Bulgarian Journal of Physics vol. **XX** (20XX) 1–9

Prediction of Total Electron Content Using ARIMA and Neural Network

Bornali Chetia, **Aidashisha Bareh**

Department of Physics, Royal Global University, Assam, India

Received *12 February 2025*

Abstract. *Objectives*: To predict short term ionospheric Total Electron Content (TEC) data derived from the International GNSS Service (IGS) station. *Methods*: A new hybrid technique based on Autoregressive Integrated Moving Average (ARIMA) and Artificial Neural Network (ANN) model for the short-term ionospheric TEC prediction is constructed. The TEC dataset in this research work is utilized from Crustal Dynamics Data Information System (CDDIS), NASA from the year 2014 to 2016 and for each day the 10 minute average TEC data have been utilized for modeling and prediction. The hybrid model than compared with existing developed model such as ANN and ARMA as well as IRI-2016 global model. *Findings*: In order to estimate the performance of new hybrid model, it is compared with the existing ANN model, ARIMA model and IRI-2016 global model. Based on the comparison results, it is observed that new hybrid model predicts well than other prediction models with RMSE of 10.21 TECU and MAPE 0.034 TECU. *Novelty*: The proposed model can recognize the usual pattern of TEC in different seasons and it can predict the TEC values with 86% accuracy up to 24 hours.

KEY WORDS: IGS, GNSS, Total electron content, ARIMA, ANN.

1 Introduction

Total Electron Content (TEC) is a very important ionospheric parameter and is used as an indicator for detecting various atmospheric irregularities. It is the determination of the total number of electrons per square meter near the line of sight from satellite to the receiver down where 1 TECU (TEC Unit) $= 1 \times 10^{16}$ electron/m^2. Variations and characteristics of TEC at different latitudes have been studied by many scientists [1, 2]. Earth ionosphere elements are a very complex visual process that reflects a variety of daily, seasonal, and latitudinal features. The seasonal variation of TEC at different latitudes depends on various factors such as solar zenith angle, medium wind direction, thermospheric neutral formation etc. [3–8]. It is noteworthy that differences in total electron content (TEC) in the equator and low latitude are significant compared to the mid-latitude regions. The prediction of TEC is quite challenging and complex since TEC also depends on the effect of sun solar activity, geomagnetic storm effects, seasonal and diurnal variations etc. International Reference Ionosphere (IRI) and Klobuchar the most widely used models of climate models in the world in comparison with VTEC values measured by GPS. Model developed by N. Elmunim [9] and A. Krankowski model [10] is used for short-term forecast of the TEC timeline at different geomagnetic times. However, these are standing time series models consider the trend in a given input. A TEC timeline that may not be suitable for prediction non-concentrated / variable patterns present in ionospheric TEC. Therefore, artificial neural networks are used by various researchers, in order to consider atmospheric weather and influential geomagnetic features of non-ionospheric patterns TEC.

In recent study, artificial neural network and ARIMA models has been utilized for consideration to design better model prediction and variations in ionospheric TEC.Dabbakuti et al has designed a model based on combining the signal extraction technique Singular Spectrum Analysis (SSA) with Autoregressive Moving Average (ARMA) for prediction of TEC [11]. Sivavaraprasad G. et al. has studied the performance of TEC prediction models based on Artificial Neural Networks (ANN) [12]. Tang Rongxin et al. has designed and compared model based on Autoregressive Integrated Moving Average (ARIMA), Long Short-Term Memory (LSTM) and Sequence-to-sequence (Seq2Seq)method [13]. Machine learning techniques has been also utilized for forcasting TEC data by different researchers around the globe [14]. In this paper an attempt is made to develop and analyze a new mixed TEC forecast method using the Autoregressive Integrated Moving Average and the Artificial Neural Network. The predicted TEC values are compared with the values observed by the IGS channel to assess the reliability and performance of the forecast model. The TEC data used in this study is used in the International GNSS Service (IGS) Lasha, China station.

