Global TEC maps based on GNNS data: 2. Model evaluation

P. Mukhtarov, D. Pancheva, B. Andonov, and L. Pashova

Received 17 January 2013; revised 18 April 2013; accepted 19 June 2013.

The present paper presents a detailed statistical evaluation of the global empirical background TEC model built by using the CODE TEC data for full 13 years, 1999-2011, and described in Part 1. It has been found that the empirical probability density distribution resembles more the Laplace than the Gaussian distribution. A further insight into the nature and sources of the model's error variable led up to building of a new error model. It has been constructed by using a similar approach to that of the background TEC model. The spatial-temporal variability of the RMSE (root mean squares error) is presented as a multiplication of three separable functions which describe solar cycle, seasonal and LT dependences. The error model contains 486 constants that have been determined by least squares fitting techniques. The overall standard deviation of the predicted RMSE with respect to the empirical one is 0.7 TECU. The error model could offer a prediction approach on the basis of which the RMSE depending on the solar activity, season and LT is predicted.


1. Introduction

Over the past three decades, the development of empirical total electron content (TEC) models, as well as other parameters like the critical frequencies of the ionosphere, the ion composition or electron and ion temperatures, has increasingly become a major focus of the space physics community. Ionospheric models of the critical frequencies have been playing an important role in specifying the ionospheric environment through which the radio waves propagate, i.e. they are very important for the HF radio communications. The effects that can be recognized as a consequence of bad knowledge for the ionosphere are: loss of communications, change in area of coverage, low signal power, fading. The TEC models are particularly important not only for scientific research on ionosphere but also for applications, such as error correction to operational systems, satellite navigation and orbit determination, satellite altimetry, determining the scintillation of radio wave, etc. Developing of forecast products and further improvement of the newcast products on the base of empirical TEC models has become an important space weather service for addressing the ionospheric space-weather effects on the global navigation satellite systems (GNSS) applications.

Early empirical models of TEC were constructed on the basis of TEC measured by Faraday rotation technique at a single site [Poulter and Hargreaves, 1981; Baruah et al., 1993; Jain et al., 1996; Gulyaeva, 1999; Unnikrishnan et al., 2002; Chen et al., 2002]. Later, regional TEC models over Europe using observations of differential Doppler measurements were built during two previous successful Actions: COST 238-PRIME (Prediction and Retrospective Ionospheric Modelling over Europe) and COST 251-IITS (Improved Quality of Service in Ionospheric Telecommunication Systems Planning and Operation) [Bradley, 1999; Hanbaba, 1999]. However, early regional empirical TEC models are limited by the sparse distributions of observational sites and the estimated vertical TECs are affected by a greater uncertainty over the place without observations [Mao et al., 2008]. Recently, the Global Positioning System (GPS) measurements obtained from the global and regional networks of International GNSS Service (IGS) ground receivers have become a major source of TEC data over large geographic areas [Wilson et al., 1995; Komjathy, 1997; Mannucci et al., 1998; Hernández-Pajares et al., 1999; Orús et al., 2005; Jakowski et al., 2011]. Using TEC data from hundreds of GNSS stations worldwide, global ionosphere maps (GIMs) were developed to produce instantaneous snapshots of the global ionospheric TEC. They have been used for monitoring global ionosphere as a key component of the space weather and for establishing of data-driven models.

A global empirical background TEC model based on the Center for Orbit Determination of Europe (CODE) TEC data [Schaer, 1999] for full 13 years, January 1999 – December 2011, was presented in the Part 1. It describes the climatological behaviour of the ionosphere under both its primary external driver, i.e. the direct photo-ionization by incident solar radiation, and regular wave particularly tidal forcing from the lower atmosphere. The spatial-temporal variability of the modeled TEC was presented as a multiplication of three separable functions which describe solar, seasonal and diurnal variabilities because of their very different time scales (at least an order of magnitude). The model offers TEC maps which depend on geographic coordinates (5°x5° in latitude and longitude) and UT.
at given solar activity and day of the year. One important aspect of the model development process is the evaluation of model performance comprehensively and objectively and this is done in the present Part 2. This means that we have to represent an objective and meaningful description of the model’s ability to reproduce reliable observations precisely or accurately, i.e. to determine the extent to which model-predicted events approach a corresponding set of reliable observations. We study in detail the solar, seasonal and diurnal variability of the error in order to gain even further insight into the nature and sources of the model's error variable and on the base of the obtained results we present an error model as well.

