JGR Space Physics

RESEARCH ARTICLE
10.1029/2019JA027065

A Neural Network-Based Ionospheric Model Over Africa From Constellation Observing System for Meteorology, Ionosphere, and Climate and Ground Global Positioning System Observations

Daniel Okoh[1,2], Gopi Seemala[2], Babatunde Rabiu[1], John Bosco Habarulema[3,4], Shuanggen Jin[5,6], Kazuo Shiokawa[7], Yuichi Otsuka[7], Malini Aggarwal[2], Jean Uwamahoro[8], Patrick Mungufeni[9], Bolaji Segun[10], Aderonke Obafaye[1], Nada Ellahony[11], Chinelo Okonkwo[12], Mpho Tshisaphungo[3], and Dadaso Shetti[13]

[1]Center for Atmospheric Research, National Space Research and Development Agency, Anyigba, Nigeria, [2]Indian Institute of Geomagnetism, Navi Mumbai, India, [3]South African National Space Agency, Space Science, Hermanus, South Africa, [4]Department of Physics and Electronics, Rhodes University, Grahamstown, South Africa, [5]School of Remote Sensing and Geomatics Engineering, Nanjing University of Information Science and Technology, Nanjing, China, [6]Shanghai Astronomical Observatory, Chinese Academy of Sciences, Shanghai, China, [7]Institute for Space-Earth Environmental Research, Nagoya University, Nagoya, Japan, [8]College of Education, Department of Mathematics and Science, University of Rwanda, Rwamagana, Rwanda, [9]Physics Department, Mbarara University of Science and Technology, Mbarara, Uganda, [10]Department of Physics, University of Lagos, Lagos, Nigeria, [11]Space Weather Monitoring Center, Helwan University, Helwan, Egypt, [12]Department of Physics & Astronomy, University of Nigeria, Nsukka, Nsukka, Nigeria, [13]Smt. Kasturbai Walchand College, Sangli, India

Key Points:
- The first regional TEC model over the entire African region using empirical observations is developed
- The model offers opportunities to conduct high spatial resolution investigations over the African region
- EIA occurrence is reduced during the June solstice, and the anomaly region shifts southwards during December solstice

Correspondence to:
D. Okoh, and G. Seemala,
okodan2003@gmail.com
gopi.seemala@gmail.com

Citation:
Okoh, D., Seemala, G., Rabiu, B., Habarulema, J. B., Jin, S., Shiokawa, K., et al. (2019). A neural network-based ionospheric model over Africa from Constellation Observing System for Meteorology, Ionosphere, and Climate and Ground Global Positioning System observations. *Journal of Geophysical Research: Space Physics*, *124*. https://doi.org/10.1029/2019JA027065

Received 22 JUN 2019
Accepted 9 NOV 2019
Accepted article online 9 DEC 2019

Abstract The first regional total electron content (TEC) model over the entire African region (known as AfriTEC model) using empirical observations is developed and presented. Artificial neural networks were used to train TEC observations obtained from Global Positioning System receivers, both on ground and onboard the Constellation Observing System for Meteorology, Ionosphere, and Climate satellites for the African region from years 2000 to 2017. The neural network training was implemented using inputs that enabled the networks to learn diurnal variations, seasonal variations, spatial variations, and variations that are connected with the level of solar activity, for quiet geomagnetic conditions (-20 nT $\leq$ Dst $\leq$ 20 nT). The effectiveness of three solar activity indices (sunspot number, solar radio flux at 10.7-cm wavelength [F10.7], and solar ultraviolet [UV] flux at 1 AU) for the neural network trainings was tested. The F10.7 and UV were more effective, and the F10.7 was used as it gave the least errors on the validation data set used. Equatorial anomaly simulations show a reduced occurrence during the June solstice season. The distance of separation between the anomaly crests is typically in the range from about 11.5 $\pm$ 1.0° to 16.0 $\pm$ 1.0°. The separation is observed to widen as solar activity levels increase. During the December solstice, the anomaly region shifts southwards of the equinox locations; in year 2012, the trough shifted by about 1.5° and the southern crest shifted by over 2.5°.

1. Introduction

Knowledge of the ionosphere, especially for lesser known areas like the African sector, is required to enhance understanding and exploitation of it. The ionosphere is the region of the Earth's upper atmosphere that contains considerable amount of charged particles to influence electromagnetic wave propagation through it. Trans-ionospheric radio propagations are affected by the ionosphere through phenomena that include refraction and signal delays. For satellite-based navigation systems, the signal delays translate to positioning errors, causing one of the greatest sources of positioning errors for single-frequency Global Navigation Satellite System (GNSS) receivers (Jin et al., 2006, 2007; Jin & Najibi, 2014; Liu et al., 2018; Otsuka et al., 2002; Perez, 2017). As a result of this predicament, interest to improve our understanding of the ionosphere and its variability has increased.

A lot of effort has been put into improving our understanding of the ionosphere and its variability. One of such efforts includes the development and continuous update of global ionospheric models like the

NeQuick (Hochegger et al., 2000; Nava et al., 2008; Radicella & Leitinger, 2001) and the International Reference Ionosphere (IRI; Bilitza, 2001; Bilitza & Reinisch, 2008) model. Although these models have been adequately shown to be reliable in predicting the trends and patterns of variations in the ionosphere (e.g., Aggarwal, 2011; Aggarwal & Sharma, 2017; Sharma et al., 2017; and many more), a number of research results (e.g., Adewale et al., 2011; Akala et al., 2013, 2015; Habarulema et al., 2007, 2009; Okoh, McKinnel, et al., 2015; Okoh, Onwuneme, et al., 2018; Oyeyemi et al., 2018; Olwendo et al., 2012; Rabiu et al., 2011, 2014) have reported certain shortfalls in the African ionosphere. One often cited reason for this shortfall is that the volume of data from the African region used in the development of these models is relatively small, and this is owing to the paucity of ionospheric data available from the region (Bilitza & Reinisch, 2008; McKinnell, 2002; Okoh et al., 2016).

The ionospheric time delay of radio signals exhibits an inverse-square relationship with the radio frequency (Klobuchar, 1996), and so, multifrequency GNSS receivers are able to compute the magnitude of ionospheric time delays using frequency information of the different radio signals. The ionospheric time delays are directly proportional to a quantity of the ionosphere known as the total electron content (TEC). The TEC, measured along the path of a GNSS radio signal, represents the total number of electrons in an imaginary cylinder of a unit cross-sectional area wrapped around the path of the GNSS radio signal. If measured in a direction that is vertically overhead at a given Earth location (that is, at an elevation angle of 90°), the quantity is referred to as the vertical TEC (VTEC). The VTEC gives a sense of the overall ionospheric electron density for a given Earth location if measured to a height above the ionosphere (e.g., up to the GNSS satellite altitudes).

The GNSS has become a popular and reliable source of TEC information. As a result of their proliferation, there has been proposal to use TEC information obtained from GNSS receivers in various locations of the African region for TEC modeling over the continent. In this work, we develop the first regional Global Positioning System (GPS) VTEC model (hereafter referred to as AfriTEC) over the entire African continent using the method of computer neural networks. Particularly, we report the quiet-time modeling of TEC in this paper. There is a proposal to carry out detailed and extensive research on quantifiable effects of geomagnetic storms on TEC and to incorporate the results into subsequent versions of AfriTEC.

There is limitation to using ground-based GNSS receivers to gather spatial information on ionospheric variation. Asides the cost that may be incurred in trying to provide these instruments at reasonable spatial intervals, there is also inherent difficulty in hosting/maintaining them at locations like oceans, deserts, and crisis-prone states. For instance, there is presently very sparse distribution of GNSS receivers in and around the Sahara desert and in the ocean waters around the African continent. This has made it difficult to accurately model the African ionosphere using ground-based GNSS receivers. The advent of space-based observations has provided alternative methods for studying the ionosphere and for overcoming the aforediscussed limitation associated with using ground-based observations. TEC obtained from integrated electron densities of radio occultation (RO) observations from GPS receivers onboard the Constellation Observing System for Meteorology, Ionosphere, and Climate (COSMIC) satellites was also incorporated into the AfriTEC model to enhance information on spatial variability of the ionosphere provided by the model.

Computer neural networks (also referred to as artificial neural networks) have been demonstrated by several authors (e.g., Habarulema, 2010; Habarulema et al., 2009; Okoh et al., 2016) as very efficient tools for ionospheric modeling in parts of the continent. The strengths and advantages of neural networks derive from their ability to represent both linear and nonlinear relationships directly from the data being modeled (Baboo & Shereef, 2010). Computer neural networks are a system of information processing techniques inspired by the manner in which the human brain works; they can learn trends and patterns in data and consequently be able to predict future trends in the data. For this reason, they are hugely applied in predictive modeling (Okoh, 2016). The methods of computer neural networks were applied in this work to learn the ionospheric variation patterns from the huge database of TEC obtained from networks of ground-based GPS receivers and from the COSMIC mission.

2. Data

Data used in this work include GPS data, indices for solar and geomagnetic activities, and data from ionospheric models used to comparatively verify/validate the model developed.

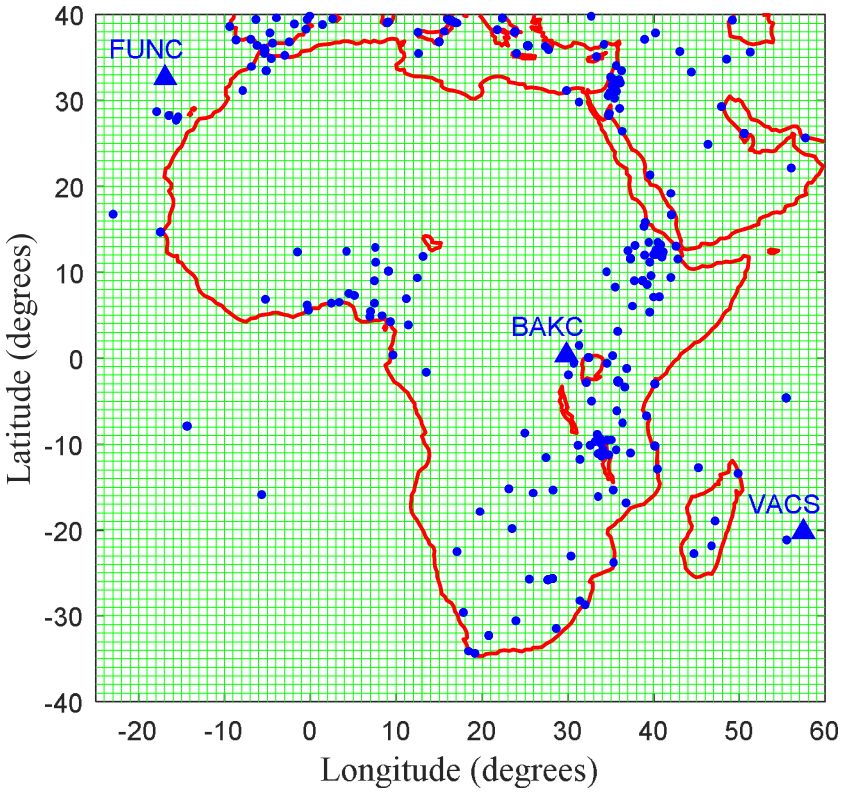

Figure 1. Map of Africa showing the locations of GPS receivers used in this work. The green background lines show 1° × 1° grids of longitude and latitude used to bin the final observations. Ground GPS data from stations marked in triangles (labeled using the four-digit station codes) and COSMIC data from the 3° × 3° grids central to the marked stations were used for testing the spatial performance of the networks.

