Annales
Geophysicae

Modeling total electron content derived from radio occultation measurements by COSMIC satellites over the African region

Patrick Mungufeni[1,2], Sripathi Samireddipalle[3], Yenca Migoya-Orué[4], and Yong Ha Kim[1]

[1]Department of Astronomy and Space Science, Chungnam National University, Daejeon, South Korea
[2]Physics department, Mbarara University of Science and Technology, P.O. Box 1410 Mbarara, Uganda
[3]Indian Institute of Geomagnetism, New Panvel, India
[4]The Abdus Salam International Centre for Theoretical Physics (ICTP) T/ICT4D, Trieste, Italy

Correspondence: Yong Ha Kim (yhkim@cnu.ac.kr)

Abstract. This study developed a model of total electron content (TEC) over the African region. The TEC data were obtained from radio occultation measurements done by the Constellation Observing System for Meteorology, Ionosphere, and Climate (COSMIC) satellites. Data during geomagnetically quiet time (Kp < 3 and Dst > -20 nT) for the years 2008–2011 and 2013–2017 were binned according to local time, seasons, solar flux level, and geographic longitude and latitude. B splines were fitted to the binned data to obtain model coefficients. The model was validated using actual COSMIC TEC data of the years 2012 and 2018. The validation exercise revealed that approximation of observed TEC data by our model produces a root mean square error of 5.02 TECU (total electron content unit). Moreover, the modeled TEC data correlated highly with the observed TEC data ($r = 0.93$). Due to the extensive input data and the applied modeling technique, we were able to reproduce well-known TEC features such as local time, seasonal, solar activity cycle, and spatial variations over the African region. Further validation of our model using TEC measured by ionosonde stations over South Africa at Hermanus, Grahamstown, and Louisville revealed r values > 0.92 and root mean square error (RMSE) < 5.56 TECU. These validation results imply that our model can estimate TEC fairly well that would be measured by ionosondes over locations which do not have the instrument. Another element of the significance of this study is the fact that it has shown the potential of using basis spline functions for modeling ionospheric parameters such as TEC over the entire African region.

1 Introduction

Among the error sources that affect the positioning in the Global Navigation Satellite System (GNSS) are propagation-medium-related errors. In particular, the ionospheric refraction is the largest contributor of the user-equivalent range error. This type of frequency-dependent error can virtually be eliminated in dual-frequency receivers by differential techniques (Hofmann-Wellenhof et al., 2007). For the case of single-frequency receivers, some GNSS (e.g., Global Positioning System (GPS) and Galileo) broadcast messages include the parameters of an ionospheric model which can be used to compute and correct the ionospheric effects (Guochang, 2007). For instance, the GPS uses the Klobuchar model, which represents the zenith delay as a constant value at night and a half-cosine function during the day (Klobuchar, 1987). In the framework of the European Galileo constellation, NeQuick G based on the NeQuick model has been proposed to be used for single-frequency positioning (see Issue 1.2, September 2016 by the European Commission, titled "European GNSS (Galileo) Open Service – Ionospheric correction algorithm for Galileo single-frequency users"). The NeQuick and its subsequent modifications (NeQuick G and NeQuick 2) are a three-dimensional, time-dependent ionospheric electron density model developed by the Aeronomy and Radio Propagation Laboratory (ARPL) of the Abdus Salam International Center for Theoretical Physics (ICTP) in Trieste, Italy, and the Institute for Geophysics, Astrophysics and Meteorology of the University of Graz, Austria (Nava et al., 2008). In addition to using models to reduce iono-

spheric refraction errors, Space Based Augmentation Systems (SBAS) such as the Wide Area Augmentation System (WAAS), the European Geostationary Navigation Overlay Service (EGNOS), and the GPS-aided Geo Augmented Navigation (GAGAN) are also used (Hofmann-Wellenhof et al., 2007).

For the international standard specification of ionospheric parameters (such as electron density, electron and ion temperatures, and equatorial vertical ion drift), the Committee on Space Research (COSPAR) and the International Union of Radio Science (URSI) recommended the International Reference Ionosphere Model (IRI) (Bilitza, 2001). IRI is an empirical model primarily based on all available experimental data (ground- and space-based) sources. However, theoretical considerations have been used in bridging data gaps and for internal consistency checks (Bilitza, 2001).

The ionospheric total electron content (TEC) is one of the important descriptive physical quantities of the ionosphere (Rama Rao et al., 1997; Ercha et al., 2012). The GNSS measurements obtained from the global and regional networks of International GNSS Service (IGS) ground receivers have become a major source of TEC data. As one of the IGS analysis centers, Center for Orbit Determination in Europe (CODE) provides Global Ionosphere Maps (GIMs) containing vertical TEC data daily using the GNSS data collected from over 200 tracking stations of IGS and other institutions. Several studies have used GIMs from CODE and other IGS analysis centers such as the Jet Propulsion Laboratory (JPL) to construct TEC models (Jakowski et al., 2011a; Mukhtarov et al., 2013; Ercha et al., 2012; Sun et al., 2017). Jakowski et al. (2011a) proposed the Global Neustrelitz TEC Model (NTCM-GL) that describes the average TEC under quiet geomagnetic conditions. The NTCM-GL was developed using GIMs during 1998–2007 provided by CODE. A global background TEC model was also built using CODE GIMs by Mukhtarov et al. (2013). The model describes the climatological behavior of the ionosphere. The GIMs from JPL were used by Ercha et al. (2012) to construct a global ionosphere model using the empirical orthogonal function (EOF) analysis method. The Taiwan Ionosphere Group for Education and Research constructed a global ionosphere model from GNSS and the Constellation Observing System for Meteorology, Ionosphere, and Climate (COSMIC) GPS radio occultation (RO) observations (Sun et al., 2017). The map of all the averaged root mean square error (RMSE) values of CODE GIMs during the years 2010–2012 presented by Najman and Kos (2014) showed high values over low-latitude African regions. This could be due to the poor distribution of IGS tracking stations over Africa and the inability of the spherical harmonics function used in GIMs to describe ionospheric structure over low latitudes.

In addition to the existing GIMs discussed in the previous paragraph, regional TEC maps and models have also been constructed. In comparison with the global models, regional TEC models might have better accuracy over the particu-

lar region for which it was constructed. Opperman (2008) stated that the higher time and spatial resolution imaging achievable with regional models permits the analysis of localized ionospheric structures and dynamics not observable in global models. Examples of studies that developed TEC models over some parts of Africa are the following. A neural network model of GNSS – vertical TEC (GNSS-VTEC) over Nigeria was developed by Okoh et al. (2016) using all available GNSS data from the Nigerian GNSS Permanent Network (NIGNET). An adjusted spherical harmonic-based TEC model was developed by Opperman (2008) using a network of South African dual-frequency GPS receivers. Habarulema et al. (2011) presented the Southern Africa TEC prediction (SATECP) model that was based on the neural network technique. The SATECP generates TEC predictions as a function of input parameters, namely, local time, day of year, solar and magnetic activity levels, and the geographical location. A neural-network-based ionospheric model was developed using GPS-TEC data over the East African sector by Tebabal et al. (2019). Recently, Okoh et al. (2019) used the neural network technique to develop a TEC model over the entire African region. In addition to using TEC obtained by the COSMIC RO technique, they used TEC measured by GPS receivers on the ground.

Due to the lack of a dense network of ground-based GNSS receivers and poor coverage of COSMIC RO data over the African region, the TEC model over the entire African region presented by Okoh et al. (2019) sometimes failed to capture the equatorial ionization anomaly (EIA) over the region. This point has been illustrated with examples in Sects. 2 and 5. In this study, we applied a data binning method to the COSMIC RO TEC data that allowed development of an improved TEC model over the region. Moreover, we demonstrate the potential of the basis spline functions to model TEC over the African region. These basis functions never vanish over limited intervals and add up to 1 at all local times and longitudes (De-Boor, 1978). Moreover, according to Scherliess and Fejer (1999), they are ideally suited to model the equatorial ionosphere, which exhibits smooth and rapid changes during daytime and near sunset, respectively, by proper placement of the mesh of nodes. In Sect. 2, the data and methods of analysis that were used in the study are described. The details of the model proposed in this study are described in Sect. 3. We present a comparison between the observed and modeled TEC in Sect. 4. The model validation and the conclusions are presented in Sects. 5 and 6, respectively.