2. General Evaluation of the Background TEC Model

[5] In the Part 1 we have already presented the overall statistical assessment of the model based on the entire data set. The model performance has been represented by the mean (systematic) error ($ME$), root mean squares error ($RMSE$) and the standard deviation error ($STDE$) calculated with the expressions (4) in the Part 1. It has been found that the background model fits to the CODE TEC input data with a zero systematic error and a $RMSE = STDE = 3.387$ TECU. The empirical probability density distribution of the model's error is shown in Figure 1a (black line). It is almost a symmetric function and bears a resemblance in some way to the Laplace distribution, shown in Figure 1a by red line (calculated at the same mathematical expectation and variance as the empirical one), but has also significant differences particularly around errors close to zero. The confidence limits of the error at a given probability are determined empirically by numerical integration of the probability density function shown by black line in Figure 1a. Figure 1b shows the probability for obtaining a given error expressed in times $STDE$ (i.e. the empirical error function). Figure 1b reveals that the 90% probability corresponds to an error interval from $-1.5 STDE$ to $1.5 STDE$, i.e. from about -5 to 5 TECU. This means that there is a 90% probability that deviations larger than 5 TECU between the model and the CODE TEC data would not occur.

[6] It is worth noting that the above found result that the empirical probability function resembles more the Laplace distribution than the well known normal or Gaussian distribution is not a strange result when the probability statistics on ionospheric quantities is considered. For example, Bradley et al. [1999] by investigating the probability statistics of the relative deviations of $f_{o}F2$ from a reference level over the European region found that the probability density distribution fits better to Laplace distribution, which has a more ‘spike’ nature, than to the Gaussian distribution. Wernik and Grzesiek [2011] reported that in situ measurements indicate that the probability distribution function of plasma density fluctuations on scales of importance to scintillation is far from the Gaussian and resemble the Laplace (double exponential) distribution. The authors revealed also that the mean delay time is much more irregular for the Laplace statistics and may be much larger than that for the Gaussian distribution.

[7] The overall statistics of the model error can be defined more precisely by showing its dependence on LT and modip latitude. Figure 2 shows the mean (systematic) error ($ME$) dependence on modip latitude and LT. It is seen that it reaches the largest values of ±0.7 TECU (insignificant error) mainly at low- and low-middle latitudes. The $ME$ variability reflects the
fact that the fifth harmonics of the solar day (4.8-hour tidal component) is not included in the background TEC model.

[8] Figure 3 shows the RMSE distribution (upper plot; the contour distance is 0.5 TECU) with respect to the modip latitude and LT. The largest errors are obtained around sunrise (~8 LT) and sunset (~18 LT). While the sunrise errors maximise above the equator (and this is normal because of the absence of equatorial ionization anomaly (EIA)) those at sunset maximise not only above the equator but also at ±30° modip latitude, i.e. at the EIA crests. The errors at the northern EIA crest are slightly larger than those at the southern crest. This result could be connected with the asymmetric behaviour of the migrating diurnal (DW1) and semidiurnal (SW2) tidal components seen at the left column of plots in Figure 2 of Part 1. Both tidal components are stronger in the Northern Hemisphere (NH) however this asymmetry for the DW1 is better expressed at high solar activity while for the SW2 is well seen at low solar activity. This asymmetric tidal behaviour is not well described by the background TEC model. The bottom left plot of Figure 3 presents the relative RMSE distribution (the contour distance is 0.025) with respect to the modip latitude and LT which is calculated by using the following expression:

$$\text{RMSE}_{rel} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\text{TEC}_{\text{mod}} - \text{TEC}_{\text{obs}})^2}$$

(1)

[9] The modip latitude-LT cross section of the mean observed TEC, used for calculating the relative RMSE, on modip latitude and LT is shown in the bottom right plot (the contour distance is 2 TECU).

Figure 3. Dependence of RMSE (upper plot; the contour distance is 0.5 TECU) and relative RMSE (bottom left plot, the contour distance is 0.025) on modip latitude and LT; the dependence of the mean observed TEC, used for calculating the relative RMSE, on modip latitude and LT is shown in the bottom right plot (the contour distance is 2 TECU).
the equatorial evening pre-reversal electric field (F-region vertical drift) and their effect on the variability of the plasma irregularities [Abdu et al., 2006; Takahashi et al., 2006, 2007].