GPS data were obtained from ground-based GPS receiver stations in different locations of the African region and also from COSMIC RO observations. The ground-based GPS receiver stations are those located within the rectangular area of region bounded by geographic longitudes −25° to 60° and geographic latitudes −40° to 40° as illustrated in Figure 1. The ground-based GPS data were obtained from the following sources:

1. African Geodetic Reference Frame (AFREF)
2. Nigeria GNSS Reference Network (NIGNET)
3. South African network of continuously operating GNSS base stations (TRIGNET)
4. University of California, San Diego, SOPAC, and CSRC GARNER GPS Archive
5. National Aeronautics and Space Administration's Crustal Dynamics Data Information System (CDDIS) Archive of GNSS products
6. Global Data Center of the International GNSS Service
7. UNAVCO Archive of GNSS Data
8. Geodetic Data Archiving Facility

The total number of ground GPS receiver stations from which data were used is 269, and all available data covering the periods from year 2000 to year 2017 were used. GPS data were obtained in Receiver Independent Exchange format and processed into TEC information using the most recent version of the GPS-TEC analysis application software (Version 2.9.5) developed by Seemala and Valladares (2011), which can be downloaded from the webpage https://seemala.blogspot.com. The GPS-TEC program reads the raw GPS data from Receiver Independent Exchange files and calculates phase and code TEC values along with corresponding elevation and azimuth angles of the satellite(s) for the epochs. The satellite differential code biases (DCB) values calculated by the Center for Orbit Determination in Europe (CODE), and published at webpage ftp://ftp.aiub.unibe.ch/CODE/, were used, and the receiver bias (DCB) is estimated by the GPS-TEC program using least squares minimization, assuming that the cumulative differences between individual satellite VTECs and average (2 sigma mean) VTECs derived from the same individual satellites will

be minimal when receiver DCB is corrected. In this program, the single shell mapping function (Langley, 2002; Mannucci et al., 1993) is used to calculate the vertical TEC at ionospheric pierce point height of 350 km (Rama Rao et al., 2006), which can be safely assumed as effective ionospheric height for the region studied in this work. TEC computed using the program is integrated up to the GPS satellite altitude of about 20,200 km.

After using this method to compute the VTECs on a 30-s time resolution, we obtained a total of 14,934,051,858 data points of the VTEC. To reduce the volume of data and to smoothen out data spikes, the computed VTEC data were averaged in 1-hr interval. To minimize multipath errors, we excluded VTEC computed from satellites with elevation angles less than or equal to 30°. Since the interest in this paper is to model TEC during quiet conditions, we also excluded TEC data for geomagnetically disturbed days. Only data for days with disturbance storm time (Dst) indices in the range $-20\ \text{nT} \leq \text{Dst} \leq 20\ \text{nT}$ were used in this work. At the end of all data processing and reductions, the volume of data was reduced to 11,191,966 data points of VTEC.

GPS data from the COSMIC mission were also used in this work. The COSMIC mission was launched in April 2006 and consists of six LEO microsatellites assigned mainly for the RO remote sensing of the atmosphere and ionosphere. Initially, the orbital altitudes of the COSMIC satellites were about 500 km, and towards the end of year 2007, the orbital altitudes were raised to the final planned altitude of about 800 km. Chen-Joe (2008) and Syndergaard et al. (2006) contain detailed information on the COSMIC constellation deployment. For consistency of the integration limits of TEC from the COSMIC mission used in this work, we have used COSMIC data starting from year 2008 to year 2017. COSMIC data used in this work consist of the second level data provided by the COSMIC Data Analysis and Archive Centre. These are the "ionprf" files that contain information about ionospheric electron density profiles. The electron density profiles for given RO events were integrated to obtain TEC, and the obtained TEC value was assigned to the geographic location that had the maximum electron density on the profile. For this study, only geographic locations within the rectangular area of region bounded by geographic longitudes −25° to 60° and geographic latitudes −40° to 40° were used. The total number of RO events obtained was 257,831. To reduce errors associated with TEC integrated from electron density profiles with wide spatial span, only TEC values integrated from electron density profiles with latitude spans less than 10° and longitude span less than 15° were used. Also, for the purpose of quiet time modeling proposed in this work, observations corresponding to instances when the Dst is outside the range $-20\ \text{nT} \leq \text{Dst} \leq 20\ \text{nT}$ were excluded. This reduced the number of data points to 114,891.

Data used for the neural network training consisted of both ground-based observations and the calibrated COSMIC observations. The observations (put together) were binned in hourly grids of $1° \times 1°$ cells as illustrated in Figure 1. The hourly medians of observations belonging to the same cells were computed and taken to represent the TEC values for the geographic location corresponding to the midpoint of the cells. The total number of data points obtained after doing this was 7,567,926.

Other data used in this work are indicators of solar and geomagnetic activity levels. The indicators of solar activity levels used in this work include sunspot number (SSN), solar radio flux at 10.7-cm wavelength (F10.7), and solar ultraviolet (UV) flux at 1 AU. The indicator of geomagnetic activity level used is the Dst index. SSN data were obtained from the WDC-SILSO, Royal Observatory of Belgium, Brussels. F10.7 data were obtained from the National Oceanic and Atmospheric Administration. Solar UV flux data were obtained from the University of South California Dornsife. Dst indices were obtained from National Aeronautics and Space Administration's OMNIWeb.

Models used in this work include the NeQuick-2 and IRI-Plas-2017 (IRI, extended to the plasmasphere) models (Gulyaeva et al., 2002). These models were used in a comparative verification/validation of the model developed in this work. The NeQuick-2 data were obtained by implementing the windows executable of the NeQuick FORTRAN source code obtained from the Ionosphere Radiopropagation Unit of the T/ICT4D Laboratory, International Centre for Theoretical Physics. The IRI-Plas data were also obtained by implementing the windows executable of the IRI-Plas FORTRAN source code obtained from the IZMIRAN Institute. The standard/default settings for each of the two models were used.

Global Ionospheric Maps (GIMs) are generated on a daily basis at a number of analysis centers using TEC data from hundreds of GPS/GLONASS sites of the International GNSS Service and other institutions. The GIMs used in this work are the ones from the CODE. The VTEC is modeled in a solar-geomagnetic

reference frame using a spherical harmonic expansion up to degree of order 15. Piecewise linear functions are used for representation in the time domain, and the time interval of their vertices is 2 hr (more recently, 1 hr). Each TEC map has a spatial resolution of 2.5° × 5° in the geographic latitude and longitude (CDDIS, 2019; Jee et al., 2010). CODE data for GIM were obtained from the CDDIS.

3. Methods

3.1. Calibrating COSMIC Observations

Since the goal of this research is to develop a TEC model that is mostly suitable for ground applications, the incorporation of COSMIC observations introduces a disparity. A major form of the disparity is that the TEC integration limits are different for the ground-based observations and for the COSMIC observations. To reduce this disparity, we first used the ground-based observations to calibrate the COSMIC observations by the method of neural networks. The calibration process involved the following:

1. Observations from the COSMIC mission and from ground-based receivers were separately binned in hourly 1° × 1° grids of longitude and latitude as demonstrated by the green cells in Figure 1.
2. The median values of hourly observations belonging to the same cells were computed differently for the ground-based observations and for the COSMIC observations.
3. For cells and hours in which there are both COSMIC and ground-based observations, the observations were matched.
4. Neural networks were then used to learn the relationships between the COSMIC observations and corresponding ground observations, treating the COSMIC observations as inputs and the corresponding ground observations as targets. The neural network procedure used here is similar as detailed in section 3.3; the same set of seven input layer neurons were used as it is shown in section 3.2 that the COSMIC-TEC/Ground-TEC ratio exhibits both temporal and spatial variations. Then an additional input layer neuron (which consisted of the COSMIC-TEC observations) was included to make a total of eight input layer neurons. The output layer neuron consisted of the Ground-TEC observations.

3.2. COSMIC Versus Ground-Based Observations

Prior to calibrating the COSMIC observations using corresponding ground-based observations, we investigated the relationship between the pair of observations. The blue dots in Figure 2 represent plots of the ratios of the COSMIC-TECs to the Ground-TECs, and the continuous black lines represent medians of the observations.

The value of correlation coefficient between the two sets of observations is 0.9267, and the mean of absolute differences is 3.28 TECU. The values indicate that the COSMIC observations correlate well with the ground-based observations and that the numerical difference between the two sets of observations is typically about 3.28 TECU. The root-mean-square value of the differences is 4.97 TECU. The continuous black lines in Figure 2 show that typical values for the ratios of COSMIC-TEC to Ground-TEC are around 1 and that the ratios can occasionally get to about 6. Statistics of the results show that the ratios of COSMIC-TEC to Ground-TEC that are greater than 2 constitute only about 5.7% of the observations. Given that the COSMIC-TECs are integrated to altitudes of about only 800 km, while the Ground-TECs are integrated to about 20,200 km, it is expected that the Ground-TECs be steadily greater than corresponding COSMIC-TECs. The cases in which the COSMIC-TECs are greater than Ground-TECs are expectedly connected with the different TEC computation techniques used for the two systems. For example, the COSMIC-TECs are integrated from electron density profiles that typically span much wider geographic spaces when compared to those of the Ground-TECs. Asides this, the asymmetry in electron density distribution around the space surrounding a geographic location will result in different TEC values depending on the path/direction considered. The calibration process is introduced to take care of these differences.

Figure 2a shows how the ratios of the two sets of observation vary over the years. This, in turn, signifies how the ratios vary over a range of solar activity levels. The figure shows that the ratio of COSMIC-TEC to Ground-TEC is lower during years of low solar activity (i.e., years 2008–2009 and towards year 2018). The ratios are higher during the high solar activity years (years 2014–2015). This is an indication that the level of solar activity affects the relationship between the COSMIC-TECs and Ground-TECs and therefore supports our inclusion of solar activity index as an additional input neuron for the neural network training

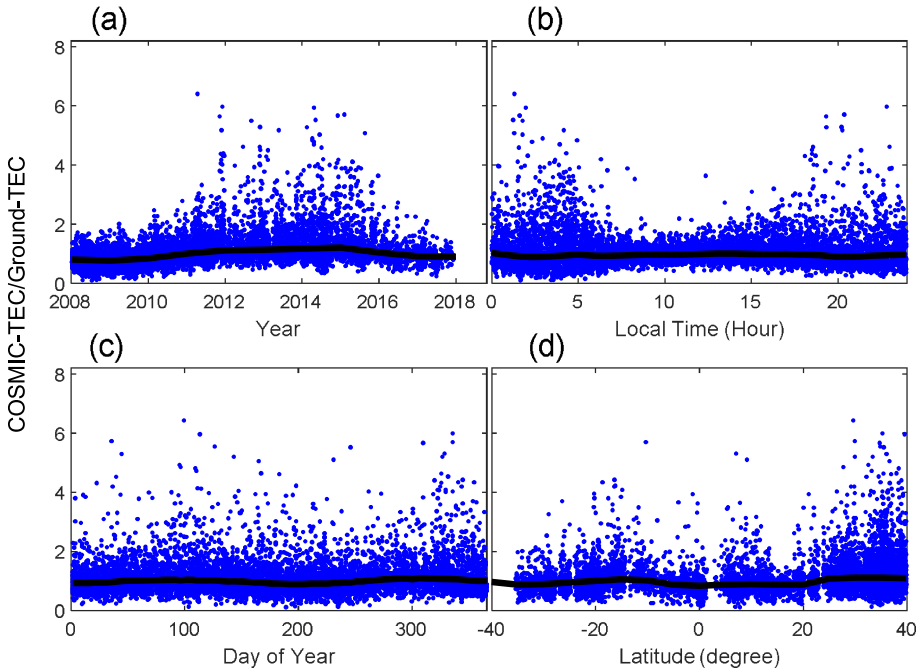

Figure 2. Plots of the ratios of COSMIC-TECs to the Ground-TECs showing how the ratios vary over (a) different years, (b) different local times (LTs), (c) different days of the years, and (d) different latitudes.

used to calibrate the COSMIC observations. Figure 2a also reveals that the observations for which the ratios of COSMIC-TECs to Ground-TECs are greater than 3 predominantly belong to the years of high solar activity. In Figure 2b, the observations for which the ratios of COSMIC-TECs to Ground-TECs are greater than 3 predominantly belong to the nights and early mornings. In Figures 2c and 2d, there are no observable systematic distributions of observations for which the ratios of COSMIC-TECs to Ground-TECs are greater than 3. Figure 2c however indicates that the ratios tend to be slightly higher during the equinoxes (around Days 90–110 and 280–300 of the years) and slightly lower during the solstices (around Days 180–200 of the years and at the start and end of the years). Figure 2d shows that the ratios are slightly higher at the midlatitudes than at the low latitudes. The plots of Figures 2b to 2d summarily indicate that the relationship between COSMIC-TEC and Ground-TEC exhibits varying degrees of temporal and spatial variations. These also support our inclusion of temporal and spatial indicators as inputs for the neural network training used to calibrate the COSMIC observations. Although Figure 2 shows that the ratio of COSMIC-TEC to Ground-TEC exhibits both temporal and spatial variations, it is evident that there is a bit of randomness associated with the data. This calls for a more expanded investigation into the relationship between the two data sets.