2 The data and methods

2.1 Data sources

In order to overcome the problem of the lack of a dense network of ground-based GNSS receivers over the African region, this study used TEC data obtained from RO measure-

ments done by the COSMIC satellites. The integrated electron density (integration being done up to the altitudes of the COSMIC satellites) which is being referred to as TEC in this study can be obtained from ionPrf files which are processed at the COSMIC Data Analysis and Archive Centre (CDAAC) (http://cosmic-io.cosmic.ucar.edu/cdaac/index.html, last access: 12 September 2020). The TEC for the individual occultation events was assigned to the geographic coordinates of NmF2 in the same file.

In order to get integrated electron density approximately up to the altitudes of GPS satellites, Okoh et al. (2019) used neural networks to learn the relationship between coincident TEC measurements done by ground-based GPS receivers and COSMIC RO. They showed that the ratio between TEC data from the two sources varies spatially. This observation implies that the neural networks may not learn the relationship between TEC measured by ground-based GPS receivers and COSMIC RO very well over locations which do not have the former data set during the entire study period. As can be seen in Fig. 1 of Okoh et al. (2019), there was a large spatial coverage with no ground-based GPS receivers. Unlike what has been done in Okoh et al. (2019) and Mungufeni et al. (2019), in the current work we used only COSMIC TEC without any adjustments.

In this regard, an analysis of coincident ground-based GNSS TEC and TEC from COSMIC occultation data performed by Mungufeni et al. (2019) reveals that the upper quartile of the differences between the two data sets may reach up to ~ 11 TECU (total electron content unit) over the northern crest of the equatorial ionization anomaly. Over the southern midlatitude region, the differences were low (~ 4 TECU). Since the upper quartiles of the differences can reach up to ~ 11 TECU, the median/mean values in the worst cases might obviously be much lower than this value. This might be the reason for observing most of the well-known ionospheric TEC features over the African region when the COSMIC RO TEC was appropriately binned as in Mungufeni et al. (2019). Therefore, this study used the TEC obtained from COSMIC occultation measurements to develop the TEC model over the African region in order to reproduce these ionospheric features. Such endeavors are important for educational purposes.

During geomagnetic storms, the variations in zonal electric fields and composition of the neutral atmosphere contribute significantly to the occurrence of negative and positive ionospheric storm effects in the low-latitude region (Rishbeth and Garriot, 1969; Buonsanto, 1999; Adewale et al., 2011). Therefore, since the ionosphere changes in a complex manner during geomagnetic storms, we only considered data on quiet days. The quiet geomagnetic days were identified by examining the 3-hourly Kp and disturbance storm time (Dst) indices that were obtained from the World Data Center of Kyoto, Japan (http://swdcwww.kugi.kyoto-u.ac.jp/, last access: 12 September 2020). A day was considered to be quiet if all the eight Kp values in that day were ≤ 3. In addition

Table 1. Distribution of number of days with data.

Year	Number of days with data
2008	219
2009	293
2010	235
2011	174
2012	169
2013	185
2014	164
2015	128
2016	151
2017	154
2018	211

to satisfying this condition, the hourly values of Dst in that day should also have values ≥ -20 nT. The two conditions were applied to ensure that both low- and mid/subauroral-latitude geomagnetic disturbances are detected by Dst and Kp indices, respectively. In future, we intend to use TEC data during geomagnetically disturbed conditions to construct a TEC model during geomagnetically disturbed periods.

2.2 Methods of data analysis

The TEC data during the years 2008–2011 and 2013–2017 were used for developing the TEC model over the African region. Due to the adequate data needed to develop an empirical model, we only reserved the data of the years 2012 and 2018 for validation. The period considered in this study represents data of both low and high solar activity level in sunspot cycles 23 and 24. The data within geographic latitude and longitude ranges of -35–$35°$ and -20–$60°$, respectively, were used to cover the African region. Table 1 presents the number of days per year when there were TEC data over the African region. Since there are many geomagnetically disturbed days in high (2012–2015) and medium (2011 and 2016) solar activity years, the number of days with data is also reduced in such years compared to low solar activity years (2008–2010, 2018).

It would be good to bin the TEC data according to geomagnetic latitudes since many structural and dynamical features of the ionized and neutral upper atmosphere are strongly organized by the geomagnetic field (e.g., Emmert et al., 2010). This may be complicated since geomagnetic latitude lines are not usually straight. For convenience and simplicity, we binned the data based on geographic coordinates. In order to observe small-scale ionospheric structures, small grid resolutions of 3 and 5° in geographic latitude and longitude, respectively, were used to bin the TEC data. These grid resolutions resulted in 24 and 16 latitudinal and longitudinal bins, respectively. Several studies (e.g., Krankowski et al., 2011 and Mengist et al., 2019) that have used COS-

MIC data commonly consider measurements with horizontal smear > 1500 km prone to errors, and they reject such measurements. We established that after applying this restriction, there were ~ 40 RO measurements per day during the year 2013 over our study area (not shown here). Based on the previous discussions, this value is far less than the 9216 (16 longitudinal, 24 latitudinal, and 24 local time) TEC data points required in all grid cells in a day. As stated in Sect. 1, this poor amount of data to represent day-of-year TEC variation might be the reason for the failure of the TEC model presented by Okoh et al. (2019) to capture in some cases the EIA over the African region. Another reason might be the discrepancy which arises due to some locations being represented by adjusted COSMIC RO TEC while others by the ground-based GPS TEC data.

Since empirical modeling requires adequate data for the mathematical functions to capture the physics inherent in the data, this study did not reject COSMIC RO TEC measurements with horizontal smear > 1500 km. Although not presented here, we observed that the COSMIC TEC data values with smear > 1500 km did not introduce alarming errors. This observation was made when we analyzed COSMIC TEC data which were coincident with TEC observed by ionosonde stations over South Africa (see details in Sect. 5.2), located at Hermanus, Grahamstown, and Louisvale. Interestingly, compared to measurements with horizontal smear > 1500 km, some measurements with horizontal smear < 1500 km were observed to be far from the linear least-squares fitting line. Further analysis of COSMIC RO observations over our study area revealed that without restricting horizontal smear, there were ~ 80 RO measurements per day during the year 2013 (not shown here). Still this value is far less than the 9216 TEC data values required to fill all spatial grid cells in a day. To partially solve this problem, instead of binning data according to year, we binned the data according to different solar flux levels as shown below.

For each spatial grid cell, the data were binned at a 1 h interval. TEC values within the bins were averaged to yield 1 h resolution TEC data over the grids. TEC data for the different days were binned according to the F10.7 flux of that day. The F10.7 flux indices were obtained from the Space Weather Prediction Center (SWPC) of the National Oceanic and Space Administration (NOAA) (http://www.swpc.noaa.gov/, last access: 12 September 2020). The F10.7 flux ranges for low solar activity (LSA), medium solar activity (MSA), and high solar activity (HSA) were < 76, 76–108, and > 108 sfu, respectively. The boundary values 76 and 108 sfu of the F10.7 flux ranges correspond to the 75th and 25th percentiles of all F10.7 flux values on the days in low (2008–2010, 2017–2018) and high (2012–2015) solar activity years, respectively.

The data within a specific solar flux bin were further binned based on months of a year. The averages of the corresponding F10.7 flux of the days used to represent seasonal

Table 2. Average monthly F10.7 flux values used in the study.

Month	F10.7 flux (sfu)		
	LSA	MSA	HSA
January	71.10	83.94	140.65
February	71.14	87.06	126.23
March	69.81	85.40	130.98
April	71.02	86.09	130.46
May	70.29	90.59	123.80
June	69.51	89.91	118.73
July	68.09	88.14	128.92
August	67.45	85.46	114.53
September	69.20	86.34	122.98
October	70.06	81.88	131.50
November	71.66	82.40	142.95
December	70.82	82.97	142.72

TEC were determined and used to capture the variation of TEC with solar flux. Table 2 presents the average F10.7 flux values that were determined in the months of a year. In summary, a total of 331 776 TEC data values were needed to exist in 16 longitudinal, 24 latitudinal, 3 solar flux, 12 monthly, and 24 hourly bins, in order to determine the model coefficients. However, from the data of the entire study period, only 121 447 bins were filled with TEC data values. The average of the standard deviations of the bins that contained more than 1 TEC data during low (sample size $= 21\,108$), medium (sample size $= 6180$), and high (sample size $= 7495$) solar flux levels were 1.28, 2.15, and 4.31 TECU, respectively.

