Abstract

Data assimilation in conventional meteorological applications uses measurements in conjunction with a physical model. In the case of the ionised region of the upper atmosphere, the ionosphere, assimilation techniques are much less mature. The empirical model known as the International Reference Ionosphere (IRI) could be used to augment data-sparse regions in an ionospheric now-cast and forecast system. In doing so, it is important that it does not introduce systematic biases to the result. Here, the IRI model is compared to ionospheric observations from the Global Positioning System satellites over Europe and North America. Global Positioning System data are processed into hour-to-hour monthly averages of vertical Total Electron Content using a tomographic technique. A period of twelve years, from January 1998 to December 2009, is analysed in order to capture variations over the whole solar cycle. The study shows that the IRI model underestimates Total Electron Content in the daytime at solar maximum by up to 37% compared to the monthly average of GPS tomographic images, with the greatest differences occurring at the equinox. IRI shows good agreement at other times. Errors in TEC are likely due to peak height and density inaccuracies. IRI is therefore a suitable model for specification of monthly averages of Total Electron Content and can be used to initialise a data assimilation process at times away from solar maximum. It may be necessary to correct for systematic deviations from IRI at solar maximum, and to incorporate error estimation into a data assimilation scheme.

Key Words

ionosphere; plasmasphere; space weather; GPS; tomography; IRI;

1. Introduction

Although large areas of the earth are now covered with dual-frequency GPS measurements, there is still not sufficient coverage to specify the ionosphere globally to a resolution of even a few hundred kilometres. This is one reason that ionospheric models are still important. The International Reference Ionosphere (IRI) 2007 is an empirical model that specifies a global ionosphere. A combination of IRI 2007 and observations could be used to constrain the electron density within a coupled physical model. This would help overcome the problem of large data gaps over the oceans. However, the possibility exists that IRI 2007 could contain biases. The purpose of this study was to look for potential large, long-term biases in IRI 2007 using GPS-derived tomographic images of electron density over Europe and the US. The comparisons were performed using twelve years of data, from 1998 to 2010, in order to capture an entire solar cycle. A global study with lower resolution was performed between 2005 and 2010, since data coverage was insufficient before then. Discrepancies between the MIDAS and IRI results were investigated using independent GPS slant TEC observations.

1.1 The ionosphere

The ionosphere is an electrically conductive region of the atmosphere above 100 km (Kivelson and Russell, 1995). It is created when Extreme Ultra-Violet (EUV) radiation in sunlight ionizes neutral particles on the dayside of the Earth, creating plasma (Kelley, 1989). Within the ionosphere, particle density decreases with increasing altitude, but the intensity of EUV radiation decreases with decreasing altitude as some of the photons get absorbed on the way down. Plasma also diffuses upwards. The result is that peak plasma density occurs at around 250 – 350 km in the F region. Plasma is neutral or nearly neutral, so electron density is almost the same as plasma density. Total Electron Content (TEC) is the total number of electrons along a vertical or slant path and is designated vertical TEC or slant TEC.
1.2 Ionospheric imaging techniques

MIDAS (Multi-Instrument Data Analysis Software) is a three-dimensional, time-dependent algorithm for imaging the ionosphere. Ionospheric imaging/data assimilation techniques provide the opportunity to mitigate the impact of active ionospheric conditions by providing delay corrections in near real time. Alternative techniques to MIDAS include IDA, developed by Bust et al. (2000, 2004), GAIM, developed by Schunk et al. (2001) at Utah State University, and GAIM, developed by the Jet Propulsion Laboratory and the University of Southern California (Mandrake et al., 2005), are both data assimilation methods that contain physics-based models. These have the potential advantage of providing an accurate forecast of the ionosphere. This is a non-trivial task, because the ionosphere is influenced both by Sun and the lower atmosphere. Consequently, a method to provide the correct external forcers to the local ionospheric physical model is also needed. Much of the current ionospheric data is derived from GPS receivers, providing estimates of slant TEC. However, GPS data cannot cover the entire Earth and so there is a need to fill in information across data-sparse regions. The research described in this paper evaluates the performance of IRI in terms of hour-to-hour monthly averages of vertical TEC. The evaluation is done using MIDAS, because ionospheric imaging provides a consistent framework to produce TEC without the distortions created by mapping functions (Meggs et al., 2005).

