

## INVESTIGATION OF THE RELATIONSHIP BETWEEN $f_0F_2$ AND THE TEC USING SINGLE STATION NEURAL NETWORK MODELS

\*Otugo, V., Okujagu, C. U., and Onwuneme, S.

Department of Physics, University of Port Harcourt, Port Harcourt, Nigeria  
 Correspondence Email: vivian.akaogu@ust.edu.ng

Received: 20-10-2021

Accepted: 12-11-2021

### ABSTRACT

*The relationship between the ionospheric critical plasma frequencies ( $f_0F_2$ ) and GNSS-TEC (Global Navigation Satellite System – Total Electron Content) measurements was investigated using an artificial neural network method. About 20 pairs of ionosonde-GNSS receiver stations from 2000 to 2016 were used. Results from this work indicate that the relationship between  $f_0F_2$  and TEC is mostly affected by the seasons, followed by the level of solar activity, and then the local time. Geomagnetic activity was the least significant of the factors investigated. The relationship between  $f_0F_2$  and TEC was also shown to exhibit spatial variation; the variation is less conspicuous for closely located stations. Single station models predicted the  $f_0F_2$  more accurately at their particular localities and clearly overestimated values of the  $f_0F_2$  ionosonde observations when used at different localities. This finding indicates that model predictions are better (in terms of reduced prediction errors) for the stations for which they are developed than for a different station. Our result visibly point out that models developed for a particular station cannot be effectively applied in another station located farther apart in space. The new approach described in this study represents an important contribution in space weather prediction.*

**Keywords:**  $f_0F_2$ , TEC, Ionosonde, GNSS, Ionosphere, Neural network

### INTRODUCTION

The ionosphere is not a stable but a dynamic system that is intensely disrupted by various events such as; neutral winds, atmospheric tides, motion due to electric and magnetic fields and auroral particles at higher latitudes (Cander, 1998, Wintoft and Cander, 2000). Ionospheric variability is predominant in the F layer of the ionosphere and the complexity of this layer poses a major constraint to the efficiency of terrestrial and earth-space communications (Stamper et al; 2004, Oyeyemi et al; 2005). The efficiency of high frequency radio communication depends on the ability of one making a precise prediction of the ionospheric conditions (Oyeyemi et al;

2005). The critical frequency of the F2 layer ( $f_0F_2$ ) is an important parameter in studying the ionosphere and in radio communication. Equally; the Total electron content (TEC) of the ionosphere is also another significant descriptive parameter in ionospheric studies (Fayose et al., 2012). TEC values are used in describing the F2 region at various locations even during magnetically disturbed periods. Any operating system that encompasses radio wave propagation through ionosphere needs TEC quantities; since TEC affects code signal transmission as well as the phase signal transmission observed in radio communication. Over the years, researchers have discovered that these two parameters;  $f_0F_2$  and TEC are

correlated (Spalla and Ciruolo, 1994; Kouris et al, 2004). And as such, recent studies adopt the use of TEC measurements for the approximation of  $f_0F_2$  values (Otugo et al; 2019). A standard formula has not been established as per the correlation between  $f_0F_2$  and TEC in different seasons and different levels of solar activity; hence, studies are still carried out in this aspect to discover ore.

In this study, we adopted a new approach; by using Artificial Neural Network to define the connection between TEC and  $f_0F_2$ . Neural network models trained with sufficient data offers a practical technique of modelling ionospheric dynamics since: they are sufficiently fast for real time processing and are able to describe non-linear occurrence actively involved in the ionospheric variations (Wintoft and Cander, 2000, Mckinnell and Poole; 2004, Habarulema et al.,2007, Athieno et al.,2017). We used  $f_0F_2$  measurements from ionosonde stations and the corresponding TEC values from closely located GNSS receiver stations to develop a single station neural network models for different locations. Each of the single station models were used to predict  $f_0F_2$  values in different locations. The performance of the models at different locations and seasons depicts the temporal and spatial connections between TEC and  $f_0F_2$ .

## **MATERIALS AND METHODS.**

### **Data sources and data pre-processing**

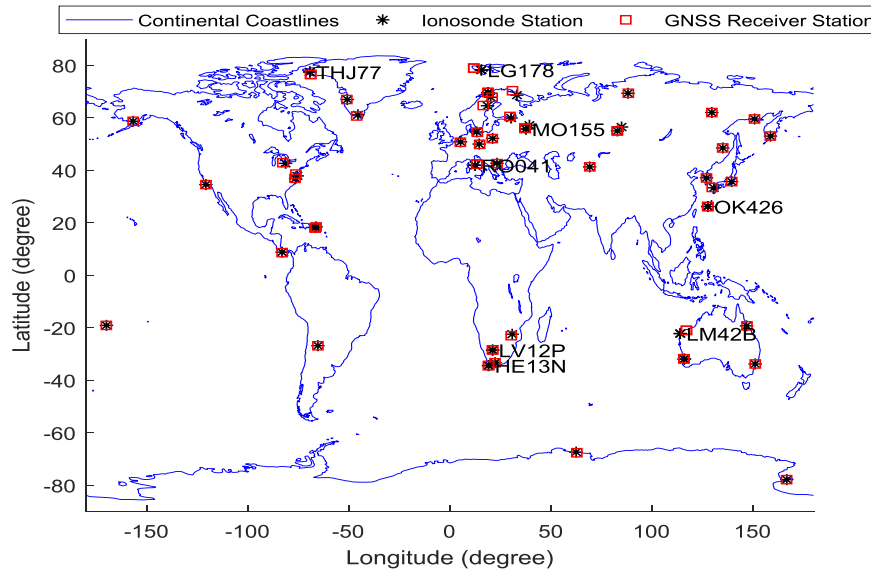
Ionosonde observations of  $f_0F_2$  values were obtained from the archives of the Space Physics Interactive Data Resources (SPIDR, <http://spidr.ngdc.noaa.gov/spidr>), and from the South African National Space Agency (SANS) Space Science. The