2 Methodology

Ionosphere characteristics depend on different geomagnetic parameters such as solar activity, sunspot number, geomagnetic indicator etc. In this paper, the average 10-minute interval of TEC values is considered as input to the model. The inclined TEC values are first converted to a straightforward TEC value to fit into the model.

Suppose, $T = \{T_v(t),\ t = 1, \ldots, n\}$, is a time series for daily TEC data at discrete time t and number of data points at time t is represented by n. Our aim is to predict the $T_v(t+1)$ value at time $t+1$. For this, we have to convert the vertical TEC to its similar volatility index with the following formula:

$$V(t) = (\log(T_v(t+1)/_T v(t))\,. \tag{1}$$

ARIMA (p, q, d) is used as the mean process. The ARIMA (p, q, d) can be represented as in Eq. (2)

$$\left(1 - \sum_{k=1}^{p} \alpha_k B^k\right)(1-B)^d X_t = \left(1 + \sum_{k=1}^{q} \beta_k B^k\right)\varepsilon_t\,, \tag{2}$$

where X_t is the time series, α and β are the parameters/coefficients of autoregressive and moving average terms with order p and q respectively. ε_t are error terms generally assumed to be independent, identically distributed variables sampled from a normal distribution with zero mean. B is the difference operator defined as follows by Eq. (3), where d is the order of the difference operator

$$\Delta X_t = X_t - X_{t-1} = (1-B)X_t\,. \tag{3}$$

In order to find the series is stationary or not, the autocorrelation of the series has to be found out. The autocorrelation between time X_t and $X_{t+\tau 1}$ is given by the Eq. (4), where X_t and $X_{t+\tau 1}$ are the time series with lag t. μ is the mean of the population. σX_t, $\sigma X_{t+\tau}$ are standard deviations at the time t and $t+\tau$. ρ_k is the autocorrelation at lag k given by the Eq. (4).

$$\mu\tau \;\mu\rho_k = \frac{E\left[(X_t-\mu)(X_{t+\tau}-\mu)\right]}{\sigma X_t \sigma X_{t+\tau}}\,. \tag{4}$$

In ANN model, single hidden layer feed forward network is the mostly utilized by the researchers for time series modeling as well as for the prediction. In the model, a collection of processing units for construction of three layers network is used. The relationship between the inputs $(y_{t-1}\,, yt-2\;\ldots\,, y_{t-p})$ and the output (y_t) has the following mathematical representation:

$$y_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j g\left(\beta_{ij} y_{t-i}\right) + \varepsilon_t\,, \tag{5}$$

where α_j $(j = 0, 1, 2, \ldots, q)$ and β_{ij} $(i = 0, 1, 2, \ldots, p; j = 1, 2, \ldots, q)$ are the connection weights; the total number of input nodes are represented by p and the total number of hidden nodes are represented by q. The activation function that is used in this research is the sigmoid function which can be represented as follows.

$$g\left(x\right) = \frac{1}{1+\mathrm{e}^{-x}} \tag{6}$$

Therefore, we can say that the ANN model executes a nonlinear functional mapping for the past observations $(y_{t-1}\,, y_{t-2}\,, \ldots\,, y_{t-p})$ to the future value y_t, i.e.,

$$y_t = f(y_{t-1}\,, y_{t-2}\,, \ldots, y_{t-p}\,, W) + \varepsilon_t\,. \tag{7}$$

Here, W is a vector for the all parameters and f is a function determined by the network structure and connection weights.

ARIMA and ANN are ideal for both direct and indirect domains that can be used to model for protection. ARIMA model estimates for complex non-linear problems may not produce equally positive results; ANN will provide mixed results on the line domain. Since many details are complex in nature, it is very important to find its features. Mixed methods are a great way for you to model both direct and indirect time series in real-time applications from a different coalesce model that can identify basic data patterns.