The bins which did not have TEC data were filled by estimation following the procedures described in the three steps below.

1. At a particular spatial grid cell, the diurnal TEC was divided into two local time sectors, namely, (i) 10:00–24:00 LT and (ii) 00:00–10:00 LT. Sector (i) which is daytime and before midnight includes the time when daily and secondary TEC peaks are expected, while (ii) which is mostly at night is when TEC varies slowly. When slow variation of TEC was expected as in sector (ii) and there were at least a few (> 2) TEC data available, a smoothing spline (De-Boor, 1978) data fitting method was used to estimate missing TEC values. In cases in which rapid TEC variations are expected as in sector (i) and at least half of the total expected number of data points were filled with TEC data, a piece-wise cubic interpolation (De-Boor, 1978) data fitting method was used to estimate missing TEC values. For example, when there were at least four measurements in sector (ii), the missing values were obtained by evaluating the fitted function through the existing TEC data values. On the other hand, when there were at least seven (half the number of hours during 10:00–24:00 LT) TEC values in sector (i), the missing values were obtained by evaluat-

ing the fitted function to the available data values. After estimating the missing TEC data from the two sections of the diurnal TEC, the entire diurnal TEC data over a particular grid cell were then considered to estimate the missing values. When there were at least 12 (half the number of hours in a day) values, the missing values were obtained by evaluating a smoothing spline function fitted to the existing data values.

2. At a particular latitude and local time, the values of TEC along all the longitudes were divided into western (− 20–20° E) and eastern (20–60° E) longitude sectors. Each of the longitude sectors contained eight bins. At night, when there were at least three TEC values over any longitude sector, the missing values were obtained by evaluating a smoothing spline function fitted to the available data points, while during the day, when there were at least four TEC values, the missing values were obtained by evaluating a smoothing spline function fitted to the available data points. After estimating the missing TEC values over the two longitude sectors, the TEC over all longitudes was then considered to estimate the missing values. At night, when there were at least eight values, the remaining values were obtained by evaluating a smoothing spline fitted to the available TEC data points. The missing values during daytime were estimated when there were at least 10 measurements available.

3. Procedure 3 is similar to 2, except for variations of TEC as a function of latitude being considered at specific values of longitude and time. TEC values over the latitudes were divided into lower (− 35–0° S) and upper (0–35° N) latitudinal sectors. There were 12 bins in each of the latitudinal sectors. To estimate missing TEC values at night over a latitudinal sector, at least four measurements were required to be available, while during the day, at least six values were required. When TEC data over the combined latitudinal sectors were considered to estimate the missing values, at least 12 values were required to be available.

After repeating procedures 1–3 three times, all the 331 776 bins were filled with TEC data. For the purposes of minimizing the effects of outliers, the diurnal TEC at spatial grid cells was then separately fitted with smoothing splines which were evaluated to obtain the TEC data that were later used to determine the model coefficients as explained in Sect. 3. In order to demonstrate the appropriateness of our estimation of missing TEC data values and its use for determining model coefficients, we present Fig. 1. Panels (a)–(c) of the figure present the available TEC data (*) and estimated (red line) TEC values during low, medium, and high solar flux levels, respectively.

The TEC data plotted in Fig. 1 correspond to January and the grid cell centered at longitude 17.5° W and latitude

34.5° S. Figure 1 clearly shows that the available and estimated TEC variations depict the well-known diurnal and solar activity level dependence patterns. Moreover, the figure shows that the available data values are in most cases close to the estimated TEC values. Therefore, the estimated TEC values were then used to obtain the model coefficients.

3 The model

The TEC over the African region was expressed as

$$\text{TEC}(t, d, F, \lambda, \varphi) = \sum_{i=1}^{24} \sum_{j=1}^{12} \sum_{k=1}^{3} \sum_{l=1}^{16} \sum_{m=1}^{24} a_{ijklm}$$

$$\times N_i(t) \times N_j(d) \times N_k(F) \times N_l(\lambda) \times N_m(\varphi) \ , \qquad (1)$$

where the linear model coefficients a_{ijklm} were determined by the least-squares fitting procedure to the 331 776 TEC data values as in Abdu et al. (2003), Jakowski et al. (2011b), and Mungufeni et al. (2015). In Eq. (1), $N_i(t)$, $N_j(d)$, $N_k(F)$, $N_l(\lambda)$, and $N_m(\varphi)$ are B splines of different orders to represent variations of TEC with local time, seasons, solar flux level, longitude, and latitude, respectively. Most of the B splines were of order 2, except for those used to represent LT and latitudinal variations which were of order 4. The order of splines used to represent LT and latitude was higher to cater for the rapid variations of TEC with these two parameters. A total of 24 local time nodes, 1, 2, ..., 24, were used. For simple interpolation between months, seasonal/monthly nodes were placed at the 15th day of each month. Solar flux nodes used in the various months are as shown in Table 2. The longitudinal nodes were separated by 5° and placed at longitudes − 17.5, 12.5 7.5, ..., 57.5°, while the latitudinal nodes were separated by 3° and placed at latitudes − 34.5, − 31.5, − 28.5, ..., 34.5°.

4 Comparison of observed and modeled TEC

In order to assess the ability of the model to describe the data used to construct it, modeled data were compared with the binned data that were used to solve Eq. (1). The results of the self-consistency check are presented in Fig. 2. It is important to note that validation using data that were not included during modeling is provided in Sect. 5. Panels in column (i) of Fig. 2 present the observed binned TEC data, while column (ii) presents the corresponding modeled TEC data. In column (iii), we present the differences between the observed and modeled TEC data, referred to as errors. In Fig. 2, rows (a), (b), and (c) correspond to LSA, MSA, and HSA, respectively. The horizontal magenta lines in Figure 2 and later also in Fig. 3 indicate the location of ∼ 0° dip latitude on the corresponding panel. As expected, Figure 2 clearly shows that the corresponding modeled TEC almost perfectly matches the observed binned TEC. This can be confirmed by the small
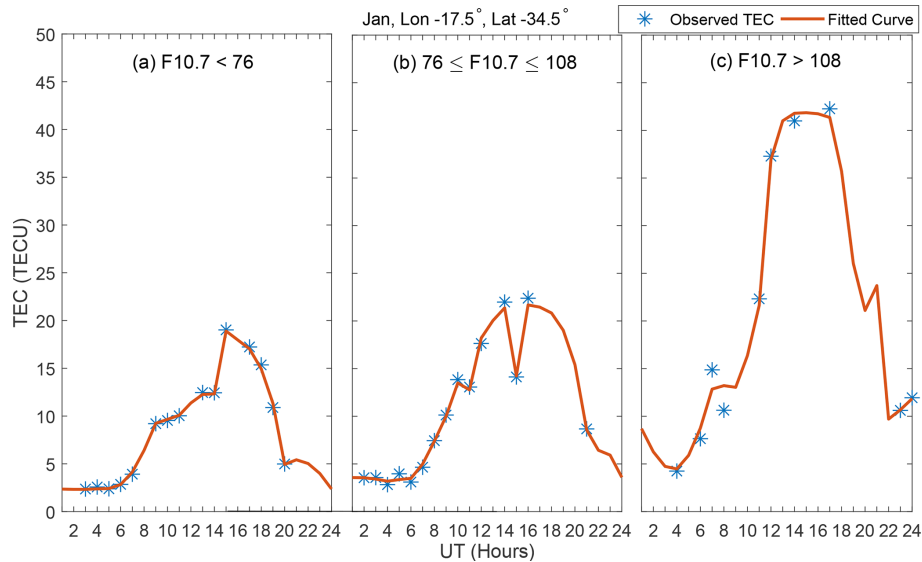
Figure 1. Panels **(a–c)** present available (*) and estimated (red line) TEC values during low, medium and high solar flux levels, respectively. The data are for the month of January and fall within the grid cell centered at longitude and latitude of 17.5° W and 34.5° S, respectively.