ionosonde  $f_0F_2$  data were averaged to get 1-hour values. GNSS data were obtained from data archives of the University of California, San Diego ([www.garner.uscd.edu](http://www.garner.uscd.edu)), and from the South African network of continuously operating GNSS base stations (TRIGNET, <http://www.trignet.co.za>). The VTEC data were obtained in RINEX (Receiver Independent Exchange) format. RINEX is a data interchange format for raw satellite system data which allows a user to post process the received data to generate a better result. RINEX gives information measured by 2 or more receivers in a differential mode. These RINEX data files were post processed into TEC files using a GPS-TEC analysis application software developed by Gopi Seemala (Seemala and Valladares, 2011) (available from <https://seemala.blogspot.com>). The software calculates the relative TEC by using both  $f_1$  and  $f_2$  and by removing errors due to clock biases and the tropospheric water vapour effect. To obtain 30-s instantaneous VTEC values for a receiver station, the average of all VTECs computed for satellites visible within the 30-s interval were computed. VTECs from satellites below 30-degree elevation angles were excluded from the computation so as to reduce multipath errors. To further reduce the volume of data, smoothen out data spikes, and obtain VTEC values that correspond to the hourly  $f_0F_2$  values, the 30-s VTEC data were averaged in 1-hour interval.

The GNSS receiver stations were selected based on the proximity to the corresponding ionosonde stations. The greatest separation between the GNSS receiver stations and corresponding ionosonde stations is less than 4 degrees of great circle. Since the state

of the ionosphere varies spatially, it is important to minimize the distance between the ionosondes and corresponding GNSS receivers so as to get minimal errors from

the modelling. The ionosonde stations and corresponding GNSS receiver stations used in this work are illustrated in Figure 1.



**Figure 1.** Locations of ionosonde and GNSS receiver stations.

We also obtained the hourly values of both the sunspot values and Dst (Disturbance storm Time) indices from the Space Physics Data Facility of the Goddard Space Flight Center, National Space Administration's omniweb (<https://omniweb.gsfc.nasa.gov/form/dx1.html>).

**Table 3.1: List of the ionosonde stations with their corresponding GPS stations**

Ionosonde Station			GPS Station		
Code	Latitude	Longitude	Code	Latitude	Longitude
LG178	78.2	15.6	NYA1	78.9	11.9
LM42B	-22.2	114.0	KARR	-21.0	117.1
PA836	34.6	-120.7	VNDP	34.6	-120.6
RO041	41.9	12.5	AQUI	42.4	13.4
THJ77	77.5	-69.2	THU2	76.5	-68.8
TR169	69.6	19.0	TRO1	69.7	18.9
TR170	69.6	19.0	TRO1	69.7	18.9
YA462	62.0	129.7	YAKT	62.0	129.7
YG431	33.4	130.7	G121	33.5	129.8
WP937	37.9	-75.4	HNPT	38.6	-76.1
SN437	37.1	127.1	OSN1	37.1	127.0
GR13L	-33.3	22.4	GRHM	-33.3	22.4
HE13N	-34.4	19.2	HNUS	-34.4	19.2
LV12P	-28.5	21.2	UPTA	-28.4	21.2

### Neural network training algorithm

Artificial Neural network ('neural network' for short) is an inter connected assembly of simple processing elements, units or neurons; which is based on the neural configuration of the brain (Gurney; 2009, Kohli et al; 2014). Neural network is a computational system designed to recognize patterns. The architectural design of a neural network consists of three layers; an input layer, a hidden layer and an output layer. Each of these layers is made up of nodes or neurons which are information processing units.

Neural networks can be trained using different algorithms. In this work, we used the Levenberg-Marquardt (LM) algorithm. The LM algorithm is designed to minimize the error functions that arise during neural network training. Prior to training, the entire available data was sequentially split in time into three sets: the earliest 70% was kept for training, the next 15% was kept for validation, and the latest 15% was kept for testing. The MATLAB implementation of the LM training algorithm was used in this work. This implementation works by performing different iterations, the validation dataset is used to check when there was no further progress in the generalization, and the training stops automatically when there is no further progress in the generalization.