3. Basic Approach of the Error Model Construction

[10] In order to assess the dependence of the error on the solar activity, seasons and LT we have to demonstrate how the model’s error changes at different conditions. For this purpose we calculated the monthly mean values of the RMSE for the considered period of time, 1 January 1999 – 31 December 2011. The left column of plots in Figure 4 shows the modip latitude-time cross sections of the monthly mean RMSE at different LT: 00LT (most upper plot), 08LT (second from above plot), 12LT (third from above plot) and 18LT (bottom plot). It is clearly evident that the model’s errors are larger during high solar activity (i.e. solar cycle dependence) at equinoxes (i.e. have seasonal dependence), and they depend on LT. All the above mentioned dependences are very similar to those of the background TEC itself. The time scales of the error variability related to the solar cycle, season and LT are very different, as it was the case with the background TEC model. Therefore for building the error model we use the same approach as that applied in constructing the background TEC model and the RMSE can be represented as:

\[ \text{RMSE}_{F,K_F,\text{month},LT} = \Psi_1(F,K_F) \Psi_2(\text{month}) \Psi_3(LT) \]  

[11] The above right hand side unknown functions \( \Psi_k \) (\( k = 1,2,3 \)) can be represented by their series expansions as it was done in Part 1; \( \Psi_f \) can be expanded in Taylor series, while \( \Psi_2 \) and \( \Psi_3 \), which are periodic functions with periods respectively a year (12 months) and a solar day (24 hours), can be expanded in Fourier series. Then at each modip
latitude the error model can be described by the following functions:

\[
RMSE(F, K_F, \text{month}, LT) = \left( a_0 + a_1 F + a_2 K_F + a_3 F^2 + a_4 F K_F + a_5 K_F^2 \right) \times \left( b_0 + \sum_{j=1}^{4} b_j \cos\left( \frac{2\pi}{12} \text{month} - \theta_j \right) \right) \times \left( c_0 + \sum_{j=1}^{4} c_j \cos\left( \frac{2\pi}{24} LT - \lambda_j \right) \right)
\]  

(3)

[12] The expression in the first right hand bracket, i.e. the Taylor series expansion up to degree of 2, represents the solar activity term which modulates the seasonal and diurnal behavior of the RMSE. Similarly to the Part 1, \( F \) is the solar radio flux at 10.7 cm wavelength (F10.7) and \( K_F \) describes the linear rate of change of F10.7. The seasonal term (expression in the second right hand bracket) includes the yearly mean \((b_0)\) and 4 subharmonics of the year, i.e. annual, semiannual, 4- and 3-month components; it modulates the diurnal behavior of the RMSE. In this case the diurnal variability of the RMSE model (expression in the third right hand bracket) is composed however only by two terms: daily mean RMSE \((c_0)\) and a term describing the migrating tides. This is due to the fact that the RMSE depends mainly on the LT. The contribution of the migrating tides in (2) includes 4 subharmonics of the solar day, i.e. 24-, 12-, 8- and 6-hour components.

![Figure 5. Latitude-LT cross sections of the real (left column of plots) and model (right column of plots) RMSE for 15 January (upper row of plots), 15 March (second from above row of plots), 15 July (third from above row of plots) and 15 October (bottom row of plots) at high solar activity 2001.](image-url)
The error model described by (3) contains 486 constants (we remind that (3) is applied at each modip latitude) and they are determined by least squares fitting techniques. Similarly to the TEC model the numbers of the included components in the Taylor and Fourier expansion series are defined by the trial and error method. We accepted only the above mentioned solar, seasonal and diurnal components because the addition of more components does not improve significantly (only after the second decimal point) the error of the error model.

The error model offers a prediction approach on the basis of which we can predict the RMSE depending on the solar activity, season and LT. Therefore, the presented companion papers present not only a global empirical background TEC model but also a global prediction of the model's error at different solar, seasonal and LT conditions. The overall standard deviation of the predicted RMSE with respect to the empirical obtained one is 0.7 TECU. The right column of plots in Figure 4 presents the same results as the left column of plots but for the model RMSE, i.e. modip latitude-time (months) cross-sections of the monthly mean RMSE. The detailed comparison between the real and model RMSE reveals some important features. The error model describes very well the real RMSE at 00 and 08 LT; it is able to reproduce not only qualitatively, but also quantitatively solar and seasonal dependences of the RMSE. The error model is able to reproduce even the hemispheric asymmetry of the RMSE well seen particularly at high solar activity; it is larger in the SH at 00 LT and in the NH at 08 LT. The error model performance at 12 and 18 LT is not as good as that at 00 and 08 LT. The model has not been able to reproduce well particularly the large errors seen in 1999, 2001 and 2011. Generally however the error model describes correctly the solar and seasonal dependences of the RMSE and its global distribution.