MIDAS uses phase data from dual-frequency Global Positioning System (GPS) receivers to measure the slant TEC between satellites and the ground. This data includes a potentially large unknown phase offset, which is assumed to remain constant while signal lock is maintained between satellite and receiver. Mitchell and Spencer (2003) created the first version of MIDAS. It balances a series of spherical harmonics and empirical orthonormal functions derived from Chapman profiles around the world and assumes the ionosphere varies linearly with time. Spencer and Mitchell (2007) developed a second version using a Kalman filter approach. This allowed physical models to be incorporated in the inversion, which meant that small-scale fast-moving structures could be imaged successfully. Physical models help MIDAS cope with the lack of GPS data in the polar regions. The lack of GPS data is a result of the lack of receiver sites. The third version of MIDAS (used in this paper) does not use a Kalman filter. As in the first version, a set of orthonormal profiles derived from Chapman functions are used to constrain the vertical profile, so only a small number of coefficients are required to determine electron densities above each geographic position. A regularisation term is included to minimise the second derivative of electron density in space and time. Also, data from satellite-receiver pairs that provide less than the maximum number of rays during the window are weighted lower, since their biases are less well determined.

1.3 The International Reference Ionosphere

IRI 2007 is an empirical model of the ionosphere based on a wide range of ground and space data, including incoherent scatter radars and topside sounders. It is the result of collaboration between the Committee on Space Research (COSPAR) and the International Union of Radio Science (URSI) that began in 1969. It produces monthly median reconstructions (Blitza, private communication) of electron density, ionized gas composition and temperature in the altitude range 50 – 1500 km (65-2000 km for electron density). The 2007 model has new topside electron density options designed to improve TEC measurements by removing an artefact that overestimated electron densities above 500 km. IRI 2007a subtracts a correction coefficient from the IRI 2001 values above 500 km while IRI 2007b uses a different topside model, also used in NeQuick. This model was developed by Radicella and Letlinger (2001). IRI 2007b produces better results than IRI 2007a according to four tests run by Blitza and Reinisch (2008). It should be noted that IRI 2007a produces better results than IRI 2007b in the equatorial region because NeQuick provides two F-region peaks even when only one is necessary.

2. Method and Results

2.1. - 12 year comparison of MIDAS and IRI 2007

Two twelve year IRI runs were performed, one with the NeQuick topside model switched on (IRI 2007b) and another with it switched off (IRI 2007a). There was minimal difference (around 1 TECU) between the two at any time in terms of a bulk TEC number across the entire grid, therefore the NeQuick (IRI 2007b) runs were used for the following comparisons. The top of the grid was set to the maximum 2000 km to minimise the plasmaspheric difference when comparing with GPS data (GPS satellites are at 20 000 km). IRI aims to produce a monthly average of the ionosphere for magnetically quiet conditions, so thirty day median hourly
MIDAS images were created for a fair comparison with IRI. For example, an IRI image for 1200 UT on the 15th of April was compared with the cell-by-cell median of the thirty MIDAS TEC images at 12:00 UT in April. The mean TEC from the cells in this median image was compared with the mean TEC in the IRI image.

Regions of good GPS data coverage were identified in North America and Europe between 1998 and 2010. At least 40 sites were available in each of these areas for most of that period. This allowed the observation of the ionosphere over a full solar cycle. GPS receiver site lists for North America and Europe were updated every three years to deal with new sites being installed and old sites being switched off. This means the spatial distribution of the data improved over the 12 years. The exact number of sites in use varied due to temporary outages. The impact of this was difficult to assess, because closely grouped sites were less useful to the reconstructions than ones that were more spread out. Reconstructions were performed hourly to capture the major diurnal variations. GPS data was supplied by the International GNSS Service (IGS). See Dow et al. (2009) for more details.

Fig. 1 shows close agreement between MIDAS and IRI over most of the 12-year period in Europe and North America. Day-side disagreement peaked in spring 2000 and particularly spring 2002, with MIDAS values significantly higher. Night-side plots showed no such peaks in disagreement, but did show slightly higher TEC for much of the 12 year period. RMS differences were low most of the time, showing that TEC maps were similar across the grid. Peaks in RMS difference were associated with peaks in overall difference between the two images. Spring 2000 and Spring 2002 peaks in disagreement coincided with peaks in sunspot number. This may be an indication that the F10.7 solar flux indicator used by IRI did not sufficiently characterise solar activity. However, the Autumn 2002 peak in disagreement in North America did not fit this trend. After 2003, MIDAS and IRI TEC generally agreed well in the regional plots.

