**Intercontinental Geoinformation Days**

http://igd.mersin.edu.tr/2020/

# Ionospheric TEC Prediction Performance of ARIMA and LSTM Methods in Different Space Weather Conditions

**Erman Şentürk*[1]**

[1]Kocaeli University, Faculty of Engineering, Department of Geomatics, Kocaeli, Turkey

**Keywords**
ARIMA
Geomagnetic Storm
Ionosphere
LSTM
TEC Prediction

**ABSTRACT**

The ionosphere has some temporal regular changes under the dominant control of the Sun. The stationary structure of the ionospheric time series (e.g. TEC, foF2) allows it to be modeled on a specific time. In this study, we tested the performance of the artificial intelligent (AI) techniques e.g. a machine learning-based method, autoregressive integrated moving average (ARIMA), and a deep learning-based method, long short-term memory (LSTM) network to the prediction of Total Electron Content (TEC) values. The TEC data of six different locations in low, middle, and high latitudes were selected from the Center for Orbit Determination in Europe – Global Ionosphere Maps (CODE-GIMs). To show the performance of the proposed methods during quiet space weather and a severe geomagnetic storm, we trained the 60 days TEC data (24 data points in one day) and forecasted the TEC data of the subsequent five days by fitted models with optimal hyperparameters. The forecasted TEC values were compared with observed TEC through some statistical metrics (RMS, MAE). The results indicated that the LSTM is more successful in TEC prediction than ARIMA. This study brings new insights into the AI techniques in the ionospheric TEC prediction.

## 1. INTRODUCTION

The ionosphere is a three-dimensional dispersive medium atmosphere layer whose primary driver is the Sun. The layer locates above approximately 50-1000 km from the Earth's surface and includes molecules with potential for photoionization. When molecules are exposed to light energy emitted from the Sun, their components are divided into atoms, which are negative electrons and positive ions. Negatively charged electrons affect the propagation of electromagnetic signals traveling between the earth and space.

The number of free electrons is described by the Total Electron Content (TEC) parameter. The TEC describes the number of free electrons in a cylinder with a 1 m$^2$ base area throughout the line-of-sight (LOS). The unit of the TEC (TECU) is equal to $10^{16}$ electron/m$^2$. TEC values have periodic temporal and spatial variations such as the diurnal, 27-day, seasonal, semi-annual, annual, and 11-year under control of the Sun (Vaishnav et al., 2019).

The TEC also increase/decrease due to space weather events such as solar winds, solar flares, geomagnetic storms (Bagiya et al., 2009), earthquakes (Şentürk et al., 2019), tsunamis (Occhipinti et al., 2013), volcanic eruptions (Dautermann et al., 2009), hurricanes/typhoons (Chen et al., 2020) and anthropogenic events (Lin et al., 2017). These events generally cause non-secular changes and affect the regular change of TEC variation.

There are some traditional time series analysis methods to modeling the TEC time series, but these methods are not adequate to simulation the previous TEC observation and the pattern that is far away from forecasting-initial-point, as any artificial intelligence (AI) algorithms. However, AI method such as ARIMA and LSTM learns the trend, seasonality, and residuals patterns in the TEC time series and successfully forecasting TEC values for a short period. Some AI-based methods were previously utilized to forecast ionospheric parameters (McKinnell and Poole, 2004; Athieno et al., 2017; Sai Gowtam and Tulasi Ram, 2017; Srivani et al., 2019; Kaselimi et al., 2020; Ruwali et al., 2020).

In this study, we discussed the advantages and disadvantages of the ARIMA and LSTM methods for ionospheric TEC forecasting. For this purpose, TEC data of CODE-GIMs were obtained in low, middle, and high latitudes of the hemispheres during quiet space weather and geomagnetic storm. The forecasting performance of the methods was compared using some statistical metrics.

* **Corresponding Author**

* (erman.senturk@kocaeli.edu.tr) ORCID ID 0000-0002-0833-7113

**Cite this study**

Şentürk E (2020). Ionospheric TEC Prediction Performance of ARIMA and LSTM Methods in Different Space Weather Conditions. Intercontinental Geoinformation Days (IGD), 32-35, Mersin, Turkey

## 2. METHOD

We have fitted two models to forecast TEC values; the ARIMA and LSTM. The ARIMA is a statistical, traditional, machine learning approach while LSTM is a more advanced version of a special kind of Artificial Neural Network (ANN) named Recurrent Neural Network (RNN). We split our analysis into three divisions: (1) data collection and pre-processing, (2) formulation of the models, and (3) implementations which are the following:

### 2.1. Data Collection and Pre-processing

The TEC time series of $75^0$N-S (high latitude), $45^0$N-S (middle latitude), and $15^0$N-S (low latitude) at prime meridian were obtained from GIMs produced by the CODE. These GIMs are gridded between $\pm 87.5^0$ N-S and $\pm 180^0$ W-E with a $2.5^0$x$5^0$ spatial resolution, respectively, and with a 1-hour temporal resolution. The gridded TEC values are published by files in the Ionosphere Map Exchange Format (IONEX), which is freely available in ftp://cddis.gsfc.nasa.gov/gps/products/ionex/.