3.3. Neural Network Procedure for the AfriTEC

Neural networks basically consist of an input layer, a hidden layer, and an output layer. Each of the layers contains one or more neurons. The input layer neurons are parameters that are relevant for learning patterns in the parameter(s) to be modeled. The parameter(s) to be modeled constitute the output layer neuron(s). The number of hidden layer neurons is determined using a performance criterion discussed in section 3.3.4. For the final neural network used to develop the AfriTEC model, we have used seven input layer neurons, 32 hidden layer neurons, and one output layer neuron. These are detailed in the sections that follow.

3.3.1. Neural Network Inputs

To decide data that will constitute the neural network input layer neurons, we first identified the ionospheric variation components as follows: (i) temporal components, which have to do with variations in the ionosphere over time, and (ii) spatial components, which have to do with variations in the ionosphere over space.

The temporal components were further classified as diurnal and seasonal. We introduced the LTs of the days as inputs to enable the neural networks learn diurnal variations in the TEC data. There is need to clarify that

the data used in this work were obtained in universal time (UT) reference and were converted to LTs using the relation in equation (1).

$$LT = UT + Long/15, \tag{1}$$

where LT is the number of hours that have elapsed from the local midnight, UT is the number of hours that have elapsed from the Greenwich midnight, and $Long$ is the longitude (in degrees) of the location for which the LT is computed. $Long$ is measured in reference to the Greenwich meridian, such that places to the east of the meridian have positive values and places to the west of it have negative values. We also clarify here that, for the purpose of universality, the final model developed in this work does also accept UT inputs from users. The UTs are then converted internally to LTs by the model software application before being supplied to the neural network.

We also introduced the Julian number of days from the start of the years as inputs to enable the neural networks learn seasonal variations in the TEC data. To allow for a numerically continuous trend in the data, the diurnal and seasonal variations, represented by day of the year (DOY) and hour of the day (HH), were each split into two cyclical components as defined by equations (2) and (3) (Habarulema, 2010; Poole & McKinnell, 2000; Williscroft & Poole, 1996).

$$DOY_s = \sin\left(\frac{2\pi \times DOY}{365.25}\right) \qquad DOY_c = \cos\left(\frac{2\pi \times DOY}{365.25}\right), \tag{2}$$

$$HH_s = \sin\left(\frac{2\pi \times HH}{24}\right) \qquad HH_c = \cos\left(\frac{2\pi \times HH}{24}\right), \tag{3}$$

where DOY_s, DOY_c, HH_s, and HH_c are respectively the sine and cosine components of the DOY and HH.

The geographic longitudes and latitudes (which characterize the Earth locations in space) were also introduced as inputs to enable the neural networks learn spatial variations. Following the need to also learn and model variations that are associated with the solar activity levels, we introduced a final input layer neuron for the networks. For this application, we considered the following three solar activity indicators: SSN, F10.7, and UV flux. The F10.7 was adopted (following a performance criterion explained later) for the final version of the model developed in this work.

Summarily, a total of seven input neurons were used in the final neural network model developed in this work. They include DOY_s, DOY_c, HH_s, HH_c, $Longitude$, $Latitude$, and $F10.7$.

3.3.2. Training Procedure

The model in this work was developed by the method of training neural networks. The Bayesian regularization back-propagation algorithm (Bayes, 1763; Burden & Winkler, 2008; Laplace, 1812; MacKay, 1992) was used for training of the networks. The Bayesian regularization back-propagation algorithm has been shown to be robust and to give better results when compared to standard back-propagation training algorithms (Aggarwal et al., 2005; Burden & Winkler, 2008). Implementation of the neural network training was carried out using MATLAB's trainbr function and other functions in MATLAB's neural network toolbox.

MATLAB's tansig function (which implements the hyperbolic tangent) was used as the transfer function between the input layer and the hidden layer, as well as between the hidden and the output layer. The neural network equations are described by equations (4a) and (4b). There is common practice to use linear functions for the output layer of neural networks, especially in the training continuous variables. The VTEC parameter modeled in this work is a quantity that should not be negative (by physical definition). Using the hyperbolic tangent function for the output layer offered the neural networks the advantage to provide this boundary and even to provide lower errors for the neural network predictions. The hyperbolic tangent function has been described to be favorable in other similar applications (e.g., Ghasemi et al., 2015; Javed et al., 2015; Okoh, Onwuneme, et al., 2018; Vasiliades et al., 2015).

$$H_m = \tanh(I_{wm} \times I_m + B_1), \tag{4a}$$

$$O_m = \tanh(H_{wm} \times H_m + B_2). \tag{4b}$$

Equation (4a) is the equation connecting the input layer matrix (I_m) to the hidden layer matrix (H_m), while equation (4b) connects the hidden layer matrix to the output layer matrix (O_m). The input layer matrix is the

matrix containing inputs for the neural network, the hidden layer matrix contains intermediary values computed within the hidden layer, and the output layer matrix contains outputs from the neural network. I_{wm} and H_{wm} are respectively weight matrices for the input layer and the hidden layer, while B_1 and B_2 are respectively bias vectors for the input layer and the hidden layer. The input weight matrices and the bias vectors contain constants for a given trained neural network. After training, it is the weight matrices and the bias vectors that characterize the trained neural network; they do not change unless the neural network is retrained. For the final neural network model developed in this work, the weight matrices and the bias vectors are as contained in the MATLAB neural network function file named net32.m, which is distributed alongside the final deployment of the AfriTEC model. On the other hand, the input layer matrix, the hidden layer matrix, and the output layer matrix are variable matrices; they depend on user choices/inputs. The user inputs are fed into the network as the input layer matrix. The input layer matrix is transformed to the hidden layer matrix using equation (4a), and the hidden layer matrix is further transformed to output layer matrix using equation (4b). The output layer matrix contains the desired neural network outputs that translate to TEC, in this case. Normalization of the training data was done using the "mapminmax" processing function, which is default for the MATLAB Bayesian regularization training algorithm used in this work. The "mapminmax" function is first applied to normalize the inputs into the range $[-1, 1]$, and then it is applied in reverse to scale the $[-1, 1]$ outputs into TEC values (see Mathworks, 2019).

We used one hidden layer in this work since it has been shown that including more than one hidden layer does not lead to much difference in the accuracy of results (Haykin, 1994), and a possible drawback in using multiple hidden layers is that they are more prone to fall in bad local minima (Liu et al., 2007; Okoh, Yusuf, et al., 2015).

3.3.3. Testing Procedure

Prior to training the neural networks, data used in this work were split into three categories: for training, validation, and testing purposes. First, all data from the $3° \times 3°$ cells that are central to each of the following three stations: FUNC (32.6°N, 16.9°W), BAKC (0.35°N, 29.9°E), and VACS (20.3°S, 57.5°E) as shown in Figure 1, were kept aside for testing purpose. These data were used to test the spatial performance of the trained networks. The three stations were selected such that they cut across different latitude and longitude bands. The number of data points from each of the three locations is 60,025, 18,328, and 68,501 respectively for the FUNC, BAKC, and VACS locations. This makes a total of 146,854 data points representing about 2% of the entire data. The rest of the data were then split sequentially in time series and according to the volume of data available for each of the $1° \times 1°$ cells illustrated in Figure 1. The earliest 70% of data belonging to each of the 1×1 cells were used for training, the next 15% were used for validation, and the last 15% were set aside for testing. The training data set is used to adjust the weights as described in equations (4a) and (4b), the validation data set is used to ensure that an increase in prediction accuracy on the training data set corresponds to an increase in accuracy for data outside of the training data set, and the testing data set is used to test the predictions of the final network.

The Bayesian regularization algorithm used in this work has its own form of validation check built in. The point of checking validation is to see if the error on the validation set gets better or worse as training progresses, and the idea is to stop training if the error gets worse. However, the Bayesian error is not just based on how well the model is performing on the data set, it is also based on how large the weights are; the larger the weights, the higher the error. So throughout the training, the Bayesian regularization algorithm does not let the network explore larger weights, even though larger weights may lead to the global minimum errors. For this purpose, the Bayesian regularization algorithm does not require a validation data set as is the case with other algorithms during the back-propagation iteration stages of training (Foresee & Martin, 1997; MacKay, 1992; Okoh, Seemala, et al., 2018). The validation data set was used in this work to decide an appropriate number of hidden layer neurons (as detailed in section 3.3.4) for the final neural network used for the AfriTEC model.

The root-mean-square error (RMSE), computed using the formula in equation (5), was used as criterion to measure the performance of the networks. The RMSE is a measure of the typical error obtained in using network predictions, that is, an indication of the typical difference between the network predictions and the TEC measurements. Another commonly used error metric that is also used in this work is the mean absolute error (MAE). As described in equation (6), the MAE is simply an average of the absolute prediction errors. It is a more natural computation of the average errors. A major difference between the two error metrics is that

the RMSE does systematically penalize larger errors, and this is a reason for which the RMSE is frequently used in evaluating model performances where large errors are undesirable.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (\mathrm{GPSobservation}_i - \mathrm{NNprediction}_i)^2}{n}}, \tag{5}$$

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} |\mathrm{GPSobservation}_i - \mathrm{NNprediction}_i|}{n}, \tag{6}$$

where $\mathrm{GPSobservation}_i$ are the GPS-TEC measurements corresponding to the neural network predictions ($\mathrm{NNprediction}_i$), and n is the number of the observation-prediction pair.

3.3.4. Deciding the Number of Hidden Layer Neurons

The number of hidden layer neurons used in this work was decided based on the performance criteria described here. We trained 40 different networks that varied only in the number of hidden layer neurons, starting from 1 to 40 in steps of 1. After each network is trained, it is used to predict the TEC values at instances of the validation data set. RMSEs and MAEs are then computed using equations (5) and (6), and the network that returned the least errors was chosen as the best network. This decision is based on the idea that the best performing network should give TEC predictions that are closest to the GPS-TEC measurements, and as such, the prediction errors for the best performing network should be the minimum. Figure 3 shows results of how the RMSEs and MAEs varied with the number of hidden layer neurons. The figure shows that the least values of RMSE and MAE were both obtained using the network with 32 hidden layer neurons. The value of RMSE was 4.88 TECU, and the value of the MAE was 3.29 TECU. This neural network is the one adopted and used in this work for the development of the AfriTEC model.

3.3.5. Deciding the Solar Indices

Prior to deciding to use the F10.7 indices as solar activity input for the neural network, we had tested the performance of the following three solar activity indicators: SSN, F10.7, and UV flux. Three different sets of neural networks were trained in which we used the following inputs for all the three sets: DOY_s, DOY_c, HH_s, HH_c, *Longitude*, and *Latitude*. The difference between the sets was that we introduced one of SSN, F10.7, and UV flux as additional inputs for the three sets. After training, we used the resulting networks from each of the three sets to predict the TEC values for the validation data set. On computing the RMSEs and MAEs, we obtained the results summarized in Figure 3.

The figure shows that the neural network set in which we used F10.7 inputs most frequently gave the least RMSEs and MAEs. Although the differences in performance between the networks trained using the F10.7 and UV are not significant, we chose to use the network trained using the F10.7 mainly because it gave the least set of errors. Some research (e.g., Maruyama, 2010) demonstrated that the UV flux was a predominantly better solar proxy for empirical TEC models. It is however also observed in the same work of Maruyama (2010) that there were cases where the F10.7 gave the best results. We emphasize that our reason to use the F10.7 is based on the observation that it gave the least RMSE and MAE for the data set used in this work. Some authors (e.g., Huang et al., 2009; Antony & Shanthi, 2015; and in the NeQuick model) have also shown that the F10.7 is a good proxy for solar activity studies in the upper atmosphere.

3.3.6. Neural Network Architecture

The neural network used for the development of the AfriTEC model is graphically illustrated in Figure 4. The network has seven input layer neurons, 32 hidden layer neurons, and one output layer neuron.

The required user inputs are Year, Day of Year (or Month, and Day of Month), Hour of Day (in UT or LT), Longitude, and Latitude. The program uses these user inputs to obtain the neural network input neurons as illustrated in Figure 4. The Year and Day of Year inputs are used to obtain the corresponding F10.7 values. As at the time of writing up this research, if the user inputs a date later than 30 September 2018, the program forecasts the F10.7 value using the Hybrid Regression-Neural Network method (Okoh, Seemala, et al., 2018); otherwise, the program fetches the appropriate F10.7 value from an accompanying solar index file included in the folder of the AfriTEC model. The Day of Year input is used to compute the values of DOY_s and DOY_c using equation (2), and the Hour of Day input is used to compute the values of HH_s and HH_c using equation (3). The neurons in one layer of the network connect with neurons in consecutive layers using the appropriate weight matrices and bias vectors as described by the transfer functions in equations (4a) and (4b).

