSS) or its Russian equivalent (CICADA) has reached the production stage. An exponential increase of the number of reconstructions can be expected for the next years.

"Space tomography", namely the application of GNSS occultation is starting just now: GPS/MET, a multichannel GPS receiver on board of the small research satellite MicroLab 1 (nearly circular orbit in a height around 750 km, inclination: about 70°, launch date: 3 April 1995; see, e.g., Ware et al., [1996]) observes setting occultations intermittently. ESA, NASA and national Space Agencies have definite plans to launch satellites with GNSS receivers in the near future. The main purposes of these receivers will be atmospheric research (use of occultations for monitoring of the troposphere/stratosphere system) and satellite navigation and orientation. Ionosphere occultation will be a side product. The plasma along the GNSS–LEO rays influences the data used for troposphere/stratosphere research even after compensation for first order influence on carrier phase. Therefore "ionospheric" expertise is needed and it is to be expected that occultation data for ray perigees between 100 km and the LEO height will be gained and preserved.

There are several possibilities to add information to obtain tomographic systems (see [Leitinger et al., 1996, Høeg et al., 1995]).

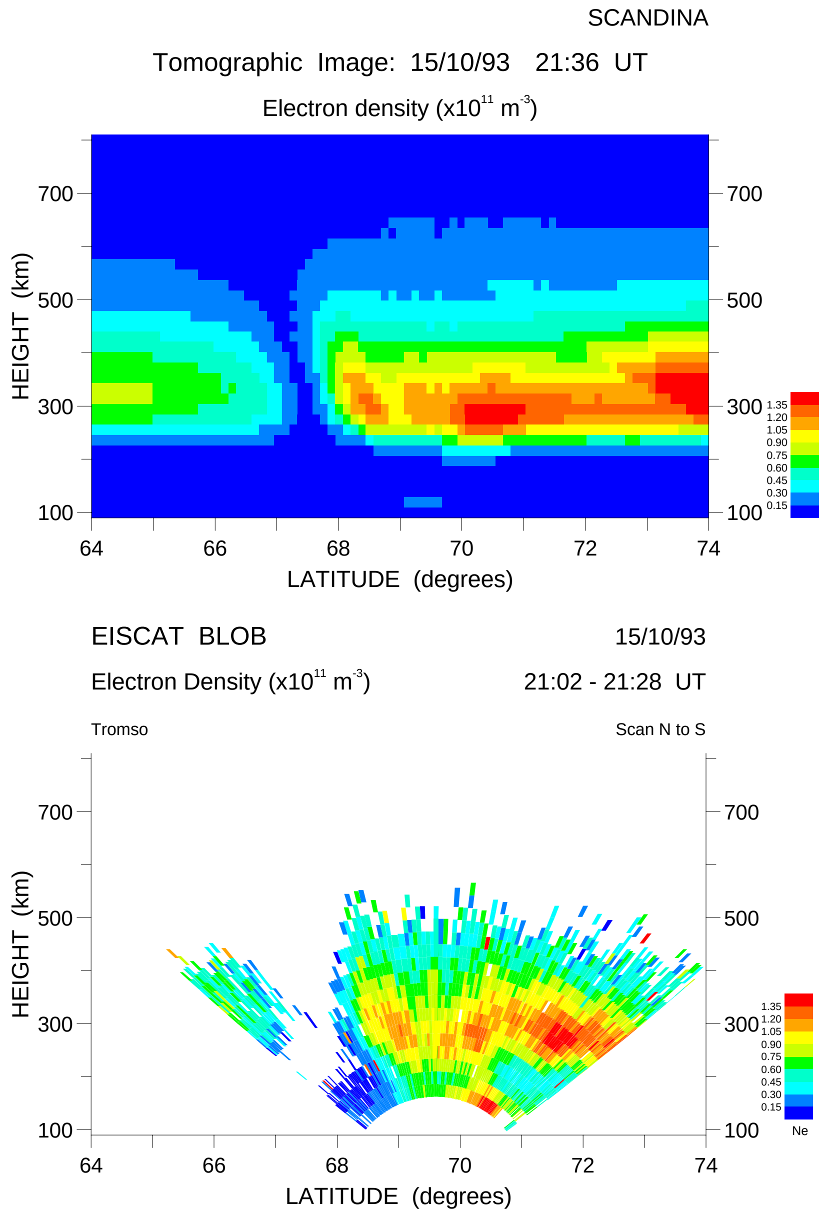

Figure 3: Example for tomographic reconstruction results: two–dimensional profiles of ionospheric electron density (isolines in a height vs. geogr. latitude system) from Mitchell et al., [1995]. sTEC *from NNSS reception by 5 stations in Scandinavia (Ny* **r***Alesund, 78.9°N, 11.9°E, Tromsø, 69.6°N, 19.2°E, Kiruna, 67.9°N, 20.4°E, Lycksele, 64.6°N, 18.8°E, Uppsala, 59.8°N, 17.6°E). Reconstructions (MART algorithm) from an NNSS pass during the evening of 15 October, 1993. A two dimensional electron density profile from an EISCAT scan made at nearly the same time is shown in the lower part of the figure. UT for the start of observations in the headers of the displays. Ionospheric structure: main trough of the F–layer and "boundary blob". Original colour coded.*