2.2. Systematic Bias

Large differences between MIDAS and IRI are evident in the daytime regional reconstructions around solar maximum. Combined GPS satellite and receiver bias estimates were extracted from the MIDAS images and compared with estimates from the Centre for Orbit Detection in Europe (CODE). Images from February 2002 were tested (fig. 2) since this was the month where MIDAS showed the greatest difference with IRI. Images from April-September 2004 were also tested (fig. 3) as a control. IRI images could not be directly tested, since IRI aims to provide only a monthly median of conditions.

The technique for deriving bias estimates from MIDAS involved subtracting independent estimates of slant TEC from the MIDAS slant TEC values. The independent data were taken from GPS phase-smoothed pseudorange estimates of TEC. These measurements provide absolute TEC values, but contain significant hardware biases from both the satellites and the receivers (see Choi et al (2011) and Wilson and Mannucci (1993)). The difference between the MIDAS and GPS slant TEC values will be that hardware bias, so long as there is no remaining bias in the MIDAS image.

Singular value decomposition was used to solve for the individual satellite and receiver biases across the network. These estimates were then added back together to form combined satellite-receiver biases. This last step was necessary because there was no way of determining how the biases were shared between satellite and receiver.

The MIDAS bias estimates were compared with estimates from CODE. Hugentobler et al. (2000) state that CODE produce inter-frequency biases for all GPS satellites and around 160 IGS receiver stations that are repeatable to around 0.285 TECU (converted from nanoseconds, as per Gaposchkin and Coster, 1993). This test showed that the MIDAS images for Europe in February 2002 had no systematic bias in TEC (fig. 2). Similar results were observed in the North American reconstructions at this time, although the number of receivers for which CODE bias estimates were available was limited in that region. No systematic bias was observed in April-September 2004 either (fig. 3). There was less random error in these measurements because the ionosphere was less variable in that period. These results indicate that the MIDAS results were not biased.

The bias in IRI TEC was investigated by examining monthly median TEC over the Rome ionosonde station in March 2002. IRI was run using manually scaled ionosonde observations of HmF2 and F0F2 and, separately, using the default coefficients. The default mode, as used in the other tests, underestimated the ionosonde-corrected TEC by 52 TECU. F0F2 inaccuracies were by far the most important source of error in this case. This demonstrates that the large systematic errors observed in IRI can be caused by poor peak
data from the coefficients. Ionosonde observations were provided by the Electronic Space Weather and Upper Atmosphere (eSWua) database of the Istituto Nationale di Geofisica e Vulcanologia (INGV), as described by Romano et al. (2008).

2.3. Diurnal Variation

Fig. 4 shows the mean diurnal TEC variation in Europe over the 12 year period. Both datasets showed the same overall pattern, with MIDAS consistently around 2 TECU higher than IRI. Similar results were observed in North America. These results were to be expected since GPS rays (the input to MIDAS) contain information on the plasmasphere. Recombination times are longer in the plasmasphere due to lower particle density, so MIDAS TEC should decrease slower than IRI after the mid-day peak.

It is clear that the plasmaspheric content in GPS TEC observations cannot be ignored in a data assimilation scheme. Coster et al. (1990) have observed over 30 TECU between 800 km and the GPS satellites (20 000 km) during a period of variable geomagnetic activity near solar maximum. Klobuchar et al. (1978) found 2-6 TECU above 2000 km during solar minimum. Ciraco and Spalla (1997) found an overall value of 3 TECU between 1100 km and 20 000 km by comparing GPS and Navy Navigation Satellite Systems (NNSS) TEC measurements over two years.

2.4. Global reconstructions

Near-global-scale reconstructions were performed for MIDAS after 2005. Prior to this date, there were insufficient data to perform a numerically stable global-scale inversion. Reconstructions were performed hourly, using 100 sites and with 5x10 degree horizontal resolution. Polar caps (>80 degrees North or South) were omitted from the global reconstruction because IRI 2007 does not perform well in the polar region. The MIDAS images used in this test were limited by a lack of GPS data over the oceans. This was mitigated by the use of lower than normal horizontal resolution and a longer time window, but those steps meant the inversion coped less well with the variable conditions seen before 2005. Fig. 5 shows the mean TEC of the IRI and monthly median MIDAS images from January 2005 to December 2009.