(< 0.1 TECU) error values presented in the panels of column (iii). The variations of the ionosphere with local time and solar flux level as well as location that are exhibited in Fig. 2 give the confidence of relying on the binned data as a good representation of the ionosphere. The physical explanations for these variations are as follows. The increase of both observed and modeled TEC that occurs when the solar flux level increases is usually attributed to increased ionizing radiations in X-ray and extreme ultraviolet (EUV) bands, which in turn leads to increased TEC in the ionosphere (Hargreaves, 1992).

The diurnal variation of TEC matches very well with the variation of photoionizing radiation. At sunrise, the electron density begins to increase rapidly, owing to photoionization (Schunk and Nagy, 2009). After this initial increase at sunrise, electron density displays a slow rise throughout the day, and then it decays at sunset as the photoionization source disappears. Another diurnal feature of variation of TEC exhibited in Fig. 2 is the existence of a secondary maximum of TEC. This can clearly be seen in panels of row (c) along the magenta lines, where the first peak occurs at $\sim 15{:}00$ LT and the second at $\sim 18{:}00$ LT. The formation of a secondary maximum of TEC that was mentioned previously may be explained as follows. During the day, the thermospheric wind generates a dynamo electric field in the lower ionosphere that is eastward (Schunk and Nagy, 2009). The eastward electric field, E, in combination with the northward geomagnetic field, B, produces an upward E × B drift of the F region plasma. As the ionosphere co-rotates with the Earth toward dusk, the zonal (eastward) component of the neutral wind increases. The increased eastward wind component, in combination with the sharp day–night conductivity gradient across the terminator, leads to the pre-reversal enhancement

in the eastward electric field (Batista et al., 1986; Schunk and Nagy, 2009). The F layer therefore rises as the ionosphere co-rotates into darkness. Although in the absence of sunlight after sunset, the lower ionosphere rapidly decays, there is high electron density at high altitudes, yielding the secondary maximum in TEC.

Panels in rows (b) and (c) of Fig. 2 demonstrate the existence of the EIA region, where there are two belts of high electron density on both sides of 0° dip latitude. The EIA is usually attributed to the upward E × B drift which lifts plasma to higher altitudes. The plasma then diffuses north and south along magnetic field lines. Due to gravity and pressure gradient forces, there is also a downward diffusion of plasma. The net effect is the formation of the EIA region (Appleton, 1946). Another feature of EIA that can be seen in panels in rows (b) and (c) of Fig. 2 is the asymmetry of the crests. Along the 120° longitude sector, Zhang et al. (2009) reported the asymmetry of EIA crests. As described later at the end of this section, the direction of neutral meridional winds in March may favor high values of electron density over the southern crest.

Generally, Fig. 2 shows that the locations outside the EIA region have lower TEC values compared to locations around and within the EIA region. The low values of TEC over locations outside the EIA region might be due to the lower elevation angle of the solar radiation flux which is responsible for the creation of electrons (Schunk and Nagy, 2009). The solar radiation flux is usually low for locations far from the subsolar point. The latter situation is dominant over locations outside the EIA region, especially in March. The closeness of the subsolar point to the locations within the EIA regions results in high solar radiations over these locations. As a re-
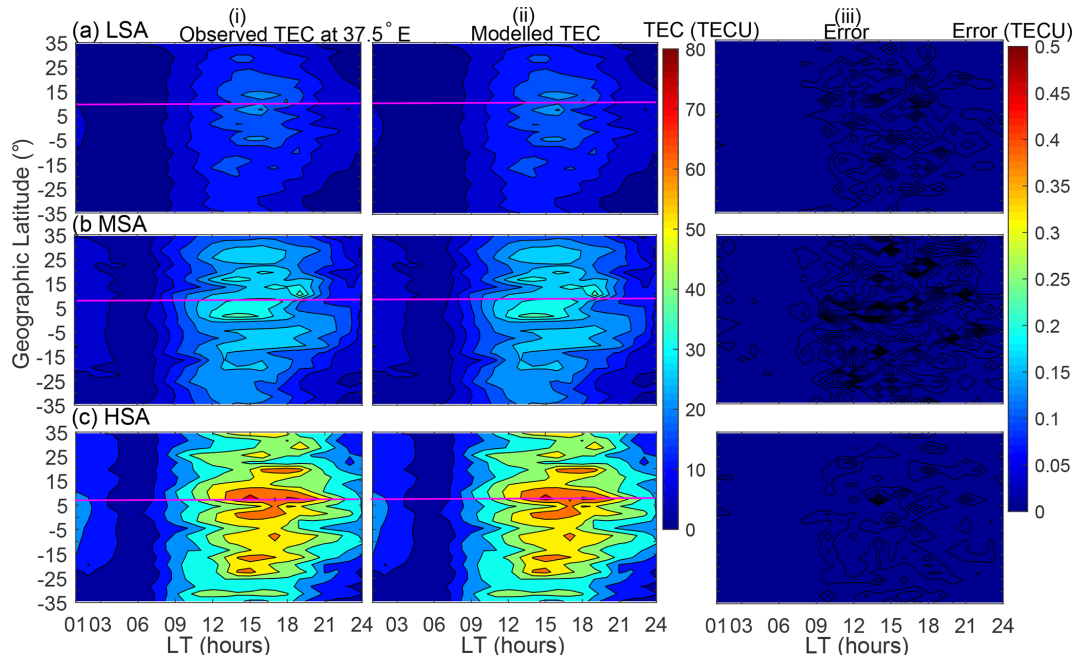
Figure 2. Variation of TEC as a function of geographic latitude and local time in the March equinox at 37.5° E. Panels in rows **(a–c)** correspond to LSA, MSA, and HSA, respectively, while panels in columns (i)–(iii) correspond to observed binned TEC, modeled TEC, and the difference between observed and modeled TEC (errors), respectively. The magenta line indicates $\sim 0°$ dip latitude.

sult, high TEC values were observed over locations within the EIA region.

To demonstrate that the modeled TEC captures TEC variation with season, we present Fig. 3. In the figure, columns (i) and (ii) present observed binned TEC and the corresponding modeled TEC, respectively. Moreover, rows **(a–d)** present TEC data during March, June, September, and December, respectively.

As already observed in Fig. 2, it can clearly be seen from Fig. 3 that the modeled TEC almost perfectly matches the observed TEC data. Among the many features of TEC exhibited by both observed and modeled TEC data, we would like to emphasize the (i) equinoxial asymmetry of TEC, (ii) occurrence of lowest TEC in June solstice, and (iii) high values of TEC in December. Features (ii) and (iii) were recently reported based on similar data by Mungufeni et al. (2019). The reader may refer to this study for more discussions. Mungufeni et al. (2016) observed equinoxial asymmetry when studying ionospheric irregularities over the African low-latitude region. They observed over the East African region that the irregularity strength in the March equinox was higher than that in the September equinox. They attributed the equinoxial asymmetry to meridional winds in March which might blow northward. Such a direction would lift plasma up where recombination is not common. On the other hand, in September, the winds might blow southward. This could lead to recombination at low altitudes.

5 Model validation

5.1 Validation using reserved COSMIC RO TEC

In addition to comparing observed binned TEC with the corresponding modeled TEC, we validated our model using observed TEC in the years 2012 and 2018. The data during these 2 years were not used in developing the model. The TEC data in the years 2012 and 2018 were binned according to local time and spatially in a similar manner to the data mentioned in Sect. 2.2. The corresponding local time, day of year, solar flux, and spatial coordinates of the data were noted and then used to generate the corresponding modeled TEC. Despite the advantages of B spline modeling mentioned in Sect. 1, one of its limitations is the inability to extrapolate. Therefore, in situations where the solar flux level is higher (lower) than the values specified in Table 2, the maximum (minimum) value in the table was used to generate the corresponding modeled TEC. This idea was also applied when the day of year, longitude, and latitude values were higher (lower) than those specified in Sect. 3.