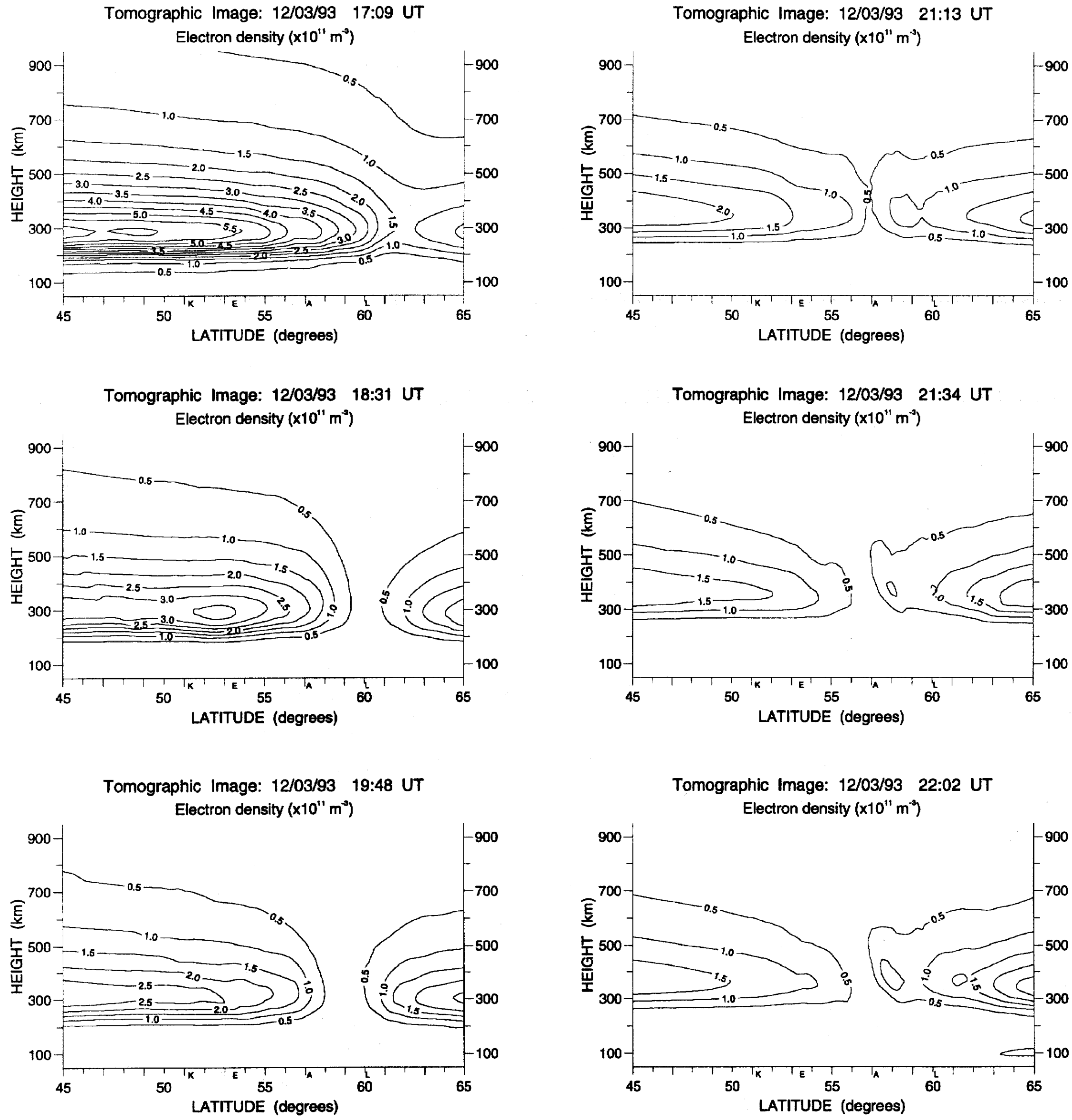

Figure 4: Example for tomographic reconstruction results: two–dimensional profiles of ionospheric electron density (isoline display in a height vs. geogr. latitude system) from Kersley et al., [1996]. Slant electron content data from reception of NNSS signals by 4 stations in the UK (Lerwick, 60.15°N, 1.15°W, Aberdeen, 57.21°N, 2.16°W, Hawick, 55.42°N, 2.82°W, Aberystwyth, 52.47°N, 4.05°W). Reconstructions (MART algorithm with careful preparation of the background ionosphere, see Heaton et al., [1995]) from 6 NNSS passes during the evening of 3 December, 1993. Universal Time for the start of observations given in the header of the display. Ionospheric structure: main trough of the F–layer.

4.1 Example for the inversion of occultation data

Figure 5 shows one example for the inversion of GPS/MET phase difference data. The occultation occurred in European mid latitudes (34.2°N, 20.7°E – Greece) on 22 October, 1995, 7:34:51 UT. The setting of the ray perigee through the ionosphere took about 6 minutes. The "ionospheric" data exist in intervals of 10 seconds. This resolution is marginally sufficient. Artificial resolution enhancement was gained by means of third order interpolation. In this way the number of data was increased by a factor of 6. The data were extrapolated from the start height to the height of the MicroLab satellite. Then linear interpolation was used to gain phase data equidistant in height (100 intervals from the peak perigee height of 738.8 km to the ground). For inversion in each height interval electron density is assumed to be a linear function of height in each height interval (polygon). The inversion procedure is a straightforward and inherently stable solution of a linear equation system with a triangular matrix for the lefthand side. The electron density values are gained from the top down. An initial oscillation of the solution damps out quickly (height range above 700 km in Figure 5). The fluctuations in the height range (450 – 700) km reflect fluctuations in the P1–P2 data which are probably instrumental. In the example of Figure 5 the results are useful down to a height of 80 km. It is interesting to note that an E layer appears despite the resolution limitation of the GPS/MET data.