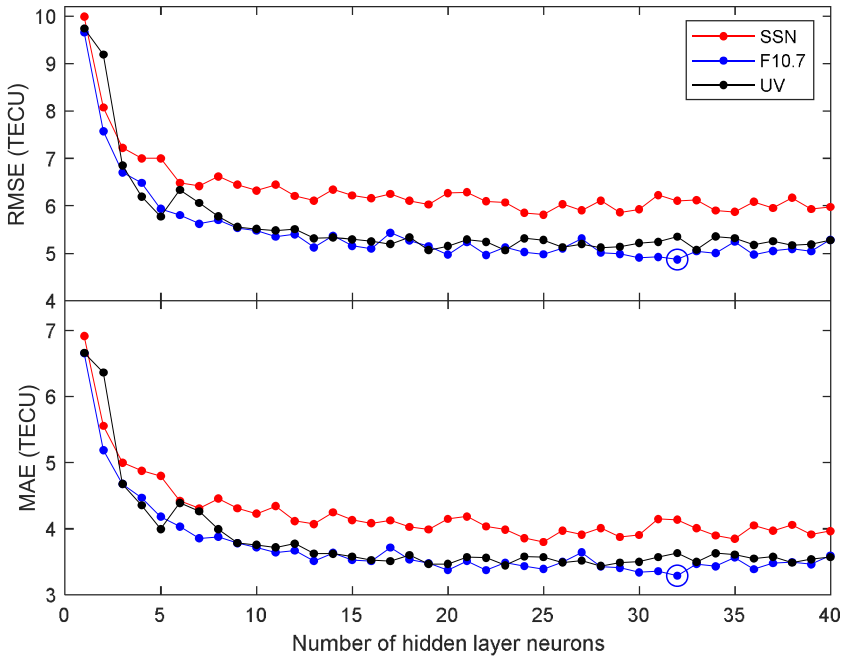

Figure 3. RMSEs and MAEs corresponding to different number of hidden layer neurons for 40 different networks trained in this work and for the three different sets of neural networks that differed in the solar activity index used as input.

4. Results and Discussions

4.1. AfriTEC Availability

Since the AfriTEC model is developed using MATLAB and for real-world applications, the model is made available for research, education, and other noncommercial applications in the following web pages:

1. Website of the Centre for Atmospheric Research (http://carnasrda.com/tec_models)
2. MATLAB Central website (https://www.mathworks.com/matlabcentral/fileexchange/69257-african-gnss-tec-afritec-model?s_tid=prof_contriblnk)

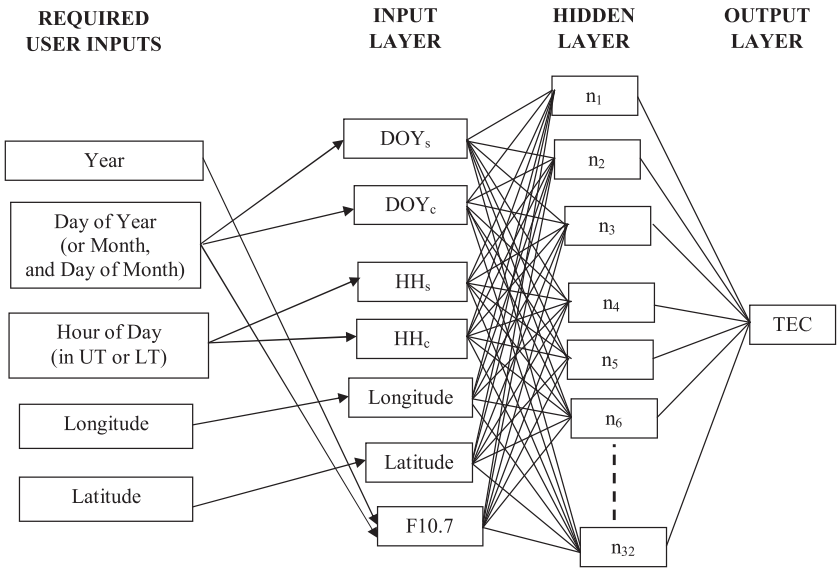

Figure 4. Graphical representation of neural network used in developing the AfriTEC model.

For user friendliness, a graphical user interface is developed. Detailed information on how to use the AfriTEC model is given on the model web pages and on a Microsoft word document (named AfriTEC ReadMe) included in the AfriTEC model package.

The user interface allows users to select whether they want to generate diurnal or spatial profiles. Diurnal profiles provide TEC information from the model for a given location over the entire period of a given day. Spatial profiles provide TEC information for the entire region (longitudes −25° to 60° and latitudes −40° to 40°) for a given day and time. The interface also allows users to generate, in one run, the diurnal profiles for a given location throughout the entire days in a given year. This is done by checking the "For Entire Year" checkbox on the user interface. Users can also make advanced settings that have to do with resolutions of the time and space parameters using the "Advanced settings" radio button on the user interface. In the program default settings, the diurnal profiles are generated in time resolutions of 1 hr, while the spatial maps are generated in longitude and latitude resolutions of 0.5° each. It is these values of resolution that are used for all graphical illustrations of results presented in this paper. Since the longitude and latitude inputs may constitute of any finite values within the model boundaries (longitudes −25 to 60° and latitudes −40 to 40°), the VTEC at any given African location may be computed. This is possible because the neural network training gives it capacity to interpolate/extrapolate over data used for the training, and this is why there is the capacity to allow users to define the spatial and temporal resolutions using any finite value.

4.2. Calibrated COSMIC Observations

Results of the calibrated COSMIC observations are shown in this section as a way of evaluating the calibration process. Figure 5 shows how the calibrated COSMIC observations compare with corresponding ground-based observations. Figure 5a is a representation for all data sets used for the calibration process. Figures 5b to 5d are respectively representations for the training, validation, and test data sets.

The values of RMSE and correlation coefficient (r) between the calibrated COSMIC-TECs and observed Ground-TECs are inscribed in the figures. For the case of using all data sets (Figure 5a), the value of correlation coefficient is 0.9712, which is greater than the 0.9267 value obtained prior to the calibration. The root-mean-square value of the differences is 2.78 TECU, which is about 44% lower than the 4.97 TECU obtained prior to the calibration. Both scenarios show that the calibration process was an effective way of unifying observations from the two sources. A comparison of the RMSEs and correlation coefficients obtained for Figures 5b to 5d show that the values are not significantly different for the training, validation, and test data sets. This is an indication that the neural network model used for the calibration process generalizes well; otherwise, it would perform significantly better on the training data set than on the validation data set and even worse on the test data set. The attribute of good generalization is required for neural networks to ensure that the networks will perform well on data that were not used for their training.

4.3. Tests and Accuracies for the AfriTEC Model

This section is devoted to present results that indicate the accuracies of the final neural network that was used to develop the AfriTEC model. Figure 6 is a plot of the TEC observations versus corresponding AfriTEC predictions for the test data sets used in this work. Figure 6a represents the scenario for the 15% test data set, while Figures 6b to 6d respectively represent the scenarios for the test data sets obtained from the 3° × 3° regions around the FUNC, BAKC, and VACS stations.

The values of correlation coefficient (r), inscribed in the figures for each of the test data sets, are generally higher than 0.85, showing that there is good correlation between the observations and corresponding predictions. The RMSE values are also lower than 4 TECU except for the scenario of the BAKC region, which is about 6 TECU. We attribute the higher value of RMSE obtained for the BAKC region to be due to the reason that the region is in the equatorial zone. The higher value of RMSE (and slightly lower correlation coefficient) obtained for the region implies that it is more difficult to model the region than the other two regions that are in the low latitudes and midlatitudes. The equatorial ionosphere is known to be more difficult to model owing to the large TEC gradients and peculiar phenomena like the equatorial anomaly (Hunt et al., 2000; Okoh, Onwuneme, et al., 2018; Uwamahoro et al., 2019). This is also evident from the detailed error analysis presented in Figure 7a. Figure 6e reveals an approximately Gaussian distribution of the errors, with maximum errors reaching about 40 TECU on one side of the distribution and about 60 TECU on the other side of the distribution. Statistics of the errors however show that 99.32% of the errors are within the

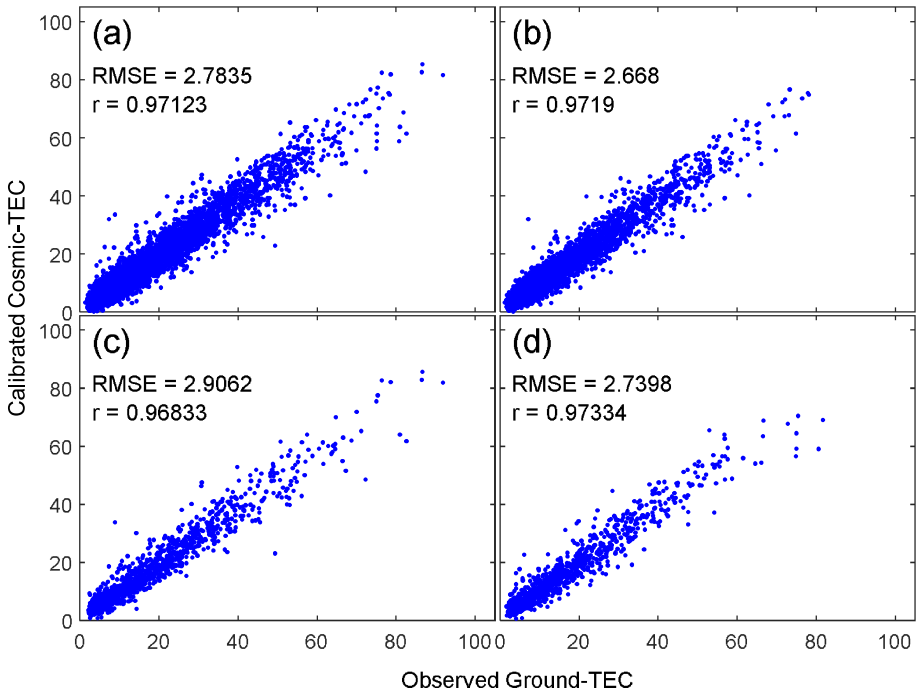

Figure 5. Plots of calibrated COSMIC-TECs versus corresponding observed Ground-TECs illustrated using (a) all data sets, (b) training data set, (c) validation data set, and (d) test data set.

range of ±10 TECU. An investigation of the large errors (especially those exceeding 30 TECU) reveal that they are predominantly from locations (e.g., latitude 4.5°N and longitude 17.5°N) where there are no ground-based GPS receivers, and so, the associated data are from the COSMIC observations. These errors exist mostly due to the wider areas covered by the RO electron density profiles from which COSMIC-TEC is integrated. Investigations of the large errors also show that they are predominantly associated with conditions of enhanced TEC observations, for example, the local middays of March equinox and high solar activity years. As can be seen in Figure 7, these conditions systemically correspond to conditions of high TEC errors.

In Figure 7, we present a detailed measure of the magnitudes of errors obtainable in using the AfriTEC model under varying conditions. Figures 7a to 7d respectively present the RMSEs computed (using all test data sets) for varying conditions of latitudes, solar activity levels, LTs, and seasons. Figure 7a shows that the RMSE values are as low as about 3 TECU at the midlatitudes and that the RMSEs can get to about 6 TECU at the equatorial zone. Figure 7b shows that the errors mostly increase as the solar activity levels increase. The RMSEs are as low as 3 TECU for low solar activity conditions (corresponding to F10.7 values of about 65 solar flux units) and as high as about 10 TECU for high solar activity conditions reaching F10.7 values of 250 solar flux units. In Figure 7c, the RMSEs are lower (about 1.5–2.0 TECU) during the nights and early mornings, and the values reach about 5.5 TECU during the day times. This shows a diurnal pattern that is in tandem with augmentation of the electrodynamics of the ionosphere in consistency with solar activity. Figure 7d shows that the RMSEs get to about 5 TECU during the equinoxes and drop to about 3.2 TECU during the June solstice. The typical scenario presented by Figure 7 is that the prediction errors are larger for conditions corresponding to larger values of TEC.

4.4. Diurnal Profiles From the AfriTEC, IRI-Plas, and NeQuick

Sample diurnal profiles of predictions from the AfriTEC, IRI-Plas, and NeQuick models are presented in this section. The default settings for the IRI-Plas and NeQuick models were used, and the TEC integration heights were set to the GPS satellite altitude of 20,200 km. The predictions are presented alongside corresponding GPS observations from the three test stations of FUNC, BAKC, and VACS. For each station, data for the 3 years of 2009, 2012, and 2014 were used to represent years of low, moderate, and high solar activity.