Figure 4 presents a scatter plot showing the observed TEC against the corresponding modeled TEC. The red line in the figure indicates the linear least-squares fit to the data in the panel. Furthermore, indicated in Fig. 4 are (i) the correlation coefficients (r) (ii) the r squared values, (iii) the number of data points (n plotted), and (iv) the root mean square error (RMSE) when the modeled TEC is used to represent the observed TEC.
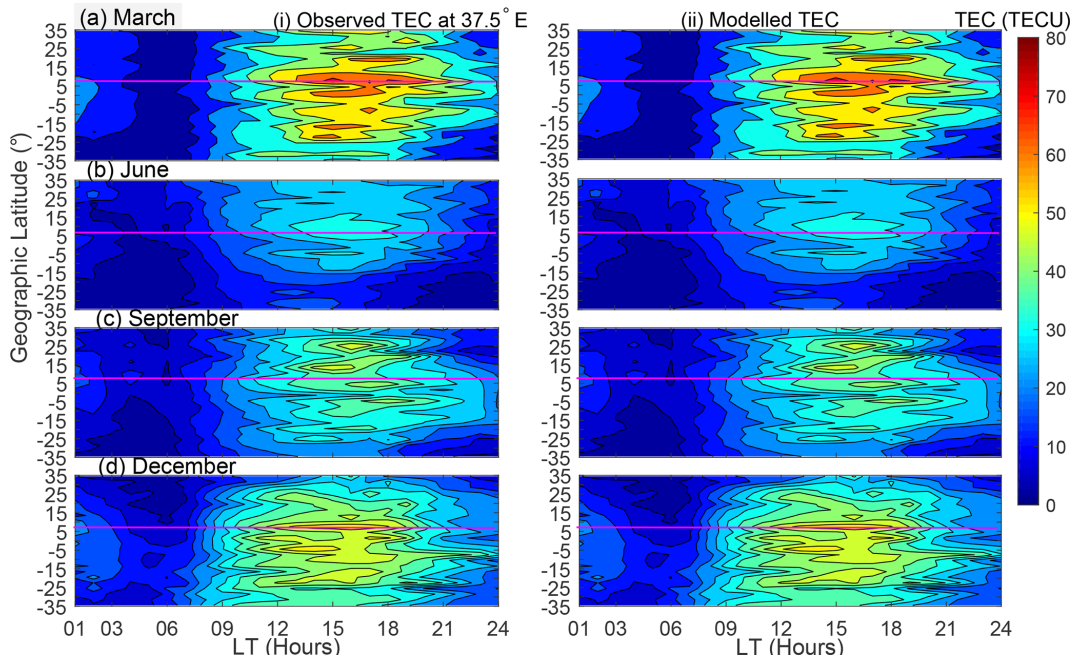
Figure 3. Variation of TEC as a function of latitude and local time in HSA at 37.5° E. Panels in rows **(a–d)** are for the March equinox, June solstice, September equinox, and December solstice, respectively, while panels in columns (i) and (ii) are observed binned and modeled TEC, respectively. The magenta line indicates 0° dip latitude.
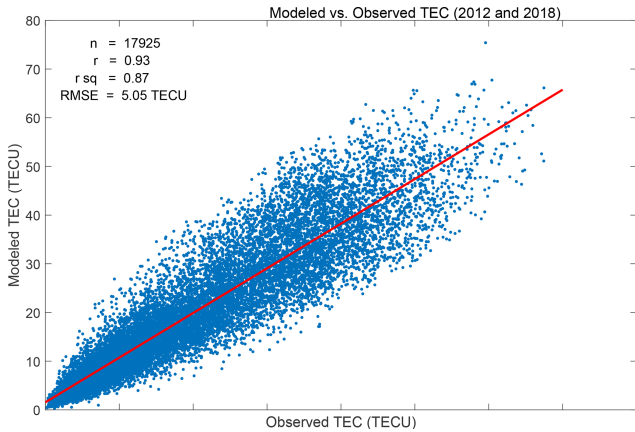

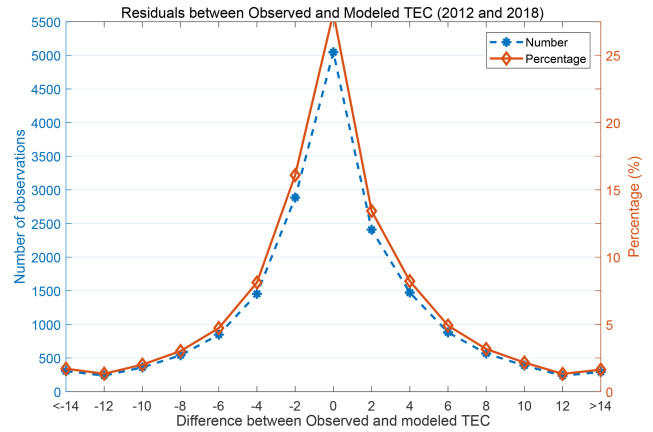**Figure 4.** Scatter plot of observed TEC against modeled TEC.

Figure 5. The blue and red curves show the distribution of the number of observed errors (difference between observed and modeled TEC) and the percentage of the errors, respectively.

The following observations can be noted from Fig. 4. (i) The modeled TEC correlates highly ($r \sim 0.93$) with the observed TEC. (ii) The r squared values indicate that high proportions ($\sim 87\%$) of the variations in the observed TEC can be predicted by the modeled TEC. (iii) The RMSE value of 5.05 TECU signifies that the modeled TEC closely approximates the observed TEC.

In order to show that the observed and modeled TEC have similar magnitudes in addition to their similar variation depicted in Fig. 4, we computed the differences between corresponding values of the data plotted in the figure. These were

referred to as errors. We also computed the percentage of the different errors. The left and right vertical axes in Fig. 5 present the distribution of the number of observed errors and their percentages, respectively. It can be seen from the figure that the errors are randomly distributed since the distribution curve is symmetric about 0 TECU. Indeed, the magnitudes of the modeled TEC values are close to those of the observed TEC since the majority of the error values are close to zero.

The cases of high error values (> 10 TECU) mostly have $< 2.5\%$ occurrence probability, as can be seen on the right

vertical axis. These high errors may be partly attributed to the limitation of the spline modeling technique (inability to extrapolate) which was discussed earlier in this subsection, Sect. 5.1.

5.2 Validation using ionosonde TEC measurements

The TEC data measured by the digisonde ionosonde stations over South Africa located at Hermanus, Grahamstown, and Louisvale can be accessed from the National Oceanic and Atmospheric Administration (NOAA) website via the link ftp://ftp.ngdc.noaa.gov (last access: 12 September 2020). The data obtained from the NOAA website are in the form of auto-scaled ionospheric parameters such as peak height in the F2 layer, critical frequency in the F2 layer, and TEC which are stored in Standard Archiving Output (SAO) format files. It should be noted that the TEC data provided in SAO files are obtained by integrating electron density profiles up to altitudes of ~ 700 km. More details about the auto-scaling program (real-time ionogram scaler with true height (ARTIST)) and the electron density profiles they produce can be found in Reinisch and Huang (2001) and Klipp et al. (2020).

The magenta lines in Fig. 6 present the diurnal patterns of TEC measured by ionosonde stations at Hermanus (panels in column i), Grahamstown (panels in column ii), and Louisvale (panels in column iii). The corresponding TEC generated by our spline modeling technique (spline), NeQuick 2, and IRI-2016 is superimposed with red, green, and blue lines, respectively. We need to mention that during computation of TEC using NeQuick 2 and IRI-2016, the height was limited to the approximate altitude of the COSMIC satellites (800 km). Moreover, for the case of IRI-2016, the NeQuick model option was specified to estimate topside electron density.

The panels in rows (a)–(c) show TEC on day of year 170 (June), 260 (September), and 350 (December), respectively. These three days of the year in 2013 were geomagnetically quiet. Preliminarily, Fig. 6 appears to reveal that IRI-2016 either overestimates (December) or underestimates (June and September) the TEC measured by the ionosonde stations. On the other hand, our spline modeling technique and NeQuick 2 seem to depict good correspondence between the observed and the modeled TEC. It can also be seen from Fig. 6 that over a particular station, the shape of curves on different days representing TEC generated by the IRI-2016 and NeQuick 2 models is similar. This is expected since these two models were meant to reproduce monthly median values of the ionosphere. This means that our model, based on spline functions, may capture the day-to-day variability of the ionosphere better.