We selected the TEC data at a two-time interval for quiet space weather (from February 27 to May 01, 2020) and a geomagnetic storm (from June 28 to August 31, 2018). The geomagnetic storm is identified by the disturbance storm-time (Dst) index which decreased to -174 nT on August 26, 2018. This Dst value indicates a severe geomagnetic storm. Also, the quiet space weather period is decided by threshold values of Dst > -20 nT, solar radio flux (F10.7) < 90 sfu. The indices are available at https://omniweb.gsfc.nasa.gov/form/dx1.html.

### 2.2. Model formulation

#### 2.2.1. Autoregressive Integrated Moving Average (ARIMA)

ARIMA is a time series model based on traditional statistical concepts and integration of two methods: Auto Regression (AR), and Moving Average (MA).

An ARIMA model can be defined as three parameters: (1) p is the number of autoregressive terms, (2) d is the number of nonseasonal differences needed for stationarity, and (3) q is the number of moving averages.

Jenkins and Box proposed a method to get the order of ARIMA using the autocorrelation function (ACF) and the partial autocorrelation function (PACF) of the sample data (Bartholomew, 2020). The parameters, q calculates using ACF plot and p obtains from the PACF plot.

We can formulate ARIMA (p, d, q) as follows:

$$Y_t = \delta + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \alpha_3 Y_{t-3} + \cdots + \alpha_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \theta_3 \varepsilon_{t-3} - \cdots - \theta_q \varepsilon_{t-q} \quad (1)$$

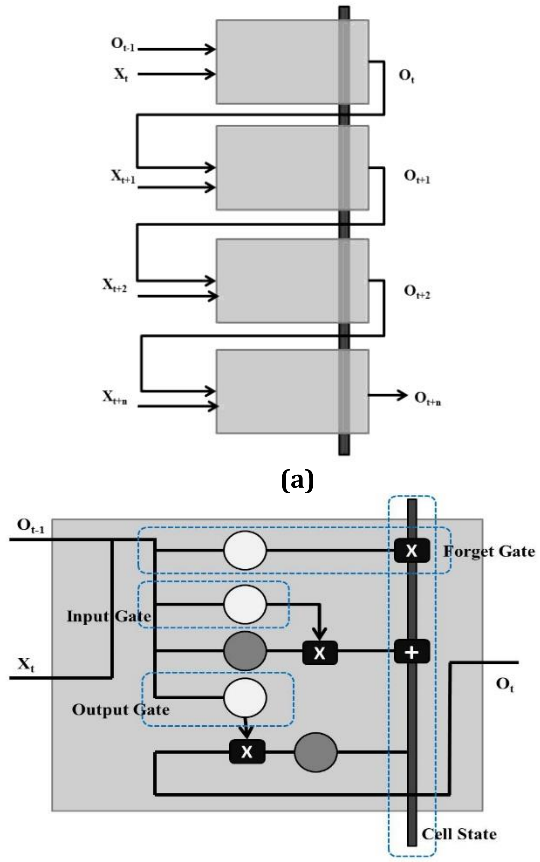
where $\delta$ is a constant, $Y_t$ is linear combinations of the previous time-series terms with the coefficients $\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_p$ and $\varepsilon_t$ is a random shock at time t.

#### 2.2.2. Long Short Term Memory (LSTM) Network

Time series is a very special type of data in which dependent and independent variables are the same, and the target attribute depends on its previous observations rather than on the independent variables.

TEC variation has the same daily min-max values with little erratum but this erratum becomes high during any special events (e.g. magnetic storms, solar activity, or earthquakes). Traditional ANN is advanced enough to capture and learn the irregularity in the TEC data and forecast upcoming value but being a time series data it is highly correlated with the previous terms. So, RNN is used to pass previous learning to the adjacent nodes. Although RNN predicts well, it does not have any memory power to keep within whatever it has learned from the nodes that are far away from the current node. This problem is also known as a vanishing gradient problem. LSTM comes into existence to overcome the problem of the vanishing gradient. The special architectures of the LSTM-RNN network (Fig. 1) made it possible to keep those learning within the network and forecast based on these learning.

LSTM networks have four components: Cell State, Input Gate, Forget Gate, and Output Gate (Fig.1b).




(b)

**Figure 1.** The architecture of LSTM. (a) The complete connected architecture of LSTM-RNN, (b) One-single-node representation of LSTM-RNN.

### 2.3. Implementations

We have split our dataset into two-part; training dataset and test dataset. The training data for quiet space weather include 60 days from Feb 27 to Apr 26, 2020, and test data includes the subsequent 5 days between Apr 27 and May 1, 2020. Also, the training data for geomagnetic storm includes 60 days from Jun 28 to Aug

26, 2018, and test data includes 5 days on Aug 27-31, 2018.

ARIMA and LSTM models were implemented in the MATLAB R2019b using Econometrics and Deep Learning Toolboxes.