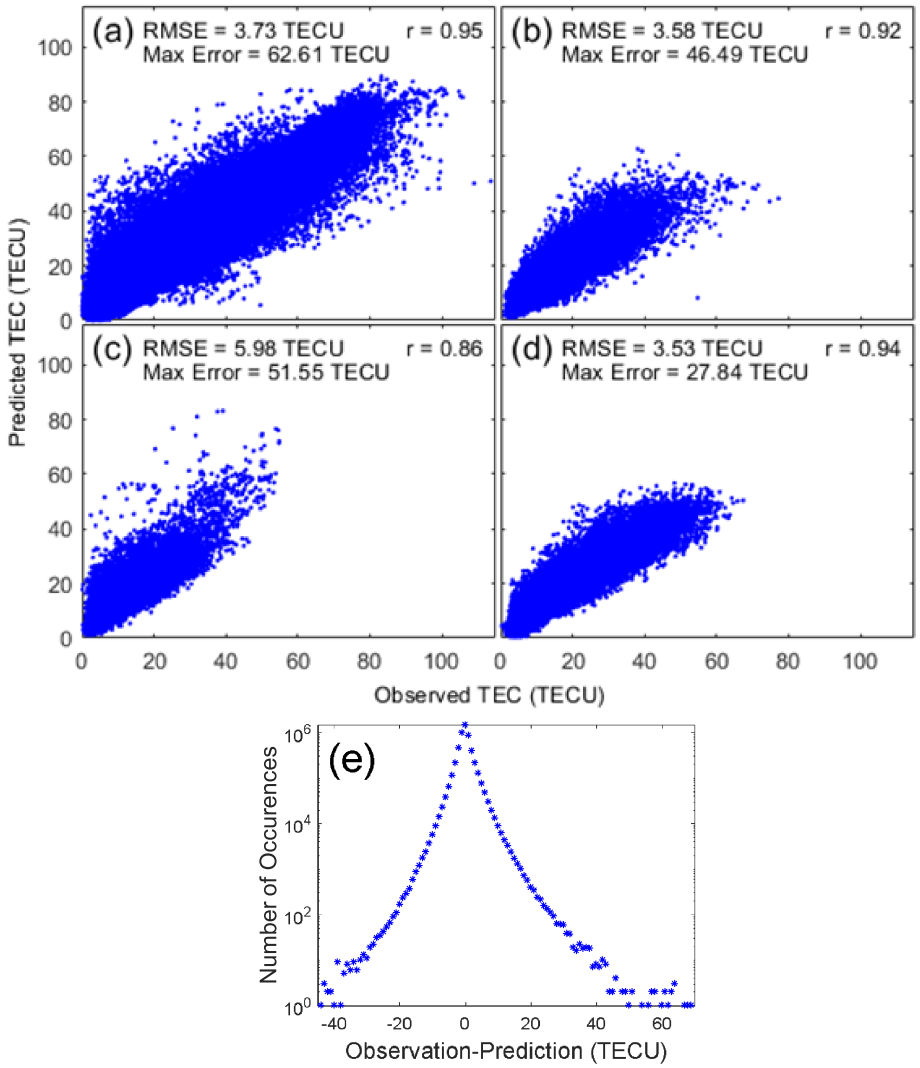

Figure 6. Observed TEC values versus corresponding AfriTEC predictions for (a) the 15% test data set used in this work and the test data sets obtained from the 3° × 3° regions around the stations of (b) FUNC, (c) BAKC, and (d) VACS. The distribution of errors for the training data set is represented in (e) using a logarithmic scale for the *y* axis.

In each year, data for the particular days of the March equinox, June solstice, September equinox, and December solstice were also used to represent TEC information for the equinox and solstice seasons. If there were no available data for any of the particular days, then data from the closest day were used. Figure 8 illustrates the results for year 2009. Plots in the first column of Figure 8 are for the FUNC station, while the second and third columns respectively contain plots of the BAKC and VACS stations. Plots in the four rows respectively represent the plots for the days of the March equinox, June solstice, September equinox, and December solstice. For year 2009, the March equinox was day number 79 of the year, while the June solstice, September equinox, and December solstice were respectively day numbers 172, 265, and 355 of the year. The actual days illustrated are shown at the top left corner for each of the plots. Figure 9 is included to illustrate that the neural network used to develop the AfriTEC model was able to capture the occurrence of the winter anomaly during the high solar activity year of 2014 at the FUNC station. The FUNC station is in the Northern Hemisphere; therefore, the months of June and December are respectively summer and winter months for the station. Going by the solar-zenith angle dependence of ionospheric ionization, TEC values are expected to be greater for this station in June (summer) than in December (winter). It is however observed that the TEC values were greater in December than in June. This is the winter anomaly that occurs as a result of seasonal changes in the density ratio of atomic oxygen to molecular nitrogen (Rishbeth, 1998). The result is that the summer ion loss rate becomes

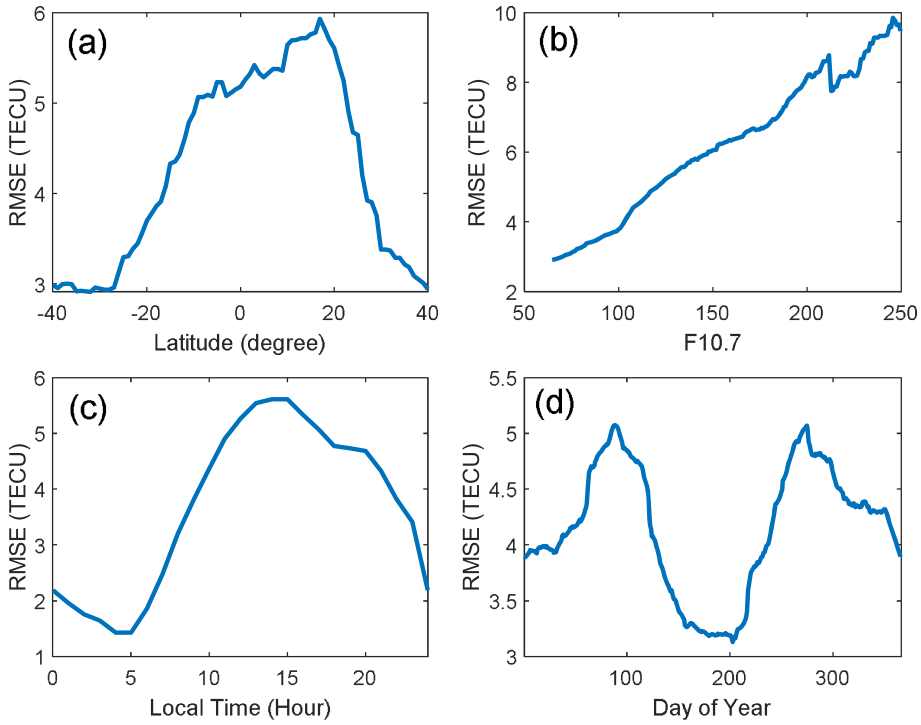

Figure 7. RMSE variations for predictions of the AfriTEC model using the test data set under conditions of varying (a) latitudes, (b) F10.7 values, (c) local times, and (d) days of the years.

higher, leading to reduced ion densities during the summer season. The higher F10.7 value of the December date (when compared to the June date) is also a reason why the December TEC values are greater than the June TEC values.

Results of correlation coefficient values computed between the model predictions of the profiles and the corresponding observations show that the correlation coefficients are mostly greater than 0.8 for all three models, an indication that the model predictions follow similar trends as the corresponding observations. This is also evident from Figures 8 and 9. The means of the correlation coefficient values are respectively 0.94, 0.93, and 0.94 for the IRI-Plas, NeQuick, and AfriTEC models. Figures 8 and 9 also show that TEC predictions from the AfriTEC model are relatively closer to the GPS values when compared to predictions from the NeQuick and IRI-Plas models. The average of the RMSEs computed between the model predictions and the corresponding GPS observations are respectively 10.44, 6.65, and 3.76 TECU for the IRI-Plas, NeQuick, and AfriTEC. The corresponding MAEs are respectively 9.26, 5.73, and 2.71 TECU. The observed minimal errors for the AfriTEC model are expected especially since the model focus is restricted to the African region, and its development is based on huge data resource from within the region. Also, the IRI-Plas and NeQuick models are based on other data sets than the GNSS TEC. These are reasons why their prediction errors may not be as low as those of the AfriTEC. Data used for the development of the AfriTEC were tailored in the direction of GNSS observations since the major motivation is to develop a high-accuracy regional model of TEC for ground-based GNSS applications across Africa. It is the first report of ionospheric study over the African region using this volume of data over space and time. The project also avails opportunity to investigate the African ionosphere at spatial resolution of about 1°. We conduct one of such investigations and present the results in the next section.

CODE GIM data for the exact same days as in Figures 8 and 9 were obtained, and the data were linearly interpolated to obtain data for the exact same three stations represented in Figures 8 and 9. The average of CODE GIM RMS errors (provided by CODE) for the days and stations represented in Figures 8 and 9 was computed, and the value is 3.91 TECU, which is slightly greater than the 3.76 value computed for AfriTEC. Using the RMS values provided by CODE eliminates the bias that would have been introduced by computing RMS errors with respect to the GPS-TEC observations used in this work.

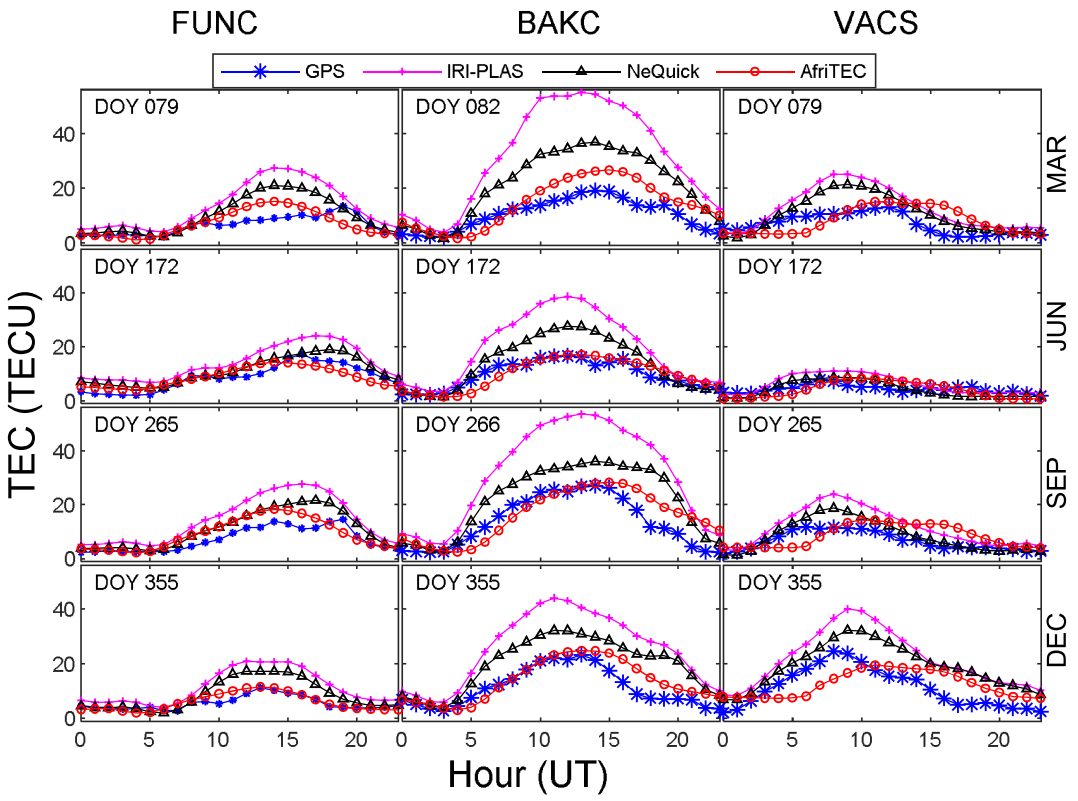

Figure 8. Diurnal TEC profiles for FUNC, BAKC, and VACS stations for the days of the March equinox, June solstice, September equinox, and December solstice, year 2009. In case of nonavailability of GPS-TEC data for any of the particular days, data from the closest day were used. The specific days of year (DOY) used are shown at the top left corner of the plots.