Various sources (e.g. Rishbeth and Muller-Wodarg, 2006) report an annual asymmetry in peak density (NmF2) values. A 30% difference in NmF2 is observed between January and July, much more than could be explained by the 7% annual variation in solar flux. It appears there is more plasma in January than in July. Mendillo et al. (2005) show that the asymmetry exists in TEC as well as in peak density. They exclude the possibility that plasma is simply being transported above the peak height - and therefore out of sight of the ionosondes - by comparing TEC from GPS data in December 2001 and June 2002 between 65 degrees north and 65 degrees south. Mendillo obtains an asymmetry index (AI) of 0.15 using the following calculation:

\[ AI = \frac{Dec01 - Jun02}{Dec01 + Jun02} \] (1)

This is a reasonable index to use over periods greater than one solar cycle. However, it is susceptible to changes in solar activity over periods of a few years or less because of solar cycle behaviour. For example, December 2004 should have less global TEC than June 2003 since sunspot numbers decreased over the same period. For consistency with the earlier ionosonde measurements of asymmetry, January and July were used instead of December and June. Each year's asymmetry index was computed from that January's data, and from the mean of the preceding and following July. We used the following calculation to achieve this:

\[ AI = \frac{Jan02 - \text{mean}(Jul01 + Jul02)}{Jan02 + \text{mean}(Jul01 + Jul02)} \] (2)

This method was used by Rishbeth and Muller-Wodarg (2006) to remove the solar cycle behaviour. MIDAS results for TEC asymmetry were quite similar to Rishbeth and Muller-Wodarg's NmF2 asymmetry from pairs of ionosondes. Fig. 6 shows that the MIDAS asymmetry index varied between 0.08 and 0.15 between 2006 and 2009, which is generally somewhat lower than Mendillo et al's (2005) result of 0.15. This is most likely because we removed the solar cycle behaviour from our results. IRI shows similar behaviour between 1999 and 2009, with an index of between 0.08 and 0.14. IRI's asymmetry index, which represents the asymmetric
content as a proportion of total ionization, remains roughly constant over the solar cycle. This means the asymmetric effect has varied at the same rate as the total plasma content.

2.5. Summary

- Regional MIDAS and IRI reconstructions showed close agreement after 2003, with peaks in disagreement between 2000 and 2002 during the daytime at equinox.
- MIDAS showed no systematic bias when compared to independent GPS data, showing that IRI underestimated TEC at periods of high solar activity.
- Diurnal TEC variations show that MIDAS had higher TEC than IRI, probably due to a contribution from the plasmasphere.
- Global MIDAS images showed close agreement with IRI after 2005, both showing positive asymmetry indices for all years studied.

3. Discussion

A twelve year comparison of IRI 2007 with monthly median GPS derived tomographic images of the ionosphere showed that the empirical model closely matched TEC over Europe and North America for most of that period. This demonstrated that monthly median ionospheric TEC at mid-latitudes can be specified by the limited range of global geomagnetic and solar parameters used by IRI 2007. Therefore it is feasible that GPS data could be combined with IRI 2007 to define a global ionospheric analysis with variable structures in regions of good data coverage. It would be beneficial for such an analysis to be used to constrain the ionospheric element of a coupled physical model of the upper atmosphere, feeding information to the little-observed thermospheric state by the well-known mechanisms of ion drag and joule heating. This in turn would allow the physical model to produce improved forecasts of both charged and neutral parameters. Forecasting the upper atmosphere is an important challenge, since it will allow us to improve the resilience of Global Navigation Satellite Systems (GNSS), protect power grids, reduce risks to flights over the polar caps and more accurately determine satellite orbits.

The comparison showed that IRI 2007 significantly underestimated TEC in February-March 2000 and 2002, compared to MIDAS images validated with independent GPS data. This was most likely caused by URSI/CCIR coefficients underestimating HmF2 and NmF2 at these times, since the use of ionosonde-derived HmF2 and NmF2 increased IRI TEC over Rome by 52 TECU in March 2002. The authors look forward to the development of a data assimilation scheme for IRI that could incorporate ionosonde data into the three-dimensional electron density images. In the meantime, the assessment of IRI peak errors can be used as an indication of errors in TEC.