The neural networks implemented in this work constitute an input layer, a hidden layer, and an output layer. Equations (2a) and 2(b) respectively contain the transfer functions used to transfer the input layer neurons to the hidden layer neurons and from the hidden layer neurons to the output layer neurons.

$$H_m = \tanh(I_{wm} \times I_m + B_1) \quad (2a)$$

$$O_m = \tanh(H_{wm} \times H_m + B_2) \quad (2b)$$

$I_m, \dots, H_m$ , and  $O_m$  are matrices which respectively contain the input layer neurons, the hidden layer neurons, and the output layer neurons.  $I_{wm}$  and  $H_{wm}$  are respectively the input and hidden layer weight matrices, while  $B_1$  and  $B_2$  are respective bias vectors for the input and hidden layers. At the start of training, the weight matrices and bias vectors are randomly initialized, and as training progresses, the training algorithm modifies the weight matrices and bias vectors in directions that lead to minimization of the sum of square errors. At the end of training, the resulting weight matrices and bias vectors contain optimized values that minimize the network prediction errors.

### Neural network training

We trained a system of single-station networks which have TEC as the only one input and  $f_oF2$  as the output. The number of hidden layer neurons was varied starting from 1 to 30 in steps of 1. For each of the 30 trained networks, we used the validation dataset to compute the root-mean-square deviations (RMSDs) of the network predictions from the ionosonde measurements using the formula in equation (3).

$$\text{RMSD} = \sqrt{\frac{\sum_{i=1}^n (X_i - P_i)^2}{n}} \quad (3)$$

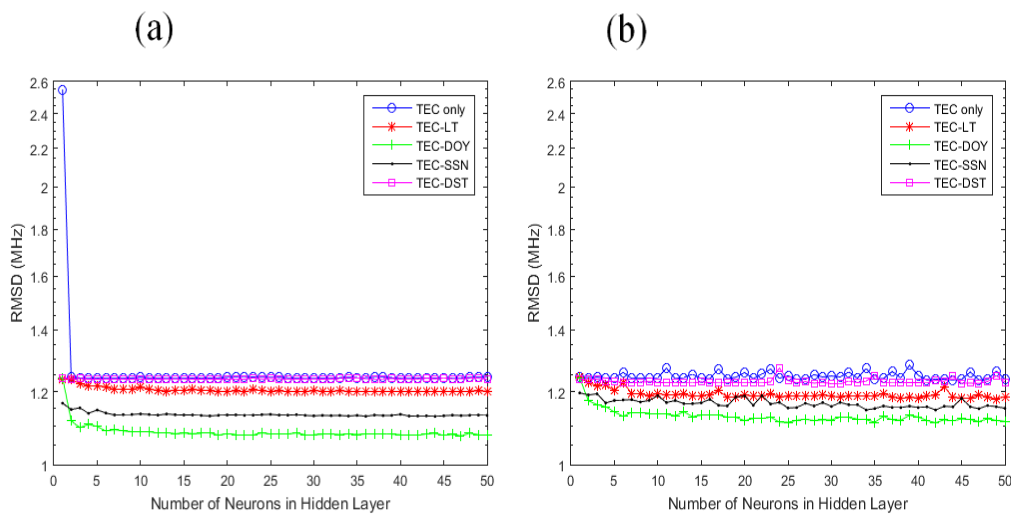
Where the  $X_i$ s and  $P_i$ s are respectively the ionosonde observations and the neural network predictions, and  $n$  is the total number of observation-prediction pairs. Results of the RMSD computations for the RO041 station are shown in Figure 2(a).

On the consideration that the relationship between  $f_0F_2$  and TEC may vary depending on local time, season, and on the levels of solar and geomagnetic activities, we introduced more input layer neurons to investigate such dependencies. The following four parameters were used to represent the factors: the number of hours from local midnight of each day (for local time), the Julian number of days from the start of each year (for seasons), the Sunspot number (SSN) (for level of solar activity), and the disturbance storm time (Dst) index (for level of geomagnetic activity).

We then trained four other systems of networks in which each of the four parameters were separately included as input layer neurons. As done for the system

of networks which had only TEC as input, we trained 30 networks for each of the four systems (with number of hidden layer neurons varied from 1 to 30 in steps of 1). The formula in equation (3) was used to compute RMSDs, and the results are as shown in Figure 2(a). The text of legends in figure 2 are as explained here: TEC only (for the networks which have TEC as the only input layer neuron), TEC-LT (for the networks which have TEC and local time as the input layer neurons), TEC-DOY (for the networks which have TEC and day of year as the input layer neurons), TEC-SSN (for the networks which have TEC and SSN as the input layer neurons), and TEC-DST (for the networks which have TEC and Dst as the input layer neurons).

## RESULTS AND DISCUSSION.



**Figure 2.** RMSDs for 5 different systems of networks trained with varying input layer neurons and varying number of hidden layer neurons (a) all available data used, (b) data used constitutes of 50% disturbed time data and 50% quiet time data. The vertical axes are plotted in logarithmic scale.

Figure 2(a) shows that the system of networks which gave the least prediction errors is that which has TEC and DOY as inputs. This is followed by the system which has TEC and SSN as inputs, and then the system with TEC and local time as inputs.