4.5. Distance Between Equatorial Anomaly Crests Over the African Region

In an earlier version of AfriTEC, the phenomena of equatorial anomaly was not conspicuous mainly due to the reason that data used did not provide good spatial coverage of the equatorial anomaly region. Only observations from ground-based GPS receivers were used, and so, the equatorial anomaly region (especially within and around the Sahara desert) was not well captured in the observations. In this version, the problem was surmounted by incorporating observations from the COSMIC mission.

Figure 10 illustrates spatial simulations of TEC from the AfriTEC model for the four different seasons of year 2012 (a year of moderate solar activity). The figures are exactly for 13:00 UT of day numbers 79, 172, 266, and 356 of the year, respectively, representing days in the seasons of March equinox, June solstice, September equinox, and December solstice.

In Figure 10, the equatorial anomaly is conspicuously observed for the days of both equinoxes and for the December solstice day but not for the June solstice day. Although we obtained few instances of reduced equatorial anomaly occurrences during days of the June solstice season, repeated simulations show that the June solstice season is not a commonplace for occurrence of the equatorial anomaly. This can be explained to be due to the seasonal variability of the drivers of the equatorial plasma fountain, which are the duo of electric field and ionization density. The eastward electric fields at the magnetic equator move the F region plasma to higher altitudes by ExB drift. The plasma then returns downward and poleward along the magnetic field line, forming the equatorial ionization anomaly. During the June solstice, the equatorial vertical drift that drives the fountain is almost absent (Scherliess & Fejer, 1999).

We further investigated variations of typical distances between the equatorial anomaly crests over the region. For this study, particular reference was given to the separation of the anomaly crests as measured along the meridian of longitude 20°E, which is somewhat central for the region. For each instance of the

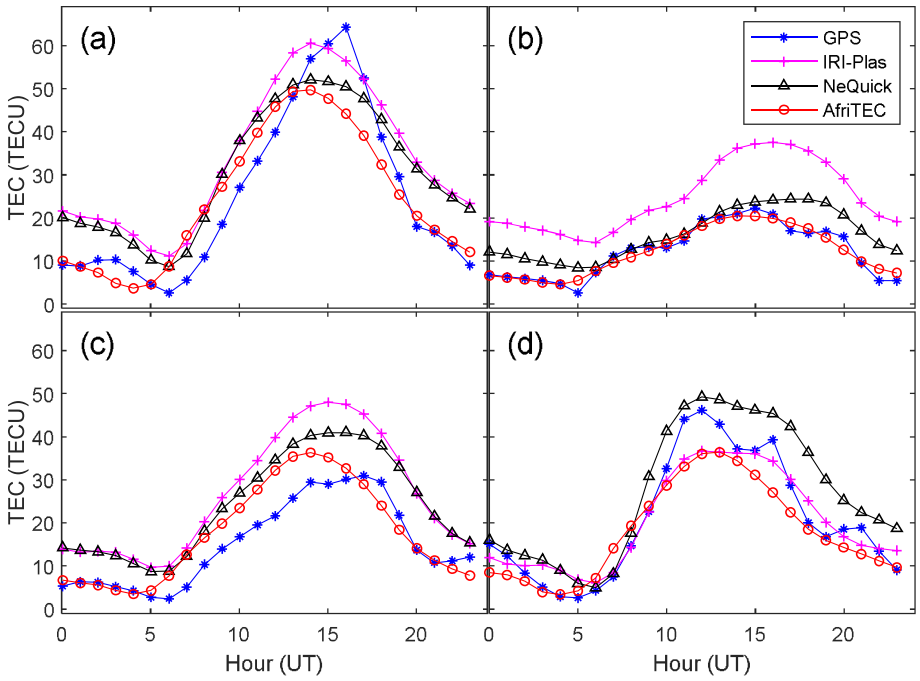

Figure 9. Diurnal TEC profiles for FUNC station for day numbers (a) 79 (F10.7 = 150), (b) 172 (F10.7 = 105), (c) 266 (F10.7 = 139), and (d) 356 (F10.7 = 173), of year 2014, which are respectively the days of the March equinox, June solstice, September equinox, and December solstice for the year.

simulations, we first decide the latitudes of the two ionization peaks and that of the sandwiched trough by using the vector representing the TEC values along longitude 20°E. The distance of separation between the peaks is then computed, and the process is repeated for the 2 days before and after the day. The means of the observations are reported in Table 1. The table reports parameters of the anomaly for two scenarios. The first scenario is according to seasons of year 2012 as reported in Figure 10, and the second scenario is according to years of varying solar activity as reported in Figures 11b to 11d. In Figure 11a, we use a sample TEC profile for longitude 20°E to illustrate our determinations of the anomaly crest and trough locations. The illustrated profile is for the March equinox day of year 2012. Figures 11b to 11d are spatial simulations of TEC from the AfriTEC model for 13:00 UT of day number 79 for years 2009, 2012, and 2014, respectively, illustrating the equatorial anomaly occurrence during years of low, moderate, and high solar activity.

Table 1 shows that the distance of separation between the anomaly crests typically varies from about 11.5 ± 1.0° to 16.0 ± 1.0°. The table reveals a scenario in which the distance of separation between the crests tends to widen as the solar activity levels increase. Yeh et al. (2001) engaged more than 350 tomographically reconstructed ionospheric images to investigate the motion of the equatorial anomaly crests. Their results indicate that the crest-trough distance is a function of the plasma vertical drift that varies with the dynamics of the ionosphere. As a further support, results from the work of Raghavarao et al. (1988) strongly suggested that the root cause of the crest motion is the equatorial dynamo electric field that depends on solar activity. Lin et al. (2007) affirmed that two basic processes that affect the anomaly formation significantly are the strength of the equatorial plasma fountain and thermospheric neutral winds (Abdu, 2001; Balan et al., 1997; Balan & Bailey, 1995; Lin et al., 2005; Rishbeth, 2000; Su et al., 1997). The solar activity therefore influences the separation of the anomaly crests since it modulates drivers of the strength of the equatorial plasma fountain and the thermospheric winds. During the years of high solar activity, there is increased production of ionized matter in the ionosphere. This leads to an increase in ionospheric electron densities and consequently an increase in the magnitude of the ExB force that drives the anomaly. The ExB force is responsible for the vertical movement of plasma, and the magnitude of the force determines how high the particles are lifted. Habarulema et al. (2019), for example, showed that the vertical drifts are proportional to the maximum electron densities in the E region. After the upward ExB lifting of the particles, they return by a downward diffusion along the magnetic field lines. The Earth's magnetic field lines are oriented such that the particles that

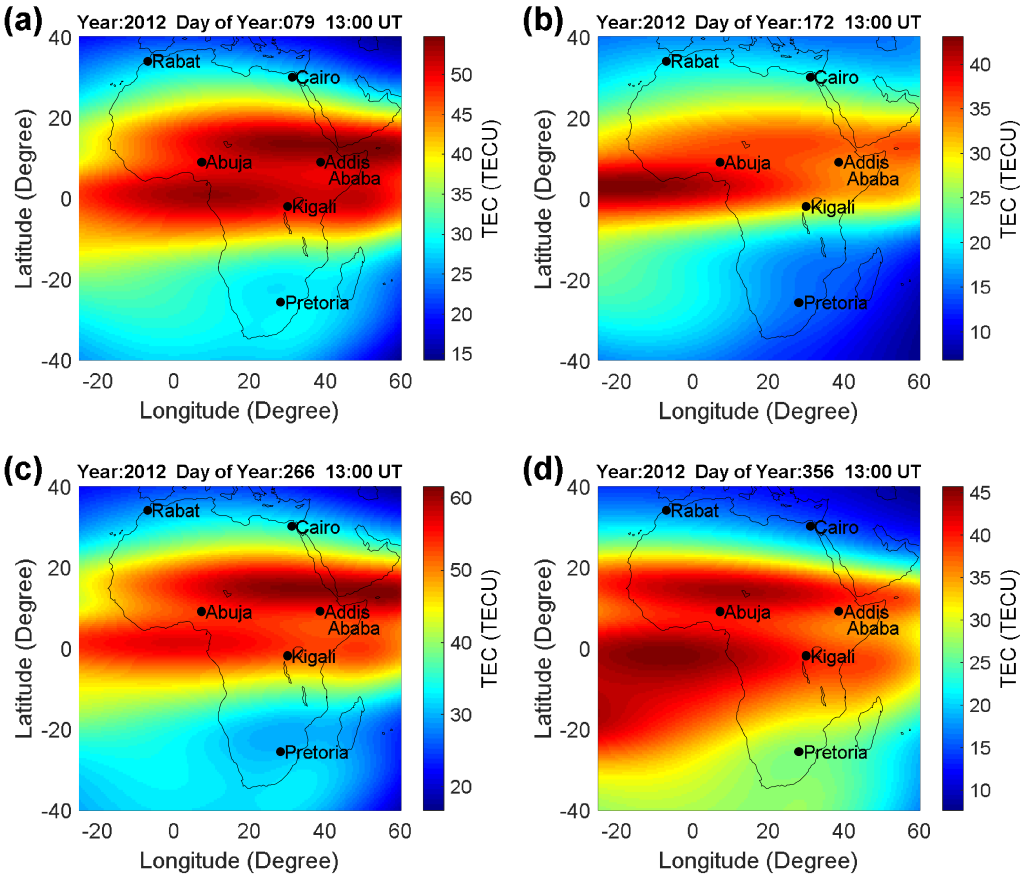

Figure 10. Spatial simulations of TEC from the AfriTEC model for 13:00 UT of day numbers (a) 79 (F10.7 = 101), (b) 172 (F10.7 = 107), (c) 266 (F10.7 = 125), and (d) 356 (F10.7 = 111), of year 2012.

are lifted to higher altitudes diffuse farther away from the geomagnetic equator; the particles are more laterally displaced as they return downwards along the magnetic field lines. In essence, the ionization peaks are formed at greater distances from the geomagnetic equator. This explains why we observe greater values of latitudinal separations of the anomaly crests during the years of higher solar activities.

Table 1 also reveals that the anomaly region is shifted southwards during the December solstice; the trough is shifted by about 1.5°, and the southern crest is shifted by over 2.5°. The December solstice is summer for the Southern Hemisphere, and this explains why the anomaly region is shifted south of the equinox

Table 1
Latitudinal Locations of the Equatorial Anomaly Crests and Troughs Measured Along the Meridian of Longitude 20°E

Season/Year	Latitude of northern crest (±0.5°)	Latitude of trough (±0.5°)	Latitude of southern crest (±0.5°)	Latitudinal separation between crests (±1.0°)
Classification by seasons of year 2012				
March equinox (F10.7 = 101)	13.5	7.0	0.5	13.0
September equinox (F10.7 = 125)	15.0	6.5	1.0	14.0
December solstice (F10.7 = 111)	13.5	5.5	−2.5	16.0
Classification by years of varying solar activity during March equinoxes				
2009 (F10.7 = 68)	12.0	6.0	0.5	11.5
2012 (F10.7 = 101)	13.5	7.0	0.5	13.0
2014 (F10.7 = 150)	15.5	7.5	0.5	15.0