We generated such data plotted in Fig. 6 for geomagnetically quiet days of the entire year 2013 and then performed statistical analysis of the observed and the model TEC data. Table 3 presents in column 3 the correlation coefficients (r) for the correlations between modeled and ionosonde TEC. Moreover, the table presents the RMSE when the ionosonde

Table 3. Correlation coefficients, r and RMSE associated with estimation of TEC observed by ionosonde stations using models.

Ionosonde station/ number of observations	Model	r (TECU)	RMSE
Hermanus ($n = 5110$)	Spline	0.92	4.64
	IRI-2016	0.86	5.45
	NeQuick 2	0.92	4.10
Grahamstown ($n = 4450$)	Spline	0.88	5.56
	IRI-2016	0.82	6.29
	NeQuick 2	0.86	5.27
Louisville ($n = 4543$)	Spline	0.94	3.82
	IRI-2016	0.87	5.62
	NeQuick 2	0.94	3.73

TEC was estimated using the models listed in column 2. The number of observations (n) over each station that were used to determine r and RMSE are put in brackets below the station name.

It can be seen from Table 3 that the r values associated with NeQuick 2 and the spline-based model are consistently better when compared with those of IRI-2016. Moreover, the RMSE values associated with IRI-2016 are the highest in all the cases. These two observations indicate that compared to the spline model and NeQuick 2, IRI-2016 poorly estimates TEC at the locations of the ionosondes. The RMSE values associated with NeQuick 2 are always slightly lower than that of the spline model, while the r values associated with the spline model are mostly comparable or slightly higher than that of NeQuick 2. These discussions demonstrate that our spline model generates TEC values consistently with that observed by ionosondes. This implies that equivalent TEC measured by ionosondes over midlatitude locations which do not have ionosonde stations can be predicted fairly well using our model. We might validate our model over the low-latitude region that falls within the current study area if ionosonde observations become available over the region in future.

5.3 Comparison of our model with existing regional models

It would be good to compare error levels produced when some measured TEC values are compared with modeled TEC generated by (i) the existing regional TEC models discussed in Sect. 1 and (ii) our TEC model of the spline technique. We may not perform such analysis since models in (i) are based on electron density integrated from the ground up to GPS satellites ($\sim 20\,200$ km), while the model in (ii) is based on electron density integrated up to ~ 800 km. However, we present Figs. 7 and 8 to compare EIA features captured by our spline modeling technique with those by the neural network technique of Okoh et al. (2019). The TEC plots based on the neural network technique can be
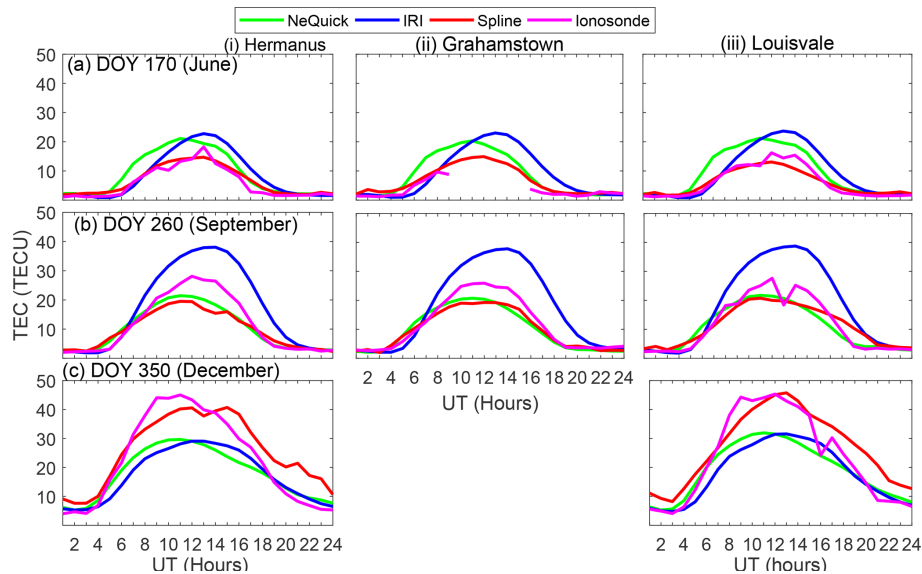
Figure 6. Magenta color shows diurnal TEC observed by ionosonde stations at Hermanus (panels in column i), Grahamstown (Panels in column ii), and Louisvale (Panels in column iii). The green, blue, and red colors show TEC estimations using NeQuick 2, IRI-2016, and spline models, respectively. Panels in rows **(a–c)** show diurnal TEC during the year 2013 on DOY 170, 260, and 350, respectively.

obtained from the MATLAB central website (Okoh et al., 2019). (https://www.mathworks.com/matlabcentral/fileexchange/69257-african-gnss-tec-afritec-model?s_tid=prof_contriblnk, last access: 12 September 2020). We present in Fig. 7 examples of TEC generated by the neural network model during the year 2012 at 11:00 UT. Over the East African sector (LT = UT + 3), this time translates to 14:00 LT and falls within the range of LT when EIA exists over the region (Mungufeni et al., 2018). Panels (a) and (b) in Fig. 7 present TEC during March (DOY 81) and September (DOY 260) equinoxes, respectively, while (c) and (d) present TEC during June (DOY 171) and December (DOY 347) solstices, respectively. It is important to mention that these four days were geomagnetically quiet.

In order to generate TEC maps using our model for purposes of comparing with TEC maps in Fig. 7, we noted and used the F10.7 flux values on the days indicated in the figure. The TEC maps generated using our model that correspond to TEC maps presented in Fig. 7 are presented in Fig. 8.

Unlike our TEC maps in Fig. 8 which clearly show the EIA trough (see magenta arrows) in all the seasons, the neural network technique TEC maps (Okoh et al., 2019) of Fig. 7 only clearly capture the EIA trough in the December solstice. As pointed out before, this shortfall in the neural network TEC model might be due to a poor amount of data to represent the day of year during model development. Another observation that can be made from Figs. 7 and 8 is that unlike the neural network model which yields smooth spatial TEC variation, the spline modeling technique does not yield smooth spatial TEC variation. In real life, measurement or observed values rarely vary smoothly. Since the spline modeling tech-

nique produces results (see Fig. 2) which demonstrate that the modeled data match the observed data almost perfectly, it is expected that the spatial variations of TEC in maps of Fig. 8 are not smooth.