By considering that the most effective of the systems should be the one which minimized the prediction errors, we use minimization of prediction error as criteria to measure effectiveness of the input layer neurons. Our results therefore indicate that the

relationship between  $f_0F_2$  and TEC is mostly dependent on the seasons, followed by the level of solar activity, and then the local time. The dependence on geomagnetic activity was the least significant. Based on the understanding that the observed insignificance of geomagnetic activity may actually be a result of the fact that the data used for training and testing are predominantly quiet-time observations, we conducted a control experiment. For the control experiment, we first selected all observations that were recorded during geomagnetically disturbed periods (with  $Dst \leq -50$ ), and then randomly selected an equal amount of observations from the remaining data. In this way, we had as much disturbed-time data as the quiet-time data (that is, 50% disturbed-time data and 50% quiet time data). Using this dataset, we repeated the neural network training and testing procedure, and the results are illustrated in Figure 2(b). The result is consistent with that of Figure 2(a), and still shows that the relationship between  $f_0F_2$  and TEC is mostly dependent on the seasons, followed by the level of solar activity, and then the local time. The dependence on geomagnetic activity was

still insignificant. Therefore, the following four input parameters: TEC, day of year, SSN, and local time, were then used to train different neural network for different single stations so that each of the resulting networks is specifically designed for use at the particular locality of the ionosonde/GNSS receiver system. Furtherly, to investigate the space-dependence of the TEC and  $f_0F_2$  relation, we used the single-station models developed for a particular locality to predict the  $f_0F_2$  values for other localities as illustrated in Figures (2) to (5). Each of the figures consist of three diurnal plots: (i) The observed  $f_0F_2$  values for the particular ionosonde station X (denoted *obs-X* on the plot legend), (ii) The predicted  $f_0F_2$  values for station X using the single-station model developed for the same station X (denoted *net-X* on the plot legend), and (iii) The predicted  $f_0F_2$  values for station X using the single-station model developed for another station Y (denoted *net-Y* on the plot legend). The stations and their coordinates are presented in Table 1 while the RMSDs of each of the single station model performance are presented in Table 2.

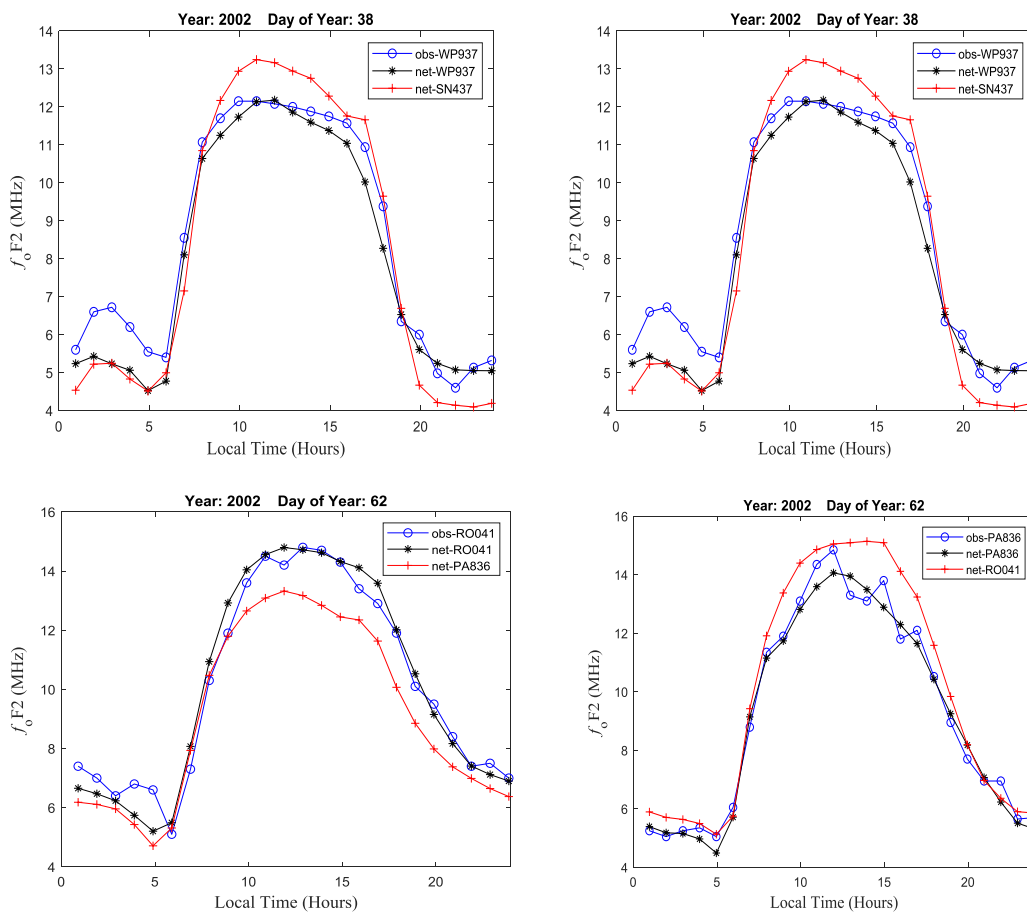
**Table 2: The RMSDs of the predictions of the single station models**