IRI’s underestimation of TEC by about 2 TECU over the whole 12-year period may be explained by topside plasma, but it has not been possible to determine whether this is above or below the 2000 km top of the model. The sources referenced in Section 2.3 show that there could be at least 2 TECU above 2000 km, the top of IRI. The alternative scenario is that the extra plasma is contained within the limits of IRI and that therefore the model is in error. If that is the case, the bias is likely to be in the topside, since IRI is largely based on observations of the bottomside and peak. This is supported by Lee and Reinisch (2006), who found that IRI 2001, an earlier version of the model, had no significant NmF2 biases when compared to digisondes at Jicamarca. In addition, Jakowski and Mayer (2009) found that IRI 2007 with NeQuick underestimated topside electron densities at low and mid latitudes from 400 km up to 2000 km. The data from this study came from assimilative reconstructions of CHAMP satellite GPS data between 2002 and 2006, which also indicated plasmaspheric TEC. In order to develop the IRI project, it would be beneficial for work to be done on improving our understanding of the topside plasma distribution, perhaps using TEC measurements from a satellite at an altitude of 2000 km or thereabouts. If significant quantities of plasma are found above that height, the model top will need to be extended, whereas a finding of only negligible densities would indicate a bias in underestimation of densities below that height.

IRI was often seen to be biased in the equinoctial months, both in the regional and the global reconstructions. This bias was not always in underestimation - figure 6 shows an overestimation around September 2007 followed by an underestimation leading up to March 2008. Global RMS errors also consistently increased around the equinoxes. It is possible that IRI failed to represent the more active conditions in this period, since the model was primarily designed for use in geomagnetically quiet conditions (Bilitza and Reinisch, 2008). Kunches and Klobuchar (2001) observed more magnetic storms during the...
equinoctial months because of the relative orientation of Earth's magnetic field within the interplanetary magnetic field. IRI's error may also be due to an inability to represent some of the processes underlying the semi-annual asymmetry. This phenomenon, observed by Pham Thi Thu et al. (2011) among others, is distinct from the annual asymmetry referred to earlier and instead describes enhanced ionization in the equinoctial months. If the semi-annual asymmetry is caused by composition changes in the neutral atmosphere, then it could be the cause of the biases in IRI at these times. This is because the topside model, NeQuick, contains no information on neutral atmospheric composition, as explained by Coisson et al. (2005). This explanation requires that the neutral atmospheric composition plays an important role in the topside ionosphere, specifically around the equinoxes. Possible causes, then, include Rishbeth and Setty's (1961) explanation, a change in the O/N2 ratio, and Zou et al.'s (2000) suggestion that the asymmetry could be due to tides propagating up from the lower thermosphere. These are not necessarily the main causes of the semi-annual asymmetry, but their variability could be the reason for the errors we observed in IRI. Jee et al. (2005) observe that IRI 2001, an earlier version of the model, had much less semi-annual asymmetry than TOPEX ocean altimetry TEC data during periods of high solar activity. We suggest that a consideration of thermospheric parameters may help IRI cope with the ionospheric variability caused by these processes.

Acknowledgements

We thank all those at the Invert group in Bath for useful discussions and comments. Thanks also to Dieter Bilitza for his advice on IRI 2007. This work was supported by the Systems Centre and the EPSRC funded Industrial Doctorate Centre in Systems (Grant EP/G037353/1) and the Met Office. Cathryn Mitchell acknowledges support from a Royal Society Wolfson award.

References


Figure Captions

{WEB}
Fig. 1 shows the mean TEC of IRI (green) and 30-day median MIDAS (blue) images between 1998 and 2010 over Europe and North America. Day-time images are based around local noon, whilst night-time images are based around local midnight. The red line shows the RMS difference between the images. The dash-dotted line in the day-time images shows the smoothed sunspot number reduced by a factor of 1.5.

{PRINT}
Fig. 1 shows the mean TEC of IRI (dashed line) and 30-day median MIDAS (solid line) images between 1998 and 2010 over Europe and North America. Day-time images are based around local noon, whilst night-time images are based around local midnight. The dotted line shows the RMS difference between the images. The dash-dotted line in the day-time images shows the smoothed sunspot number reduced by a factor of 1.5.

Fig. 2 compares MIDAS and CODE estimates of combined satellite-receiver biases in February 2002, a period of intense geomagnetic activity.

Fig. 3 compares MIDAS and CODE estimates of combined satellite-receiver biases in April-September 2004, a period of moderate geomagnetic activity.

Fig. 4 shows the mean diurnal variation of TEC, based on European data from 1998 – 2010. The dashed line indicates IRI, whilst the solid line indicates MIDAS.

Fig. 5 shows the mean TEC of IRI (dashed line) and 30-day median MIDAS (solid line) images between 2005 and 2010 over all longitudes and up to 80° North and South. The dotted line shows the RMS difference between the images.

Fig. 6 shows the annual asymmetry index (AI) taken from global IRI and 30-day median MIDAS images in January and July. The index gives an indication of the imbalance in TEC over this period.
Average Total Electron Content (TECu)
Fig. 2