6   Conclusions

This study developed a model of TEC measured by COS-MIC satellites. The TEC data were binned according to local time, seasons, solar flux level, and spatially. The coefficients of B splines that were fitted to the binned data were determined by means of the least-squares procedure. As expected, the modeled TEC almost perfectly matched the corresponding observed binned TEC data. The model was validated with independent data that were not used in the model development. The validation revealed that (i) the observed and the modeled TEC correlate highly ($r = 0.93$), (ii) the coefficient of determination R^2 which is the proportion of variance in the observed data predicted by our model was 87 %, and (iii) the modeled TEC closely approximates the observed TEC (RMSE of 5.05 TECU). Due to the extensive input data and the modeling technique applied, we were able to reproduce the well-known features of TEC variation over the African region. Further validation of our model using TEC obtained from ionosonde stations over South Africa at Hermanus, Grahamstown, and Louisville reported r values > 0.92 and RMSE < 5.56 TECU. These validation results imply that our model can estimate TEC fairly well that would be measured by ionosondes over locations which do not have the instrument.
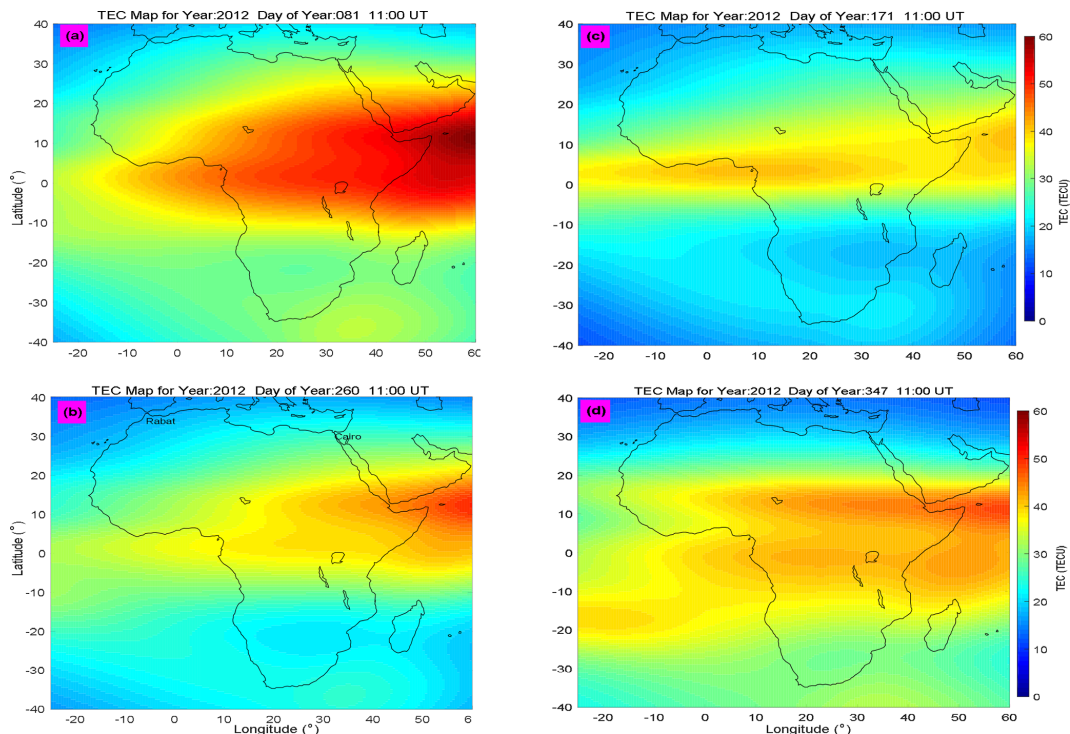
Figure 7. Neural network TEC maps during the year 2012 at 11:00 UT. Panels **(a)** and **(b)** are for March (DOY 81) and September (DOY 260) equinoxes, respectively, while **(c)** and **(d)** are for June (DOY 171) and December (DOY 347) solstices, respectively.
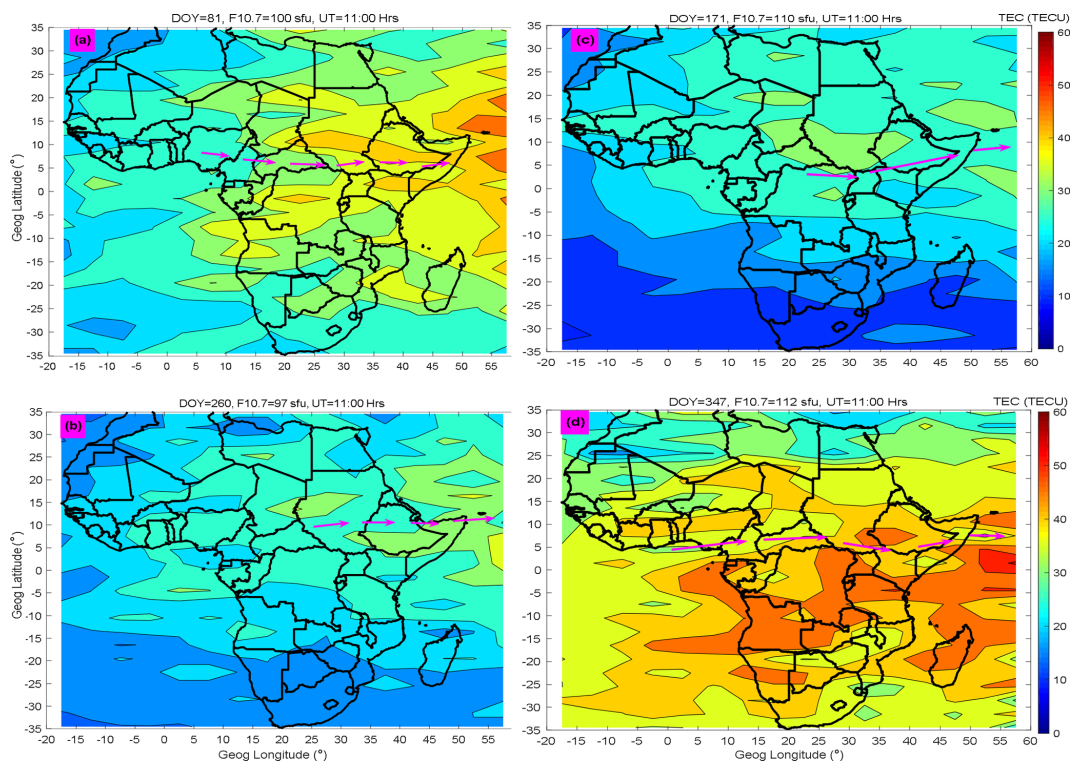

Figure 8. Similar to Fig. 7 but generated by the spline modeling technique. Magenta arrows indicate approximate locations of the EIA trough.

P. Mungufeni et al.: Modeling total electron content

References

Abdu, M. A., Souza, J. R., Batista, I. S., and Sobral, J. H. A.: Equatorial spread F statistics and empirical representation for IRI: A regionalmodel for the Brazilian longitude sector, Adv. Space Res., 31, 703–716, 2003.

Adewale, A. O., Oyeyemi, E. O., Adeloye, A. B., Ngwira, C. M., and Athieno, R.: Responses of equatorial

F region to different geomagneticstorms observed by GPS in the African sector, J. Geophys. Res., 116, A12319, https://doi.org/10.1029/2011JA016998, 2011.

Appleton, E. V.: Two Anomalies in the Ionosphere, Nature, 3995, p. 691, 1946.

Batista, I. S., Abdu, M. A., and Bittencourt, J. A.: Equatorial F region vertical plasma drifts: seasonal and longitudinal asymmetries in the American sector, J. Geophys. Res., 91, 12055–12064, 1986.

Bilitza, D.: International Reference Ionosphere 2000, Radio Sci., 36, 757–767, 2001.

Bolaji, O., Owolabi, O., Falayi, E., Jimoh, E., Kotoye, A., Odeyemi, O., Rabiu, B., Doherty, P., Yizengaw, E., Yamazaki, Y., Adeniyi, J., Kaka, R., and Onanuga, K.: Observations of equatorial ionization anomaly over Africa and Middle East during a year of deep minimum, Ann. Geophys., 35, 123–132, https://doi.org/10.5194/angeo-35-123-2017, 2017.

Buonsanto, M. J.: Ionospheric Storms – A Review, Space Sci. Rev., 88, 563–601, 1999.

COSMIC data analsysis center: Level 2 – ionospheric profiles, available at: http://cosmic-io.cosmic.ucar.edu/cdaac/index.html, last access: 12 September 2020.

De-Boor, C.: A Practical Guide to Splines, Springer, New York, USA, 1978.

Emmert, J. T., Richmond, A. D., and Drob, D. P.: A computationally compact representation of Magnetic-Apex and Quasi-Dipole coordinates with smooth base vectors, J. Geophys. Res., 15, A08322, https://doi.org/10.1029/2010JA015326, 2010.

Ercha, A., Zhang, D., Ridley, A. J., Xiao, Z., and Hao, Y.: A global model: Empirical orthogonal function analysis of total electron content 1999–2009 data, J. Geophys. Res., 117, A03328, https://doi.org/10.1029/2011JA017238, 2012.

Gamache, R. R., Galkin, I. A., and Reinisch, B. W.: A Database Record Structure for Ionogram Data, University of Lowell Center for Atmospheric Research, available at: ftp://ftp.ngdc.noaa.gov, last access: 12 September 2020.

Guochang, X.: GPS. Theory, Algorithms, and Applications, Springer, New York, USA, 2007.

Habarulema, J. B., McKinnel, L., and Opperman, B.: Regional GPS TEC modeling; attempted spatial and temporal extrapolation of TEC using neural networks, J. Geophys. Res.-Space Phys., 116, A04314, https://doi.org/10.1029/2010JA016269, 2011.

Hargreaves, J. K.: The Solar-Terrestrial environment, Cambridge University Press, New York, USA, 1992.