Practically a time series may be considered to have a linear component and a non-linear component as shown

$$y_t = l_t + n + t \tag{8}$$

where l_t is the linear part and n_t is the nonlinear part. The two parts can be measured separately from the data. First the linear part is determined that is separated from the time series. This residual is fitted with the ANN model. This is fundamental strategy behind this model. Let e_t be the residuals which can be obtained from the time series by subtracting forecasted value l_t from ARIMA model as shown

$$e_t = y_t - l_t \,. \tag{9}$$

If there is a linear correlation left over, then linear models are not sufficient for predicting data. It is noted that residual analysis is the best option for capturing indirect patterns in data. Therefore, if the model has passed a diagnostic test, the model may not be eligible because the indirect relationship has not been properly modeled. Indirect relationships between data can be captured if modeling of fossils is done with the help of ANN. Total mean percentage error (MAPE), mean total error (MAE) and root mean square (RMSE) error can be used to calculate predictive accuracy.

3 Dataset and Simulation Model

The ionospheric TEC data for the year 2014 to 2016 were considered for analysis and for each day the 10 minute average TEC data have been utilized for modelling and prediction. For each month the first ten quietest days are considered for modelling and the next day is considered for checking the validity of the prediction. The prediction of TEC data on 01 January 2015 and (b) 06 April 2015 are presented in Figure 1 (a) and (b) respectively.

(a)

(b)

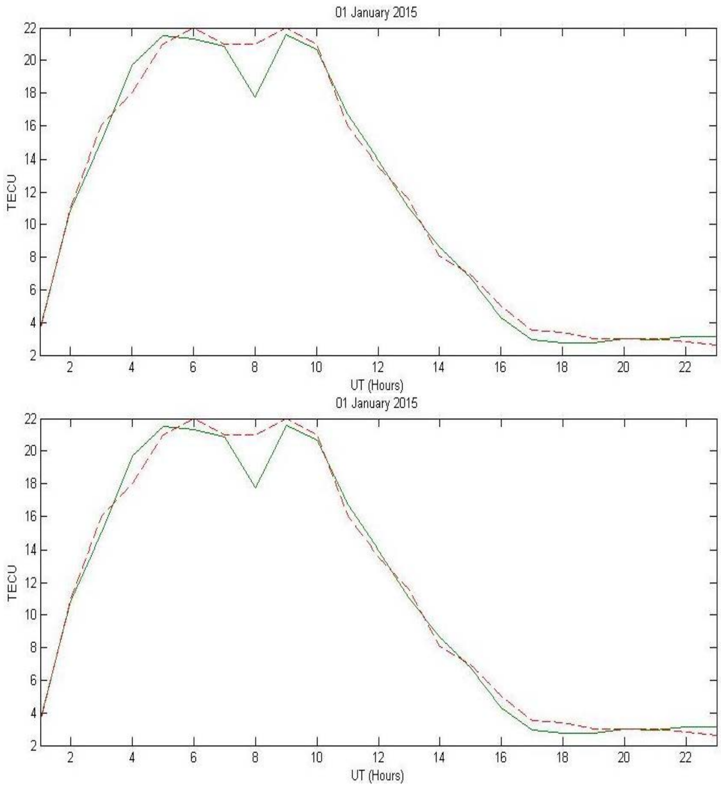

Figure 1. Comparison of the predicted TEC values by the model and the observed TEC values for: (a) 01 January 2015; and (b) 06 April 2015. The green line indicates the observed TEC collected at Lasha, China station and the red line indicates the predicted TEC.

4 Results and Discussion

4.1 Model performance

From Table 1, one can observe that the results of the statistical deviation (SD) are astonishing. The SD between the model prediction and the marked TEC 24

Table 1. Precision statistics of the TEC predicted by models (δ = predicted value - observed value)

Forecast time	$\delta \leq 1$ TECU	1 TECU $\leq \delta \leq$ 2 TECU	2 TECU $\leq \delta <$ 3 TECU
24 hours	86%	33%	18%
48 hours	66%	35%	14%
72 hours	43%	45%	9%

hours is more accurate than predictable, 48 hours and 72 hours forecast. By analyzing SD it can be concluded that the model produces very good TEC predictions over time and has good performance. From the survey it was also noted that the performance of TEC short-term price estimates is better than long-term estimates.