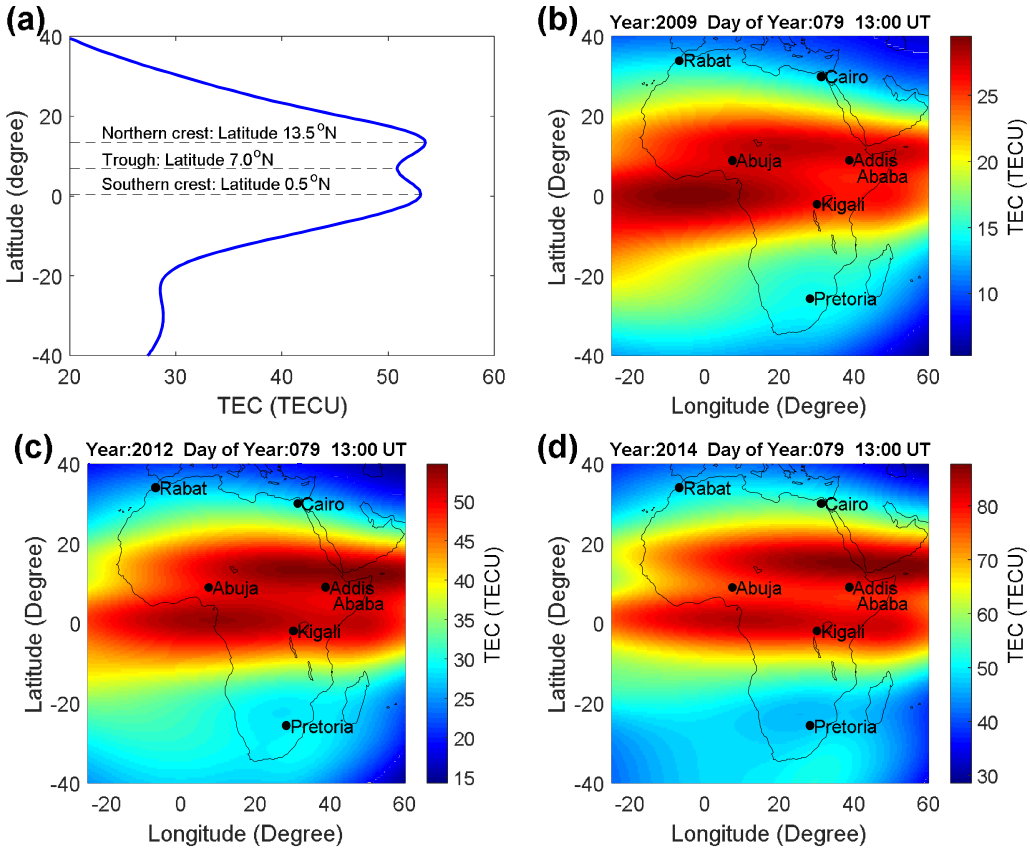

Figure 11. (a) Sample TEC profile for longitude 20°E illustrating the determination of anomaly crest and trough locations. The illustrated profile is for the March equinox day of year 2012. (b) to (d) are spatial simulations of TEC from the AfriTEC model for 13:00 UT of day number 79 of years 2009, 2012, and 2014, respectively. The F10.7 values are respectively 68, 101, and 150.

positions. The southward shifting of the anomaly region during the December solstice is expected, but this is the first report on the magnitudes of the shifting over the African region using empirical observations with high spatial resolution.

5. Conclusion

The development of a TEC model over the African region (known as AfriTEC model) using the method of neural networks is reported. TEC measurements are obtained from 269 ground-based GPS receivers as well as from the COSMIC mission for various parts of the African region. The observations span the periods from years 2000 to 2017 and represent the most extended data sets that have been used to study the African ionosphere. The 1° spatial resolution used also offers opportunity to study the African ionosphere in greater details. Data from other satellite navigation systems (such as GLONASS, GALILEO, and BEIDOU) will be included in a future improvement of the model as this will significantly increase the data coverage.

The three solar activity indicators of SSN, F10.7, and UV flux were tested for effectiveness as neural network inputs in learning TEC variations associated with the solar activity. Results from the study revealed that the F10.7 and UV flux were more effective for the TEC modeling procedure. The F10.7 was used for the AfriTEC model as it often gave the least errors, and it particularly gave the least error for the 32 hidden layer neurons finally used.

Results of equatorial anomaly simulations using the AfriTEC show that there is reduced occurrence of the phenomenon during the June solstice season. The distance of separation between the anomaly crests is

typically in the range from about 11.5 ± 1.0° to 16.0 ± 1.0°. The separation widens as the solar activity level increases. The anomaly region shifts south of the equinox location during the December solstice.

Acknowledgments
This research was funded and supported by the CV Raman International Fellowship (DST/INT/CVRF/2016), the Centre for Atmospheric Research, Nigeria, and the Indian Institute for Geomagnetism. We also thank the Indian Institute of Geomagnetism for making their High Performance Computing (HPC) System available for this work. K. S. and Y. O. are supported by Japan Society for the Promotion of Science KAKENHI (15H05815 and 16H06286). We appreciate the developers of the IRI-Plas Model and the NeQuick for making their models available. We thank the providers of all the data used in this work for making their data available. Data used in this work were obtained from the African Geodetic Reference Frame (http://afrefdata.org), the Nigeria GNSS Reference Network (www.nignet.net), the South African network of continuously operating GNSS base stations (http://www.trignet.co.za), the University of California, San Diego, SOPAC, and CSRC GARNER GPS Archive (ftp://garner.ucsd.edu), the National Aeronautics and Space Administration's CDDIS Archive of GNSS products (ftp://cddis.gsfc.nasa.gov), the Global Data Center of the International GNSS Service (ftp://igs.ensg.ign.fr), the UNAVCO Archive of GNSS Data (ftp://data-out.unavco.org), the Geodetic Data Archiving Facility (ftp://geodaf.mt.asi.it), the COSMIC Data Analysis and Archive Centre (http://cosmic-io.cosmic.ucar.edu/cdaac/index.html), the WDC-SILSO Royal Observatory of Belgium-Brussels (website: http://www.sidc.be/silso/datafiles), the National Oceanic and Atmospheric Administration (ftp://ftp.ngdc.noaa.gov/STP/space-weather/solar-data/solar-features/solar-radio/noontime-flux/penticton/penticton_observed/listings/), the University of South California Dornsife (https://dornsifecms.usc.edu/space-sciences-center/download-sem-data), NASA's OMNIWeb (https://omniweb.sci.gsfc.nasa.gov/form/dx1.html), the Ionosphere Radiopropagation Unit of the T/ICT4D Laboratory, International Centre for Theoretical Physics (https://t-ict4d.ictp.it/nequick2/source-code), the Crustal Dynamics Data Information System (ftp://cddis.gsfc.nasa.gov/gnss/products/ionex/), the IZMIRAN Institute (http://ftp.izmiran.ru/pub/izmiran/SPIM/), and the Center for Orbit Determination in Europe (ftp://ftp.aiub.unibe.ch/CODE/). Thanks to Mr. Richard Isheke of the Michael and Cecilia Ibru University for making time to severally test the AfriTEC model. We thank the *JGR* editor and reviewers of

References

Abdu, M. A. (2001). Outstanding problems in the equatorial ionosphere–thermosphere electrodynamics relevant to spread F. *Journal of Atmospheric and Solar - Terrestrial Physics*, *63*(9), 869–884. https://doi.org/10.1016/S1364-6826(00)00201-7

Adewale, A. O., Oyeyemi, E. O., Adeniyi, J. O., Adeloye, A. B., & Oladipo, A. O. (2011). Comparison of total electron content predicted using the IRI-2007 model with GPS observations over Lagos, Nigeria. *Indian Journal of Radio & Space Physics*, *40*, 21–25.

Aggarwal, K. K., Singh, Y., Pravin, C., & Puri, M. (2005). Bayesian regularization in a neural network model to estimate lines of code using function points. *Journal of Computer Science*, *1*(4), 505–509. https://doi.org/10.3844/jcssp.2005.505.509

Aggarwal, M. (2011). TEC variability near northern EIA crest and comparison with IRI model. *Advances in Space Research*, *48*(7), 1221–1231. https://doi.org/10.1016/j.asr.2011.05.037

Aggarwal, M., & Sharma, D. K. (2017). Seasonal characteristics of COSMIC measurements over Indian sub-continent during different phases of solar activity. *Advances in Space Research*, *59*(9), 2279–2294. https://doi.org/10.1016/j.asr.2017.02.018

Akala, A. O., Seemala, G. K., Doherty, P. H., Valladares, C. E., Carrano, C. S., Espinoza, J., & Oluyo, S. (2013). Comparison of equatorial GPSTEC observations over an African station and an American station during the minimum and ascending phases of solar cycle 24. *Annales de Geophysique*, *31*(11), 2085–2096. https://doi.org/10.5194/angeo-31-2085-2013

Akala, A. O., Somoye, E. O., Adewale, A. O., Ojutalayo, E. W., Karia, S. P., Idolor, R. O., et al. (2015). Comparison of GPSTEC observations over Addis Ababa with IRI-2012 model predictions during 2010–2013. *Advances in Space Research*, *56*(8), 1686–1698. https://doi.org/10.1016/j.asr.2015.07.017

Antony, D. T. S., & Shanthi, G. (2015). The study of solar activity in relation with high frequency variations of solar radio flux. *American Journal of Astronomy and Astrophysics*, *3*(6), 87–92. https://doi.org/10.11648/j.ajaa.20150306.11

Baboo, S. S., & Shereef, K. I. (2010). An efficient weather forecasting system using artificial neural network. *International Journal of Environmental Science and Technology*, *1*(4), 321–326.

Balan, N., & Bailey, G. J. (1995). Equatorial plasma fountain and its effects: Possibility of an additional layer. *Journal of Geophysical Research*, *100*(A11), 21,421–21,432. https://doi.org/10.1029/95JA01555

Balan, N., Bailey, G. J., Abdu, M. A., Oyama, K. I., Richards, P. G., MacDougall, J., & Batista, I. S. (1997). Equatorial plasma fountain and its effects over three locations: Evidence for an additional layer, the F3 layer. *Journal of Geophysical Research*, *102*(A2), 2047–2056. https://doi.org/10.1029/95JA02639

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions*, *53*, 370–418. https://doi.org/10.1098/rstl.1763.0053

Bilitza, D. (2001). International reference ionosphere 2000. *Radio Science*, *36*(2), 261–275. https://doi.org/10.1029/2000RS002432

Bilitza, D., & Reinisch, B. W. (2008). International reference ionosphere 2007: Improvements and new parameters. *Advances in Space Research*, *42*(4), 599–609. https://doi.org/10.1016/j.asr.2007.07.048

Burden, F., & Winkler, D. (2008). Bayesian regularization of neural networks. In D. J. Livingstone (Ed.), *Artificial neural networks. Methods in molecular biology* (Vol. 458, pp. 23–42). New York City: Humana Press. https://doi.org/10.1007/978-1-60327-101-1_3

CDDIS (2019). Index of/gnss/products/ionex/. ftp://cddis.gsfc.nasa.gov/gnss/products/ionex/

Chen-Joe, F. (2008). Constellation Deployment for the FORMOSAT-3/COSMIC Mission. *IEEE Transactions on Geoscience and Remote Sensing*, *46*(11), 3367–3379.

Foresee, F. D., & Martin, T. H. (1997). Gauss-Newton approximation to Bayesian learning. *Proceedings of the International Joint Conference on Neural Networks*.

Ghasemi, S. A., Hofstetter, A., Saha, S., & Goedecker, S. (2015). Interatomic potentials for ionic systems with density functional accuracy based on charge densities obtained by a neural network. *Physical Review B*, *92*, 045131. https://doi.org/10.1103/PhysRevB.92.045131

Gulyaeva, T. L., Huang, X., & Reinisch, B. W. (2002). Plasmaspheric extension of topside electron density profiles. *Advances in Space Research*, *29*(6), 825–831. https://doi.org/10.1016/S0273-1177(02)00038-8

Habarulema, J. B. (2010). A contribution to TEC modelling over Southern Africa using GPS data. *PhD Thesis,* Rhodes University.

Habarulema, J. B., Dubazane, M., Katamzi-Joseph, Z., Uwamahoro, J. C., Moldwin, M. B., & Yizengaw, E. (2019). Progress towards data improvisation for low latitude vertical drifts modeling. *IRI 2019 Workshop, Frederick University, Nicosia, Cyprus*, 13 September, 2019.

Habarulema, J. B., McKinnell, L. A., & Cilliers, P. J. (2007). Prediction of Global Positioning System total electron content using neural networks over South Africa. *Journal of Atmospheric and Solar - Terrestrial Physics*, *69*(15), 1842–1850. https://doi.org/10.1016/j.jastp.2007.09.002

Habarulema, J. B., McKinnell, L. A., Cilliers, P. J., & Opperman, B. D. L. (2009). Application of neural networks to South African GPS TEC modeling. *Advances in Space Research*, *43*(11), 1711–1720. https://doi.org/10.1016/j.asr.2008.08.020

Haykin, S. (1994). *Neural networks, a comprehensive foundation*. New York-USA: Macmillan College Publishing Company.

Hochegger, G., Nava, B., Radicella, S. M., & Leitinger, R. (2000). A family of ionospheric models for different uses. *Physics and Chemistry of the Earth*, *25*(4), 307–310. https://doi.org/10.1016/S1464-1917(00)00022-2

Huang, C., Liu, D., & Wang, J. (2009). Forecast daily indices of solar activity, F10.7, using support vector regression method. *Research in Astronomy and Astrophysics*, *9*(6), 694–702. https://doi.org/10.1088/1674-4527/9/6/008

Hunt, S. M., Close, S., Coster, A. J., Stevens, E., Schuett, L. M., & Vardaro, A. (2000). Equatorial Atmospheric and Ionospheric Modeling at Kwajalein Missile Range. *Lincoln Laboratory Journal*, *12*(1), 45–64.