STN 1	STN 2	RMSD1A	RMSD 1B	RMSD 2A	RMSD 2B
<b>SN437</b>	<b>WP937</b>	0.66	1.34	0.66	1.11
<b>RO041</b>	<b>PA836</b>	0.45	1.32	0.58	1.53
<b>HE13N</b>	<b>LG178</b>	0.33	1.27	0.53	0.96
<b>LG178</b>	<b>RO041</b>	0.32	0.84	0.40	0.77
<b>YA462</b>	<b>YG431</b>	0.20	1.30	0.42	1.51
<b>LV12P</b>	<b>GR13L</b>	0.37	0.40	0.29	0.37
<b>TR169</b>	<b>TR170</b>	0.28	0.35	0.26	0.32
<b>LM42B</b>	<b>LG178</b>	0.34	2.22	0.52	1.08
<b>THJ77</b>	<b>PA836</b>	0.23	2.15	0.38	2.74

In table 2 above; RMSD1A represents the Root mean square deviation (RMSD) of the predicted  $f_0F_2$  values for station X using the single-station model developed for the same station X. RMSD1B is the RMSD of the predicted  $f_0F_2$  values for another station Y using the single-station model developed for station X. The inverse scenario is given

by RMSD2A which is the RMSD of the predicted  $f_0F_2$  values for station Y using the single-station model developed for the same station Y. RMSD2B is the RMSE of the predicted  $f_0F_2$  values for station X using the single-station model developed for another station Y

### Stations closely located in latitude but far apart in longitude



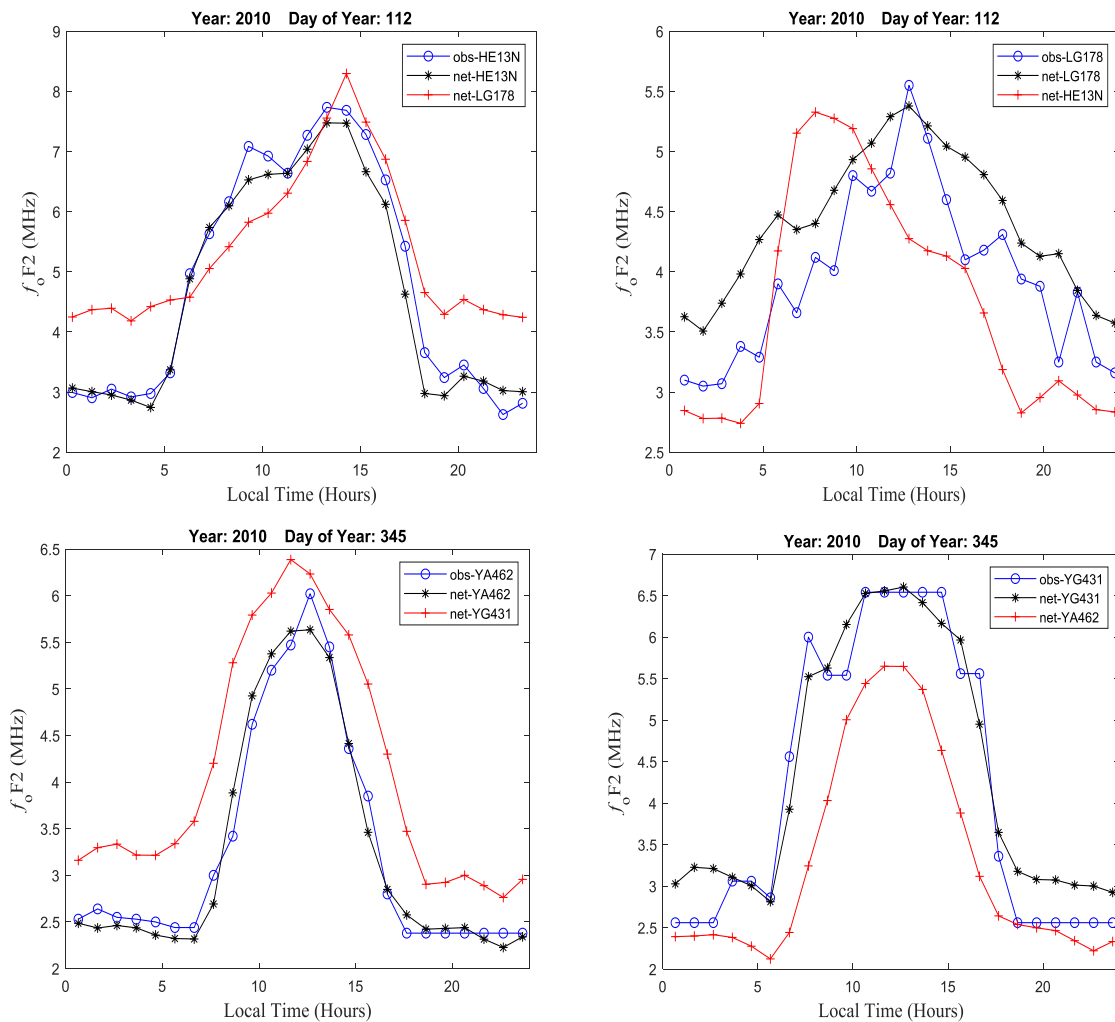
**Figure 2:**  $f_0F_2$  observations and predictions for similar latitude but different longitude stations