To evaluate the performance of the new model based on ARIMA and Neural Network, the forecast results are compared with existing models namely ANN, ARMA, IRI-2016.

From the comparison results in Figure 1 and Table 1, it is observed that, the new model prediction plot matches well with true TEC values than existing ANN, ARMA, IRI-2016 predictions. The results in Table 2 show that the RMSE of new hybrid model is 10.21 TECU and MAPE is 0.034 TECU. This result specifies that the new model is able to simulate daily TEC data with an accuracy of a degree in comparison to the ANN, ARIMA, IRI-2016 models. Xu Lin et al. [14] has utilized LSTM as one of the modelling tools for GPS TEC, but the new hybrid model based on ARIMA produced lower error values than a model based on LSTM for short term time series prediction which indicated that ARIMA is more successful than the LSTM for short term time series prediction [15]. Thus, the new algorithm for GPS TEC prediction can be able to recognize the pattern of the input data to provide good predictions of the daily variations of TEC data. Therefore, it can be concluded that the hybrid modelling approach can give more reliable predictions of total electron data than the other existing modelling approach.

Table 2. MAPE and RMSE of ARMA and ANN, ARIMA and ANN model

Models	MAPE	RMSE
ARIMA + ANN	0.034	10.21
ANN	0.068	19.33
ARMA	0.152	33.42
IRI	0.086	13.33

4.2 Seasonal Variation in model performance

Past studies indicate that the performance of TEC prediction model varies seasonally [16–18] Therefore, the model is operated in each season to investigate the performance throughout the solar year. The seasons are classified as winter (December to February), spring (March to May), summer (June to August) and autumn (September to November).

Tables 3 indicates that the ARIMA-ANN model delivered the best performance in each season for the testing data from spring, summer, autumn, and winter, respectively. The IRI 2016 model delivered the worst performance. This indicates

Table 3. Model performance during seasonal variations

	SEASONS	R2	RMSE	MAE
ARIMA-ANN	Winter	0.812	13.89	5.11
	Spring	0.927	6.342	4.63
	Summer	0.982	6.775	4.12
	Autumn	0.843	12.35	5.91
ARMA	Winter	0.625	16.33	12.46
	Spring	0.722	7.36	7.47
	Summer	0.735	8.66	8.45
	Autumn	0.561	15.89	11.34
IRI 2016	Winter	0.338	19.34	14.89
	Spring	0.638	8.32	8.24
	Summer	0.585	8.89	8.53
	Autumn	0.395	17.87	12.47

that relatively advanced models achieved significantly better TEC-estimation accuracy than did simpler models based on their consideration of more space environment parameters. Overall, prediction error was largest in winter, although the winter TEC is smaller than those in spring and summer. This may be due to the effects of solar activities. The performances of utmost models on the seasonal scale were considerably better than those on the periodic scale, thus highlighting the effect of seasons on TEC estimation. The R^2 values obtained using conventional models (ARIMA, and IRI 2016) at the seasonal scale (especially in spring and summer) were significantly improved over those obtained at the annual scale, thus verifying that conventional models are more appropriate for seasonal observations.

5 Conclusions

Although the results in this study shows that ARIMA-based ANN models can be used for short term ionospheric TEC predictions, future studies should focus on the inclusion of more parameters to the model such as foF2, Dst etc. to estimate TEC in the long term. Pre-processing techniques, such as the combination of wavelet decomposition with machine learning, could also be explored for TEC forecasting.

References

[1] Siti Syukriah Khamdan, et al. (2019) Seasonal Variations of Equatorial Anomaly Crest using GPS Ionospheric Tomography. *J. Phys.: Conf. Ser.* **1152** 012013.

[2] D. Pandit, et al. (2021) Climatology of ionosphere over Nepal based on GPS total electron content data from 2008 to 2018. *An. Geo.* **39**(3) 743-758.

[3] G. Ma, et al. (2020) A study of ionospheric irregularities with spatial fluctuation of TEC. *J. Atmos. Sol.-Terr. Phys.* **211** 105485.