Hofmann-Wellenhof, B., Lichtenegger, H., and Wasle, E.: Global Navigation Satellite Systems, GPS, GLONASS, Galileo and more, Springer, New York, USA, 2007.

Jakowski, N., Hoque, M. M., and Mayer, C.: A new global TEC model for estimating trans-ionospheric radio wave propagation errors, J. Geod., 85, 965–974, 2011a.

Jakowski, N., Schlüter, S., and Sardon, E.: Total electron content models and their use in ionosphere monitoring, Radio Sci., 46, RS0D18, https://doi.org/10.1029/2010RS004620, 2011b.

Klipp, T. D. S., Petry, A., de Souza, J. R., de Paula, E. R., Falcão, G. S., and de Campos Velho, H. F.: Ionosonde total electron content evaluation using International Global Navigation Satellite System Service data, Ann. Geophys., 38, 347–357, https://doi.org/10.5194/angeo-38-347-2020, 2020.

Klobuchar, J. A.: Ionospheric time-delay algorithm for single frequency GPS users, IEEE Trans. Aerosp. Electron. Syst., 23, 325–331, https://doi.org/10.1109/TAES.1987.3-10829, 1987.

Krankowski, A., Zakharenkova, I., Krypiak-Gregorczyk, A., Shagimuratov, I. I., and Wielgosz, P: Ionospheric electron density observed by FORMOSAT-3/COSMIC over the European region and validated by ionosonde data, J. Geodesy, 85, 949–964, https://doi.org/10.1007/s00190-011-0481-z, 2011.

Leva, J. L., de Haag, M. U., and Dyke, K. V.: Performance of standalone GPS, in: Understanding GPS: Principles and Applications, edited by: Kaplan, E. D. and Hegarty, C. J., Artech House, Inc. 685 Canton Street Norwood, MA 02062, 66–112, 2006.

Nose, M., Iyemori, T., Sugiura, M., and Kamei, T.: Geomagnetic Dst index, World Data Center for Geomagnetism, Kyoto, available at: http://swdcwww.kugi.kyoto-u.ac.jp/, last access: 12 September 2020.

Matzka, J. and Stolle, C.: Indices of Global Geomagnetic Activity, World Data Center for Geomagnetism, Kyoto, available at: ftp://ftp.gfz-potsdam.de/pub/home/obs/kp-ap/, last access: 12 September 2020.

Mengist, C. K., Ssessanga, N., Jeong, S.-H., Kim, J.-H., Kim, Y. H., and Kwak, Y.-S: Assimilation of multiple data types to a regional ionosphere model with a 3D-Var algorithm (IDA4D), Space Weather, 17, 1018–1039, https://doi.org/10.1029/2019SW002159, 2019.

Mukhtarov, P., Pancheva, D., Andonov, B., and Pashova, L.: Global TEC maps based on GNSS data: 1. Empirical background TEC model, J. Geophys. Res.-Space, 118, 4594–4608, 2013.

Mungufeni, P., Jurua, E., Habarulema, J. B., and Anguma, S. K.: Modeling the probability of ionospheric irregularity occurrence over African low latitude region, J. Atmos. Sol.-Terr. Phy., 128, 46–57, 2015.

Mungufeni, P., Habarulema, J. B., and Jurua, E. A.: Trends of ionospheric irregularities over African low latitude region during quiet geomagnetic conditions, J. Atmos. Sol.-Terr. Phy., 14, 261–267, 2016.

Mungufeni, P., Habarulema, J. B., Migoya-Orué, Y., and Jurua, E.: Statistical analysis of the correlation between the equatorial electrojet and the occurrence of the equatorial ionisation anomaly over the East African sector, Ann. Geophys., 36, 841–853, https://doi.org/10.5194/angeo-36-841-2018, 2018.

Mungufeni, P., Rabiu, A. B., Okoh, D., and Jurua, E.: Characterization of Total Electron Content over African region using Radio Occultation observations of COSMIC satellites, Adv. Space Res., 65, 19–29, https://doi.org/10.1016/j.asr.2019.08.009, 2019.

Najman, P. and Kos, T.: Performance Analysis of Empirical Ionosphere Models by Comparison with CODE Vertical TEC Maps, in: Mitigation of Ionospheric Threats to GNSS: an Appraisal of the Scientific and Technological Outputs of the TRANSMIT Project, edited by: Notarpietro, R., InTech Open Science Publications, 5 Princes Gate Court, London, SW7 2QJ, United Kingdom, 162–178, https://doi.org/10.5772/58774, 2014.

Nava, B., Coïsson, P., and Radicella, S. M.: A new version of the NeQuick ionosphere electron density model, J. Atmos. Sol.-Terr. Phy., 70, 1856–1862, 2008.

Okoh, D., Owolabi, O., Ekechukwu, C., Folarin, O., Arhiwo, G., Agbo, J., Bolaji, S., and Babatunde, R.: A regional GNSS-VTEC model over Nigeria using neural networks: A novel approach, Geodesy and Geodynamics, 7, 19–31, 2016.

Okoh, D., Seemala, G., Rabiu, B., Habarulema, J. B., Jin, S., Shiokawa, K., Otsuka, Y., Aggarwal, M., Uwamahoro, J., Mungufeni, P., Segun, B., Obafaye, A., Ellahony, N., Okonkwo, C., Tshisaphungo, M., and Shetti, D.: A Neural Network-Based Ionospheric Model Over Africa From Constellation Observing System for Meteorology, Ionosphere, and Climate and Ground Global Positioning System Observations, J. Geophys. Res.-Space, 124, https://doi.org/10.1029/2019JA027065, 2019.

Opperman, B.: Reconstructing ionospheric TEC over South Africa using signals from a regional GPS network, PhD Thesis, Rhodes University, Grahamstown, South Africa, 30 pp., 2008.

Rama Rao, P. V. S., Jayachandran, P. T., Sri Ram, P., Ramana Rao, B. V., Prasad, D. S. V. V. D., and Bose, K. K.: Characteristics of VHF radiowave scintillations over a solar cycle (1983–1993) at a low-latitude station: Waltair (17.7° N, 83.3° E), Ann. Geophys., 15, 729–733, https://doi.org/10.1007/s00585-997-0729-3, 1997.

Reinisch, B. and Huang, X.: Deducing topside profiles and total electron content from bottomside ionograms, Adv. Space Res., 27, 23–30, https://doi.org/10.1016/S0273-1177(00)001368, 2001.

Rishbeth, H. and Garriot, O. K.: Introduction to Ionospheric Physics, Academic Press, New York, USA, 1969.

Scherliess, L. and Fejer, B. G.: Radar and satellite global equatorial F-region vertical drift model, J. Geophys. Res., 104, 6829–6842, 1999.

Schunk, W. R. and Nagy, F. A.: Ionospheres: Physics, Plasma Physics, and Chemistry, Cambridge University press, New York, USA, 2009.

Sun, Y., Liu, J., Tsai, H., and Krankwski, A.: Global ionospheric map constructed by using TEC from ground-based GNSS receiver and FORMSAT-3/COSMIC GPS occultation experiment, GPS solutions, 21, 1583–1591, https://doi.org/10.1007/s10291-017-0635-4, 2017.

Tebabal, A., Radicella, S. M., Damtie, B., Migoya-Orue', Y., Nigussie, M., and Nava, B.: Feed forward neural network based ionosphericmodel for the East African region, J. Atmos. Sol.-Terr. Phy., 191, 1–10, https://doi.org/10.1016/j.jastp.2019.05.016, 2019.

Thébault, E., Finlay, C. C., Beggan, C. D., Alken, P., Aubert, J., Barrois, O., Bertrand, F., Bondar, T., and Bones, A.: International Geomagnetic Reference Field: the 12th generation, Earth Planet. Sp., 79, 67–79, https://doi.org/10.1186/s40623-015-0228-9, 2015.

U.S. Dept. of Commerce, NOAA: Daily Solar Data, Space Environment Center, available at: http://www.swpc.noaa.gov/, last access: 12 September 2020.

Zhang, M.-L., Wan, W., Liu, L., and Ning, B.: Variability study of the crest-to-trough TEC ratio of the equatorial ionization anomaly around 120° E longitude, Adv. Space Res., 43, 1762–1769, 2009.