In this part, the performances of the models of different stations which are closely located in latitudes but are far apart longitudinally were examined. In Figure 2(left) the observations illustrated are for the WP937 station, and the networks trained for this station, as well as for the SN437station, were used to predict the  $f_0F_2$

values for the WP937 station. The inverse scenario is illustrated in Figure 2(right); the observations in this case are for the SN437 station, and the networks trained for this station, as well as for the WP937 station, were also used to predict the  $f_0F_2$  values for the SN437station. These two stations are closely located in latitudes but are spaced

out in longitude. The RMSD obtained in using the WP937-station model to predict  $f_0F_2$  values in its station gives an RMSD of 0.66 MHz for its prediction; but yields much higher RMSDs of 1.46 MHz at SN437. On the other side, SN437-station model predicted  $f_0F_2$  values with an RMSDs of just 0.66 MHz at its particular station but gives a much higher RMSDs of 0.94 MHz when applied at WP937. Both of the Single station models predicted the  $f_0F_2$  more accurately at their particular localities and

clearly overestimated values of the  $f_0F_2$  ionosonde observations when used at different localities. This finding indicates that model predictions are better (in terms of reduced prediction errors) for the stations for which they are developed than for a different station. This result, visibly point out that models developed for a particular station cannot be effectively applied in another station located farther apart in space.



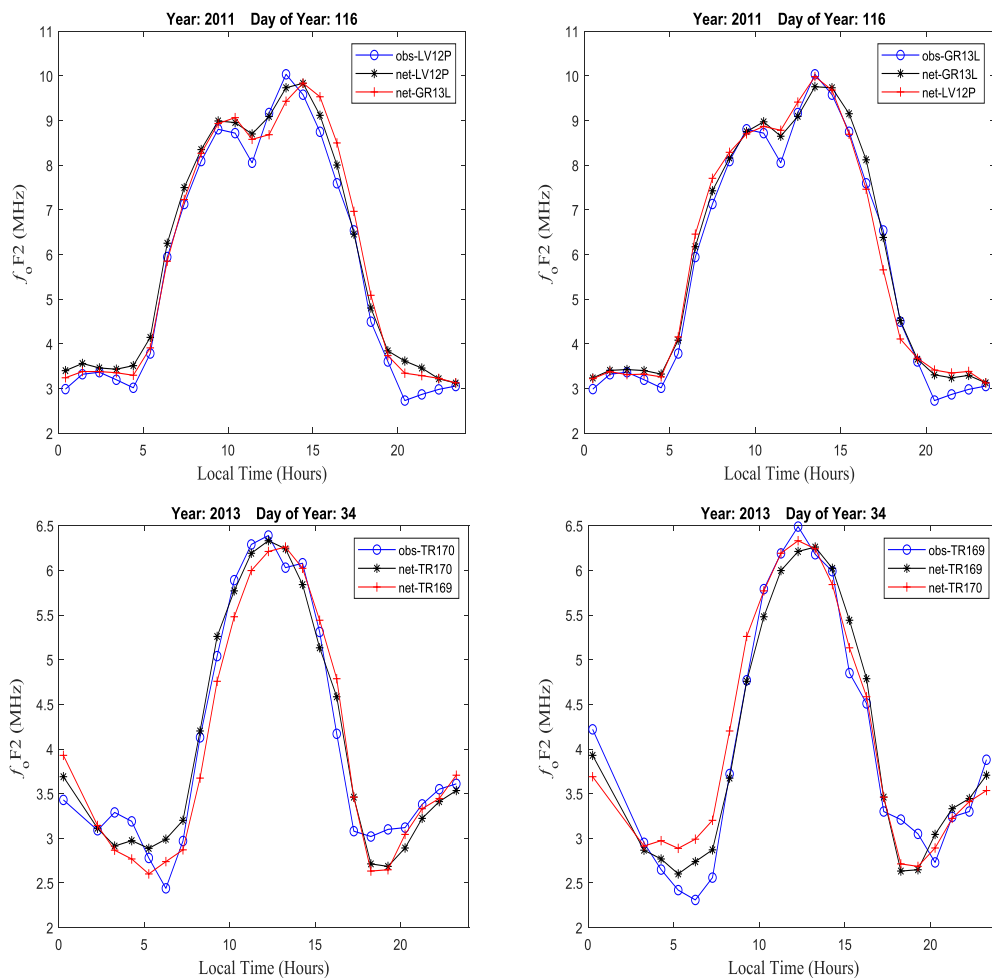
**Figure 3:** Stations closely located in longitudinally but separated in latitude.



In this part, it was demonstrated that there are observable differences in the performances of the two models for two stations with similar longitude but different latitudes. To do this, an investigation of the relationship between  $f_0F_2$  and TEC at HE13N, LG178, stations are illustrated. Both stations are separated by over 44 degrees in latitude and separated just by less than 8 degrees in longitudes. HE13N is a mid-latitude station while LG178 station is a high latitude station. The RMSD obtained in using the HE13N-station model to predict  $f_0F_2$  values at its locality is just 0.33 MHz while its predictions at LG178-station

model produced an RMSDs of 1.27 MHz. Also, using the LG178-station model to predict  $f_0F_2$  values at its locality gives an RMSE of 0.53 MHz and 0.96 MHz when used in HE13N. The results show that there are observable differences in the performances of the two models when used at either of the locations. The models perform more accurately at the stations for which they are developed than at the alternate stations. Conducting repeated tests using other similarly located stations showed that there were latitudinal differences in the relationships between  $f_0F_2$  and TEC.