4. Application of the Error Model

The global empirical background TEC model, described in Part 1, offers TEC maps which depend on geographic coordinates (5°x5° in latitude and longitude) and UT at given solar activity and day of the year. The error model however does not depend on the geographic longitude because only the contribution of the migrating tidal components is considered in the model. In this way the error maps depending on the geographic latitude and LT at given solar activity and month of the year have to be constructed. The conversion of the modip latitude to geographic one is done...
at the Greenwich meridian. The error values assigned to both poles are obtained by interpolation between the known model points at the highest northern and southern latitudes. The interpolation is done simultaneously with converting the results from modip to geographic frame by using the Inverse Distance Method [Schaer, 1999].

In order to demonstrate the ability of the error model to reproduce the spatial-temporal features of the real RMSE at different solar activity and seasons we use the examples given in the Part 1 (Figures 5, 6 and 7). This means that we will compare the real and model RMSE for 2001, 2004 and 2008 as years representing high, middle and low solar activity and months January, March, July and October as typical winter, spring, summer and autumn months.

Figure 5 presents latitude-LT cross sections of the real (left column of plots) and model (right column of plots) RMSE for 15 January (upper row of plots), 15 March (second from above row of plots), 15 July (third from above row of plots) and 15 October (bottom row of plots) at high solar activity year 2001. The latitude-LT distributions of the real and model RMSE in January (upper row of plots) are very similar not only qualitatively but quantitatively as well. As usually the largest errors are seen at both plots around sunrise (~6-8 LT) and sunset (~16-20 LT); large errors are found at both plots also near 20°N mainly during the daytime (~6-20 LT) and above the equator at sunset. It is worth clarifying that the NH EIA crest is situated around 20°N. The degree of similarity between the real and model RMSE in March (second from above row of plots) is also very high; at both plots the errors are symmetrically distributed with respect to latitude of ~10°N (because of difference between modip and geographic latitudes). Again the largest errors at both plots are seen near sunrise (~7 LT) and sunset (~18 LT) but while the maximum real RMSE is 11 TECU that of the model RMSE is slightly weaker, i.e. it is 10 TECU. The comparison between summer (July) real and model RMSE (third from above row of plots) again demonstrates high degree of similarity, qualitatively and quantitatively. In this case the largest errors are seen only in the morning hours (~8-10 LT) and at ~20°N, but not during the sunset. The winter and summer increase of the RMSE during daytime and at ~20°N is a consequence of the hemispheric asymmetry of the diurnal components DW1 and SW2 contribution to the EIA which is not well reproduced by the background TEC model. This feature was seen also in the upper plot of Figure 7. The same as Figure 5 but at middle solar activity 2004.
The qualitative similarity between the real and model RMSE in October is very good however the maximum model errors are smaller than those of the real RMSE, i.e. they are 8.1 TECU and 11 TECU respectively.

Figure 6 shows the same comparison as that in Figure 5 but for low solar activity 2008. In this case both real and model RMSE drastically decrease. The model describes qualitatively very well the latitude-LT distribution of the real RMSE at all months; there is some quantitative difference mainly during the equinoxes. As it has been expected the largest errors are found at equinoxes both in real and model RMSE but the error model underestimates the RMSE; in March the model and real RMSE are respectively 4 and 6 TECU, while for October the difference is smaller and they are 3.2 and 4.5 TECU. Some hemispheric assymetry of both real and model RMSE is seen in winter (January) and summer (July) here as well but it is not predominantly in the NH as it was at high solar activity 2001 (Figure 5). In January both real and model RMSE at sunrise and morning hours are stronger in the NH while at afternoon and sunset hours they are larger in the SH. The opposite asymmetry is seen in July.