Javed, S., Murthy, Y., Baig, R., & Rao, P. (2015). Development of ANN model for prediction of performance and emission characteristics of hydrogen dual fueled diesel engine with Jatropha Methyl Ester biodiesel blends. *Journal of Natural Gas Science and Engineering*, *26*, 549–557. https://doi.org/10.1016/j.jngse.2015.06.041

Jee, G., Lee, H. B., Kim, Y. H., Chung, J. K., & Cho, J. (2010). Assessment of GPS global ionosphere maps by comparison between CODE GIM and TOPEX/Jason TEC data: Ionospheric perspective. *Journal of Geophysical Research*, *115*, A10319. https://doi.org/10.1029/2010JA015432

Jin, S. G., Cho, J., & Park, J. (2007). Ionospheric slab thickness and its seasonal variations observed by GPS. *Journal of Atmospheric and Solar - Terrestrial Physics*, *69*(15), 1864–1870. https://doi.org/10.1016/j.jastp.2007.07.008

Jin, S. G., & Najibi, N. (2014). Sensing snow height and surface temperature variations in Greenland from GPS reflected signals. *Advances in Space Research*, *53*(11), 1623–1633. https://doi.org/10.1016/j.asr.2014.03.005

our manuscript for their constructive comments/suggestions, which have helped to improve this paper.

Jin, S. G., Park, J., Wang, J., Choi, B., & Park, P. (2006). Electron density profiles derived from ground-based GPS observations. *Journal of Navigation*, *59*(3), 395–401. https://doi.org/10.1017/S0373463306003821

Klobuchar, J. A. (1996). Ionospheric effects on GPS. In B. W. Parkinson & J. J. Spilker (Eds.), *Global Positioning System: Theory and applications 2*, *Progress in Astronautics and Aeronautics* (Chap. 12, pp. 485–515). Washington, DC: American Institute of Aeronautics and Astronautics.

Langley, R. (2002). Mapping the low-latitude ionosphere with GPS. *GPS World*, *13*(2), 41–47.

Laplace, P. S. (1812). *Theorieanalytique des probabilities*. Paris: Ve. Courcier.

Lin, C. H., Liu, J. Y., Fang, T. W., Chang, P. Y., Tsai, H. F., Chen, C. H., & Hsiao, C. C. (2007). Motions of the equatorial ionization anomaly crests imaged by FORMOSAT-3/COSMIC. *Geophysical Research Letters*, *34*, L19101. https://doi.org/10.1029/2007GL030741

Lin, C. H., Richmond, A. D., Heelis, R. A., Bailey, G. J., Lu, G., Liu, J. Y., et al. (2005). Theoretical study of the low- and midlatitude ionospheric electron density enhancement during the October 2003 superstorm: Relative importance of the neutral wind and the electric field. *Journal of Geophysical Research*, *110*, A12312. https://doi.org/10.1029/2005JA011304

Liu, Y., Janusz, A., & Starzyk, Z. (2007). Optimizing number of hidden neurons in neural networks. *Proceedings of the 25th IASTED International Multi-Conference: Artificial Intelligence and Applications* (pp. 121–126). Innsbruck, Austria: ACTA Press Anaheim, CA, USA.

Liu, Z., Li, Y., Li, F., & Guo, J. (2018). Near real-time PPP-based monitoring of the ionosphere using dual-frequency GPS/BDS/Galileo data. *Advances in Space Research*, *61*(6), 1435–1443. https://doi.org/10.1016/j.asr.2017.12.038

MacKay, D. (1992). Bayesian interpolation. *Neural Computation*, *4*(3), 415–447. https://doi.org/10.1162/neco.1992.4.3.415

Mannucci, A. J., Wilson, B. D., & Edwards, C. D. (1993). A new method for monitoring the Earth's ionospheric total electron content using the GPS global network. *Proceedings of the 6th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GPS 1993)* (pp. 1323–1332). Salt Lake City, UT.

Maruyama, T. (2010). Solar proxies pertaining to empirical ionospheric total electron content models. *Journal of Geophysical Research*, *115*(A4). https://doi.org/10.1029/2009ja014890

Mathworks (2019). Configure neural network inputs and outputs. Retrieved from https://in.mathworks.com/help/nnet/ug/configure-neural-network-inputs-and-outputs.html

McKinnell, L. A. (2002). A neural network based ionospheric model for the bottomside electron density profile over Grahamstown, South Africa. *PhD Thesis,* Rhodes University.

Nava, B., Coïsson, P., & Radicella, S. M. (2008). A new version of the NeQuick ionosphere electron density model. *Journal of Atmospheric and Solar - Terrestrial Physics*, *70*(15), 1856–1862. https://doi.org/10.1016/j.jastp.2008.01.015

Okoh, D. (2016). *Computer neural networks on MATLAB*. North Charleston, SC, USA: Createspace. ISBN-13: 978-1539360957

Okoh, D., McKinnel, L. A., Cilliers, P., Okere, B., Okonkwo, C., & Rabiu, B. (2015). IRI-vTEC versus GPS-vTEC for Nigerian SCINDA GPS stations. *Advances in Space Research*, *55*(8), 1941–1947. https://doi.org/10.1016/j.asr.2014.06.037

Okoh, D., Onwuneme, S., Seemala, G., Jin, S., Rabiu, B., Nava, B., & Uwamahoro, J. (2018). Assessment of the NeQuick-2 and IRI-Plas 2017 models using global and long-term GNSS measurements. *Journal of Atmospheric and Solar-Terrestrial Physics*, *170*, 1–10. https://doi.org/10.1016/j.jastp.2018.02.006

Okoh, D., Owolabi, O., Ekechukwu, C., Folarin, O., Arhiwo, G., Agbo, J., et al. (2016). A regional GNSS-VTEC model over Nigeria using neural networks: A novel approach. *Geodesy & Geodynamics*, *7*(1), 19–31. https://doi.org/10.1016/j.geog.2016.03.003

Okoh, D., Yusuf, N., Adedoja, O., Musa, I., & Rabiu, B. (2015). Preliminary results of temperature modelling in Nigeria using neural networks. *Weather*, *70*(12), 336–343. https://doi.org/10.1002/wea.2559

Okoh, D. I., Seemala, G. K., Rabiu, A. B., Uwamahoro, J., Habarulema, J. B., & Aggarwal, M. (2018). A Hybrid Regression-Neural Network (HR-NN) method for forecasting the solar activity. *Space Weather*, *16*, 1424–1436. https://doi.org/10.1029/2018SW001907

Olwendo, O. J., Baki, P., Mito, C., & Doherty, P. (2012). Characterization of ionospheric GPS total electron content (GPS-TEC) in low latitude zone over the Kenyan region during a very low solar activity phase. *Journal of Atmospheric and Solar - Terrestrial Physics*, *84-85*, 52–61. https://doi.org/10.1016/j.jastp.2012.06.003

Otsuka, Y., Ogawa, T., Saito, A., Tsugawa, T., Fukao, S., & Miyazaki, S. (2002). A new technique for mapping of total electron content using GPS network in Japan. *Earth, Planets and Space*, *54*(1), 63–70. https://doi.org/10.1186/BF03352422

Oyeyemi, E., Akala, A., Bolaji, S., Adewale, A., Olugbon, B., & Amaechi, P. (2018). On the response of equatorial ionospheric anomaly during geomagnetic storm events and comparisons with IRI model predictions. *42nd COSPAR Scientific Assembly,* Held 14-22 July 2018, in Pasadena, California, USA.

Perez, R. O. (2017). Ionospheric error contribution to GNSS single-frequency navigation at the 2014 solar maximum. *Journal of Geodesy*, *91*(4), 397–407. https://doi.org/10.1007/s00190-016-0971-0

Poole, A. W. V., & McKinnell, L. A. (2000). On the predictability of f_0F_2 using neural networks. *Radio Science*, *35*(1), 225–234. https://doi.org/10.1029/1999RS900105

Rabiu, A. B., Adewale, A. O., Abdulrahim, R. B., & Oyeyemi, E. O. (2014). TEC derived from some GPS stations in Nigeria and comparison with the IRI and NeQuick models. *Advances in Space Research*, *53*(9), 1290–1303. https://doi.org/10.1016/j.asr.2014.02.009

Rabiu, A. B., Groves, K., Abdulrahim, R. B., Fayose, R. S., Adeniyi, J. O., Ariyibi, E. A., et al. (2011). TEC derived from some GPS stations in Nigeria and comparison with the IRI. *United Nations International Meeting on the Applications of GNSS*, 12–16, December, Vienna, Austria

Radicella, S. M., & Leitinger, R. (2001). The evolution of the DGR approach to model electron density profiles. *Advances in Space Research*, *27*, 35–40. https://doi.org/10.1016/S0273-1177(00)00138-1

Raghavarao, R., Sridharan, R., Sastri, J. H., Agashe, V. V., Rao, B. C. N., Rao, P. B., & Somayajulu, V. V. (1988). The equatorial ionosphere. In C. H. Liu & B. Edwards (Eds.), *WITS handbook*, *World ionosphere/thermosphere study* (Vol. 1, pp. 48–93). Urbana, Illinois, USA: SCOSTEP Secretariat, University of Illinois.

Rama Rao, P. R., Niranjan, K., Prasad, D., Krishna, S. G., & Uma, G. (2006). On the validity of the ionospheric pierce point IPP altitude of 350 km in the Indian equatorial and low-latitude sector. *Annales Geophysicae*, *24*(8), 2159–2168. https://doi.org/10.5194/angeo-24-2159-2006

Rishbeth, H. (2000). The equatorial F-layer: Progress and puzzles. *Annales de Geophysique*, *18*(7), 730–739. https://doi.org/10.1007/s00585-000-0730-6

Rishbeth, H. (1998). How the thermospheric circulation affects the ionosphere. *Journal of Atmospheric and Solar-Terrestrial Physics*, *60*, 1385–1402.

Scherliess, L., & Fejer, B. G. (1999). Radar and satellite global equatorial F region vertical drift model. *Journal of Geophysical Research*, *104*(A4), 6829–6842. https://doi.org/10.1029/1999JA900025

Seemala, G. K., & Valladares, C. E. (2011). Statistics of total electron content depletions observed over the South American continent for the year 2008. *Radio Science*, *46*, RS5019. https://doi.org/10.1029/2011RS004722

Sharma, D. K., Aggarwal, M., & Bardhan, A. (2017). Variability of ionospheric parameters during solar minimum and maximum activity and assessment of IRI model. *Advances in Space Research*, *60*(2), 435–443. https://doi.org/10.1016/j.asr.2016.11.027

Su, Y. Z., Bailey, G. J., Oyama, K. I., & Balan, N. (1997). A modelling study of the longitudinal variations in the north-south asymmetries of the ionospheric equatorial anomaly. *Journal of Atmospheric and Solar - Terrestrial Physics*, *59*(11), 1299–1310. https://doi.org/10.1016/S1364-6826(96)00016-8

Syndergaard, S., Schreiner, W. S., Rocken, C., Hunt, D. C., & Dymond, K. F. (2006). Preparing forcosmic: Inversion and analysis of ionospheric data products. In U. Foelsche, G. Kirchengast, & A. K. Steiner (Eds.), *Atmosphere and Climate: Studies by Occultation Methods* (pp. 137–146). New York: Springer.

Uwamahoro, J. C., Habarulema, J. B., & Buresova, D. (2019). Highlights about the performances of storm-time TEC modeling techniques for low/equatorial and mid-latitude locations. *Advances in Space Research*, *63*(10), 3102–3118. https://doi.org/10.1016/j.asr.2019.01.027

Vasiliades, L., Galiatsatou, P., & Loukas, A. (2015). Nonstationary frequency analysis of annual maximum rainfall using climate covariates. *Water Resources Management*, *29*(2), 339–358. https://doi.org/10.1007/s11269-014-0761-5

Williscroft, L. A., & Poole, A. W. V. (1996). Neural networks, f_0F_2, sunspot number and magnetic activity. *Geophysical Research Letters*, *23*(24), 3659–3662. https://doi.org/10.1029/96GL03472

Yeh, K. C., Franke, S. J., Andreeva, E. S., & Kunitsyn, V. E. (2001). An investigation of motions of the equatorial anomaly crest. *Geophysical Research Letters*, *28*(24), 4517–4520. https://doi.org/10.1029/2001GL013897