[4] F. Abbas, et al. (2023) Evaluation of ionospheric and solar proxy indices for IRI-Plas 2020 model over Pakistan and Japan during different solar activity epochs. *Adv. Space Res.***72** 5551-5562.

[5] G. Manju, R.P. Aswathy (2020) Ionospheric Planetary Wave Activity and Its Role in Equatorial Spread F Day-to-Day Variability. *JGR Space Phys.* **125** e2020JA027960.

[6] A. Calabia, S. Jin (2020) New Modes and Mechanisms of Long-Term Ionospheric TEC Variations From Global Ionosphere Maps. *JGR Space Phys.* **125** e2019JA027703.

[7] S. Kumar, S. Priyadarshi, S. Gopi Krishna, A.K. Singh (2012) GPS-TEC variations during low solar activity period (2007-2009) at Indian low latitude stations. *Astrophys. Space Sci.* **339**(1) 165-178.

[8] O.A. Maltseva, T.V. Nikitenko, A.K. Singh (2021) Ionospheric Response to the Geomagnetic Storm of 20th April 2020 over Different Latitudinal Areas. In: *Proc. 2021 Radiation and Scattering of Electromagnetic Waves (RSEMW),* pp. 292-295.

[9] N.A. Elmunim, M. Abdullah, A.M. Hasbi, S.A. Bahari (2013) Short-term forecasting ionospheric delay over UKM, Malaysia, Using the Holt- Winter method. In: *Proc. IEEE International Conference on Space Science and Communication (IconSpace), Melaka, Malaysia*, pp. 106-109.

[10] A. Krankowski, W. Kosek, L. Baran, W. Popinski (2005) Wavelet analysis and forecasting of VTEC obtained with GPS observations over European latitudes. *J. Atmos. Solar-Terr. Phys.* **67**(12) 1147-1156.

[11] J.R.K. Kumar Dabbakuti, M. Yarrakula, S.K. Panda, P. Jamjareegulgarn, M.A. Haq (2022) Total electron content prediction using singular spectrum analysis and autoregressive moving average approach. *Astrophys. Space Sci.* **367** 8.

[12] G. Sivavaraprasad, V.S. Deepika, D. SreenivasaRao, M. Ravi Kumar, M. Sridhar (2020) Performance evaluation of neural network TEC forecasting models over equatorial low-latitude Indian GNSS station. *Geod. Geodyn.* **11**(3) 192-201.

[13] R. Tang, F. Zeng, Z. Chen, J.-S. Wang, C.-M. Huang, Z. Wu (2020) The Comparison of Predicting Storm-Time Ionospheric TEC by Three Methods: ARIMA, LSTM, and Seq2Seq. *Atmosphere* **11** 316.

[14] Yi Han, Lei Wang, Wenju Fu, Haitao Zhou, Tao Li, Ruizhi Chen (2022) Machine Learning-Based Short-Term GPS TEC Forecasting During High Solar Activity and Magnetic Storm Periods. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **15** 115-126.

[15] Rui Zhang, Hejia Song, Qiulan Chen, Yu Wang, Songwang Wang, Yonghong Li (Jan. 14, 2022) Comparison of ARIMA and LSTM for prediction of hemorrhagic fever at different time scales in China. *Plos One*.

[16] R. Mukesh, V. Karthikeyan, P. Soma, P. Sindhu (2020). Forecasting of ionospheric TEC for different latitudes, seasons and solar activity conditions based on OKSM. *Astrophys. Space Sci.* **365**(1) 13.

[17] A. Ruwali, A.J.S. Kumar, K.B. Prakash, G. Sivavaraprasad, D.V. Ratnam, D. V. (2020) Implementation of hybrid deep learning model (LSTM-CNN) for ionospheric TEC forecasting using GPS data. *IEEE Geosci. Remote Sens. Lett.* **18**(6) 1004-1008.

[18] R. Song, X. Zhang, C. Zhou, J. Liu, J. He (2018) Predicting TEC in China based on the neural networks optimized by genetic algorithm. *Adv. Space Res.* **62**(3) 745-759.