#### Stations closely located both in latitude and longitude

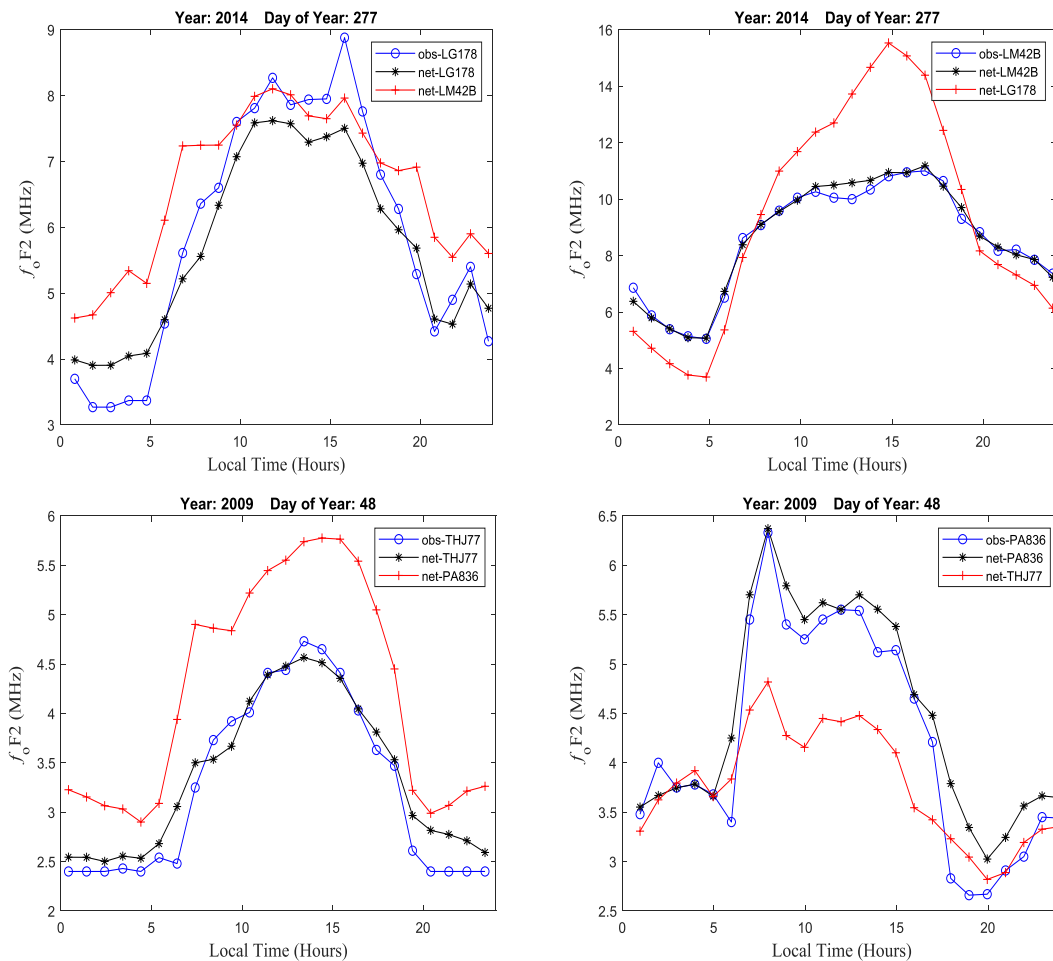


**Figure 4:** Stations closely located in latitudes and in longitudes.

More extensively, further investigation on the effect of latitudinal and longitudinal differences on the accuracy of singles stations models was done. Figures 4 illustrate a scenario of two stations (LV12P and GR13L) which are more closely located. The two stations are both in South Africa and are just 2 degrees apart. The figures show that the errors in the network predictions are not as high as the earlier cases demonstrated. The RMSD obtained in using the LV12P-station model to predict  $f_0F_2$  values for the LV12P station is 0.37 MHz while using the GR13L-station model

for the same LV12P station prediction gives an RMSD of 0.40 MHz. On the other hand, using the GR13L-station model to predict  $f_0F_2$  values for the GR13L station gives an RMSD of 0.29 MHz while using the LV12P-station model for the GR13L station prediction gives an RMSD of 0.37 MHz. The results therefore show that the relationships between  $f_0F_2$  and TEC are similar when the two pairs of stations are more closely located in space. This supports the idea of using neural networks to spatially interpolate the relationships between  $f_0F_2$  and TEC.

### Stations located far apart both in latitude and in Longitudes



**Figure 5:** Stations Far apart in latitudes and in longitudes.

Finally, to validate that there are spatial differences in the relationship between  $f_0F_2$  and TEC, single stations that are located far apart both latitudinally and longitudinally were studied. This is demonstrated in figure 5. The two stations studied; LM42B and LG178, are separated by a distance of 12,620 km. The RMSD obtained in using the LM42B-station model to predict  $f_0F_2$  values for the LM24B station is 0.34 MHz while using the LG178-station model for  $f_0F_2$  prediction at the same LM42B station gives a much higher RMSE value of **2.22 MHz**. On the other hand, using the LG178-station model to predict  $f_0F_2$  values for the LM42B-station gives an RMSD of 0.52 MHz; while using the LM42B-station model for the LG178-station  $f_0F_2$  prediction also gives a much higher RMSE of **2.03 MHz**. Obviously, it can be observed that there is greater error, when a model developed for a particular station is used for  $f_0F_2$  prediction in another station located far away from the former. Likewise, further analysis of similarly related stations portrays the same trend observed in the two Stations discussed above. The idea is to allow the neural network to use these two input neurons to learn spatial variations that are inherent in the relationship between  $f_0F_2$  and TEC.