Figure 5 (bottom) shows two electron density profiles. The one with the higher values was gained directly from the P1–P2 difference. A correction for ionization above the MicroLab height was applied for the lower profile: Before inversion a fixed value was subtracted from all data. For this constant we chose 90% of the P1–P2 value at the MicroLab height. There is no doubt that the correction improves the profile in greater heights. The correction has nearly no effect on the profile below 350 km and does not influence the F2 and the E layer peak heights.

5 Conclusions

The former Faraday observations are replaced by ground and space observations of GNSS signals and ground observations of NNSS signals. This is not only a replacement but offers important new research and monitoring possibilities. We lose the excellent temporal resolution and continuity of the Faraday observations as well as their simplicity and straightforward evaluation and data interpretation. We gain profile information and (at least in some regions of the world) spatial resolution. Continuity in time is still needed but is provided by ionosonde data. Major organizational efforts are necessary to make optimal use of the new research and monitoring capabilities.

Ionospheric tomography will enhance our knowledge on electron density profiles and variability and will strengthen the foundations for precise now casting for trans–ionospheric applications. Both ground based and occultation based tomography should be developed into easy to use tools. Both have advantages and disadvantages and both are not suited to replace important ground facilities like ionosondes and Incoherent Scatter installations. All have their specific places and roles in global and regional ionospheric research.