Figure 7 presents a comparison between real and model RMSE maps at middle solar activity 2004. It is seen that both the real and model RMSE at all months are between those at high (Figure 5) and low (Figure 6) solar activity, as it is expected in advance. At all months the largest values of the real and model errors are similar but in March they are almost the same, i.e. 6.9 and 6.8 TECU. The largest difference is seen in July when the maximum real RMSE is 4.5 TECU while the model one is 3.6 TECU. During the daytime almost at all months the NH errors are larger than those in the SH; only in January both the real and model RMSE distribution and the real RMSE in March are more hemispheric symmetric. Similarly to high solar activity the increase of RMSE during daytime and at ~20°N is a consequence of the hemispheric assymetry of the diurnal components DW1 and SW2 contribution to the EIA which is not well reproduced by the background TEC model. While the model reproduces comparatively well the night time (~2-4 LT) amplification of the RMSE in March at low modip latitudes it underestimates that observed in January.

We have to note however that the above demonstrated ability of the error model to reproduce the spatial-temporal features of the real RMSE at different solar activity and seasons does not mean a validation of the error model; it is applied to the data that have been used for generating the model. Figures 5, 6 and 7 display actually the quality assessment of the constructing model procedure, i.e. provide a compact means of reproducing the RMSE from 1999 to 2011. The validation of both background and error TEC models, as well as their improvements will be the subject of a future paper.

5. Summary

A detailed statistical evaluation of the global empirical background TEC model, presented in Part 1, is done in Part 2. The model performance has been described by its mean (systematic) error (ME), root mean squares error (RMSE) and the standard deviation error (STDE). It was found that the background model fits to the CODE TEC input data with a zero systematic error and a RMSE = STDE = 3.387 TECU. Based upon this overall error measures we may confidently conclude that this model is able to reproduce accurately the CODE TEC input data. It was found that the empirical probability density distribution (Figure 1a) resembles more the Laplace than the normal, or Gaussian, distribution. This result could be probably related to the non-Gaussian statistics of the ionospheric irregularities [Bradley et al., 1999; Wernik and Grzesiak, 2011]. The empirical error function shown in Figure 1b revealed that there is only 10% probability that deviations larger than 5 TECU between the model and the CODE TEC data would occur. The modip latitude-LT distributions of the model’s error showed predominantly known features, as: (i) the small ME observed mainly at low latitudes reflect the fact that the fifth harmonics of the solar day (4.8-hour tidal component) is not included in the background TEC model (Figure 2); (ii) the RMSE are large at sunrise and sunset time (Figure 3 upper plot), and (iii) the relative RMSE amplifications shown in the bottom left plot of Figure 3 are related to comparatively stable ionospheric anomalies which are present at some local times and latitudes (as broad plasma anomaly after midnight and evening pre-reversal plasma irregularities at equatorial latitudes) and to areas of significant reduction of the mean observed TEC (bottom right plot of Figure 3).

In order to gain further insight into the nature and sources of the model’s error variable we studied in detail the solar, seasonal and diurnal variability (LT) of the model’s error. On the base of the obtained results we built an error model. It could offer a prediction approach on the basis of which the RMSE depending on the solar activity, season and LT is predicted. The error model was constructed by using a similar approach to that of the background TEC model itself. Similarly to the TEC model the time scales of the error variability related to the solar cycle, season and LT are very different as well. Then the spatial-temporal variability of the RMSE was presented as a multiplication of three separable functions (as it is shown in (3)). The solar cycle and seasonal dependences of the RMSE are described in the same way as in the background TEC model. The Taylor series expansion up to degree of 2 represents the solar activity function while the seasonal function includes the contribution of 4 subharmonics of the year, i.e. annual, semiannual, 4- and 3-month components. The RMSE depends mainly on the LT and due to this its diurnal variability is described only by the migrating tides; four subharmonics of a solar day, 24-, 12-, 8- and 6-hour components are included in the error model. It contains 486 constants which have been determined by least squares fitting techniques. The overall standard deviation of the predicted RMSE with respect to the empirical one is 0.7 TECU. The detailed comparisons between real and model RMSE shown in Figures 5, 6 and 7 clearly demonstrate that the error model describes correctly and precisely the spatial-temporal variability of the RMSE.

In conclusion it is important to note that these two companion papers present not only a global empirical background TEC model but also a global prediction of the model’s error at different solar, seasonal and LT conditions. At given solar activity and day of the year the background TEC model offers TEC maps which depend on geographic coordinates (5°x5° in latitude and longitude) and UT. The error model offers a prediction approach on the basis of which the RMSE depending on the solar activity, season and LT can be predicted.
References