## CONCLUSION

A neural network based method for estimating ionospheric  $f_0F_2$  from GNSS-TEC measurements is implemented. By using GNSS-TEC, and certain other parameters, as input for the neural networks, the networks were trained to learn the relationship between the  $f_0F_2$  and TEC. Available  $f_0F_2$  and TEC data obtained from about 20 pair of ionosonde and GNSS

receiver stations for the period from year 2000 to year 2016 was used.

Results from the work indicate that the relationship between  $f_0F_2$  and TEC is mostly dependent on the seasons, followed by the level of solar activity, and then the local time. The dependence on geomagnetic activity was the least significant. The relationship between  $f_0F_2$  and TEC also vary with space, and the variation is less for closely located stations. Considering these number of factors that affect the relationship between  $f_0F_2$  and TEC, our study presents an elegant way of using neural networks to relate  $f_0F_2$  and TEC. The method offers an effective and refined way of relating the  $f_0F_2$  and TEC.

## Acknowledgement

This research was supported by the University Support Program (USP) of the Centre for Atmospheric Research (CAR), National Space Research and Development Agency, Nigeria. We thank the South African National Space Agency (SANSA) Space Science, and the Space Physics Interactive Data Resources (SPIDR) from which we obtained ionosonde data. We also thank the University of California, and the South African network of continuously operating GNSS base stations (TRIGNET) from which we obtained GNSS data. We appreciate developers of the IRI Model for making their model available. Thanks to Boston College and Dr. Gopi Seemala for providing us with the GPS-TEC analysis application software used in this work. We also appreciate the Space Physics Data Facility of the Goddard Space Flight Center, National Space Administration's omniweb from which we obtained sunspot number and Dst data.

**REFERENCES**

- Athieno, R., Jayachandran, P. T., and Themes, D. R. (2017). A neural network based model for single station in the polar cap, *Radio Science*, 52, doi:10.1002/2016RS006192
- Cander, L.R. (1998). Artificial neural network application in ionospheric studies. *Annals of Geophysics* :41(5), <https://doi.org/10.4401/ag-3817>.
- Fayose R.S, Rabiou, B., Olakunso, O & Grooves K. (2012). Variation of TEC and their effect on GNSS over Akure, Nigeria. *Applied Physics Research*, 4, 2.
- Gurney, K. (2009). *An introduction to Neural Networks*. Ist Edition. University of Sheffield UK. UCL press. ISBN 1857285034.
- Habarulema, J.B., Mckinnell, L. A., and Cilliers, P. J. (2007). Prediction of global positioning system total electron content using neural networks over South Africa, *Journal of solar and Terrestrial Physics*, 69, 2442-2451.
- Kohli, S., Miglani, S., and Rapariya, R. (2014). Basics of Artificial Neural Network. *International Journal of Computer Science and Mobile Computing*, 3,745-751.
- Kouris, S., Polimeris, K., &Cander, Lj. (2006). Specifications of TEC variability. *Advances in space research* 37, 983-1004
- Mannucci, A. J., Wilson, B. D., and 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)*, 1323–1332, Salt Lake City, UT.
- Mckinnell, L. A., and Poole, W.V. (2004). Neural network based ionospheric modelling over the South African region, *South African Journal of Science*, 100
- Otugo, V., Okoh, D., Okujagu, C., Onwuneme, S., Rabiou, B., et al. (2019). Estimation of ionospheric critical plasma frequencies from GNSS-TEC measurements using artificial neural networks. *Space weather* 17. <https://doi.org/10.1029/2019SW002257>
- Oyeyemi, E. O., Poole, A. W. V., and McKinnell, L. A. (2005), On the global short-term forecasting of the ionospheric critical frequency  $f_oF_2$  up to 5 hours in advance using neural networks, *Radio Sci.*, 40, RS6012, doi:[10.1029/2004RS003239](https://doi.org/10.1029/2004RS003239).
- Seemala, G. K., and Valladares, C. E. (2011). Statistics of total electron content depletions observed over the South American continent for the year 2008. *Radio Sci.*, 46, RS5019, doi:10.1029/2011RS004722.
- Spalla, P. & Ciraolo, L. (1994). TEC and  $f_oF_2$  comparison. *Annals of Geophysics*, 37, 929-938
- Stamper, R., Belehaki, A., Buresova, D., Cander, L.R., Kutiev, I., Pietrella, M., Stanislawski, I., Stankov, S., Tsagouri, I., Tulunay, Y. K and Zolesi, Z (2004). Nowcasting, forecasting and warning for ionospheric propagation: tools and methods. *Annals of Geophysics, Supplement to vol. 47, N. 2/3, 2004*.
- Wintoft, P.& Cander, L. R (2000). Twenty-four-hour prediction of  $f_oF_2$  using time delay neural networks. *Radio Science*, 35 (2): 395-408. DOI:10.1029/1998R02149