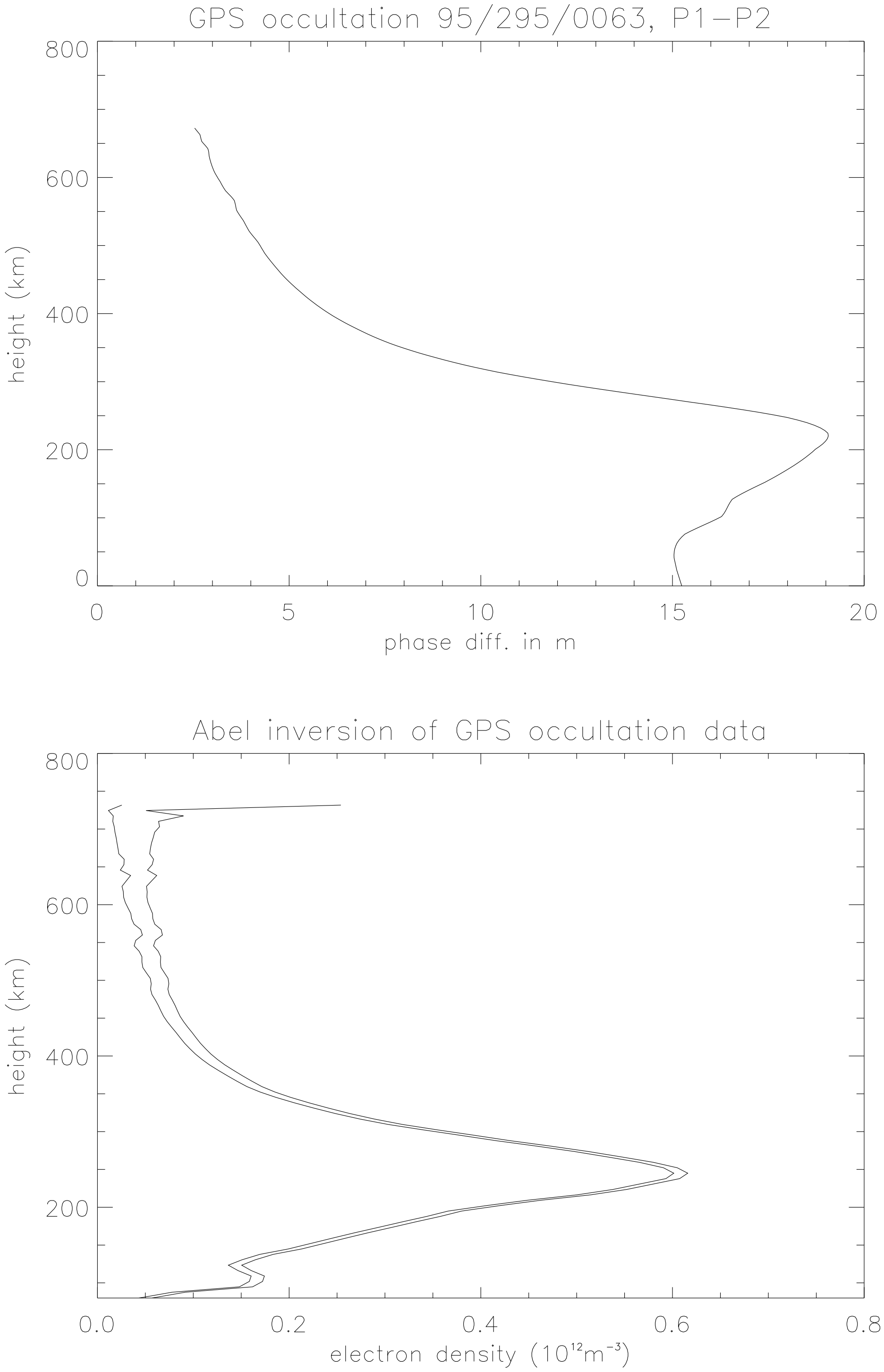

Figure 5: Inversion of GPS/MET phase difference data, occultation on 22 Oct. 1995, 07:34:51 UT. Height profiles of P1–P2 phase difference for the GPS occultation (top) and for the inversion results (bottom). Higher inversion values: without any correction for ionization above the LEO orbit; lower ones: subtraction of 90% of the electron content for the top ray. Ground point of occultation ray: 34.2°N, 20.7°E (Greece). GPS BII–02 in the north–west (end of occultation: 51.75°N, 266.96°E), MicroLab 1 in the south–west (end of occultation: 12.05°N, 36.00°E).

The main advantage of ground based tomography (with signals from polar orbiting LEO satellites) is its very good resolution for small scale structures, whereas the strength of GNSS occultation is in precise height determination and (with tomographic reconstruction) in its potential to reveal large scale structures. The combination of both, namely the use of occultation data together with well co–located ground based LEO beacon observations will give optimal results but with restricted temporal and spatial resolution. Every effort should be made to maintain NNSS and other polar orbiting satellites with coherent VHF/UHF signals in operation.

Acknowledgements: The author is indebted to GPS/MET program manager M. L. Exner, principal investigator R. Ware, and the GPS/MET data center at UCAR, Boulder, for making available the GPS/MET data used in this study.
Part of this work was supported by the European Space Agency (ESA) under contract nos. 11024/94/NL/CN and 11818/96/NL/CN.