Two statistical metrics employed to evaluate the performance of the proposed models; Maximum Absolute Error (MAE) and Root Mean Square (RMS) error.

$$\mathrm{MAE} = \max(|\mathrm{TEC}_{\mathrm{Forecast}} - \mathrm{TEC}_{\mathrm{Observed}}|) \quad (2)$$

$$\mathrm{RMS} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\mathrm{TEC}_{\mathrm{Forecast}}\text{-}\mathrm{TEC}_{\mathrm{Observed}})^2} \quad (3)$$

## 3. RESULTS

In this section, we showed our ARIMA and LSTM results only at $15^0$ N and $45^0$ S for both quiet space weather and geomagnetic storm, respectively.

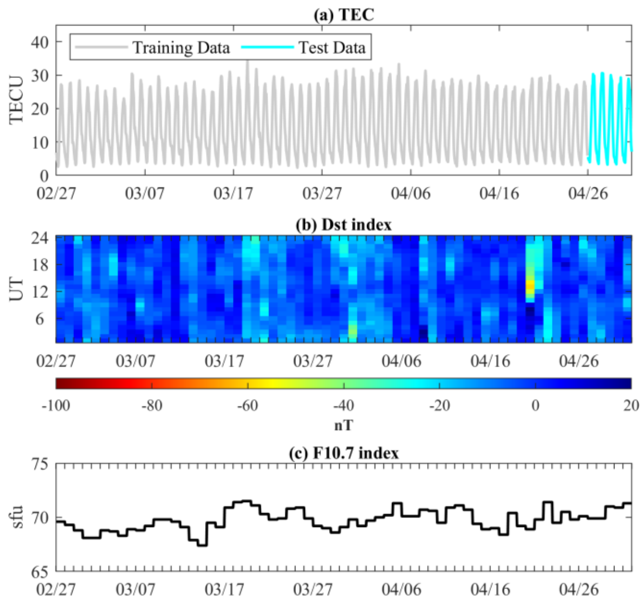



**Figure 2.** (a) TEC time series at $15^0$ N, (b) Dst index, (c) F10.7 index from Feb 27 to May 1, 2020. The gray and cyan lines indicate training and test data, respectively.

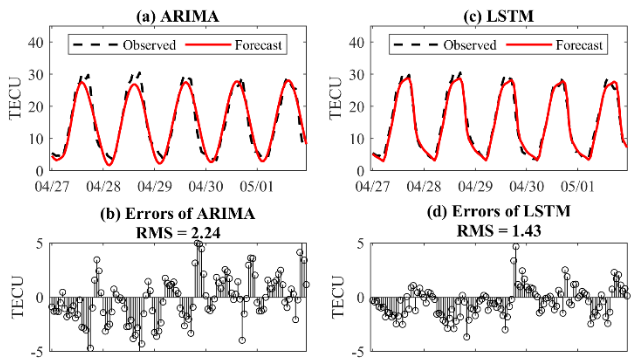



**Figure 3.** (a-c) Observed TEC at $15^0$ N and Forecast TEC calculated by ARIMA and LSTM methods (b-d) Errors from Apr 27 to May 1, 2020.

In Fig. 2, we showed the time series of TEC, Dst, and F10.7 for the quiet space weather period. The Dst values range between -59 nT and 20 nT and F10.7 values range between 67.4-71.5 sfu. These values indicate quiet space weather for ionospheric variation except for Apr 20, 2020. A moderate geomagnetic storm (Dst < -50 nT between 11-13 UT) occurred on the relevant day.

In Fig. 3, we showed the observed TEC, forecast TEC, and RMS errors of proposed methods for the quiet space weather period. We forecasted TEC values with an accuracy of 2.24 and 1.43 TECU for ARIMA and LSTM methods, respectively.

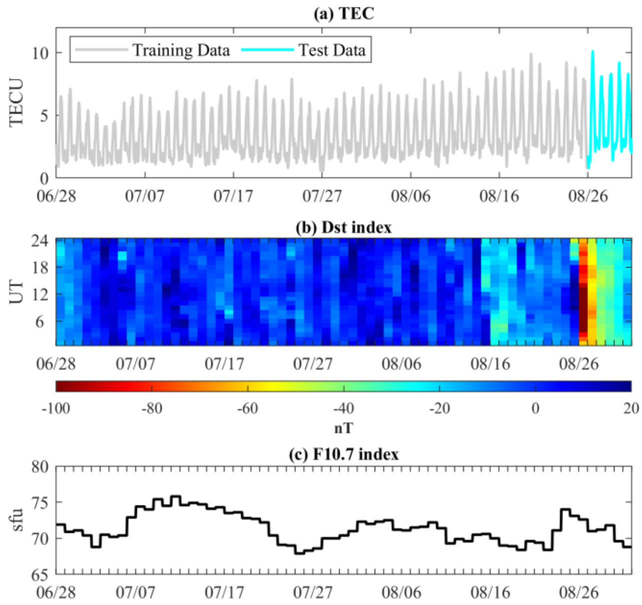



**Figure 4.** (a) TEC time series at $45^0$ S, (b) Dst index, (c) F10.7 index from Jun 28 to Aug 31, 2018. The gray and cyan lines indicate training and test data, respectively.

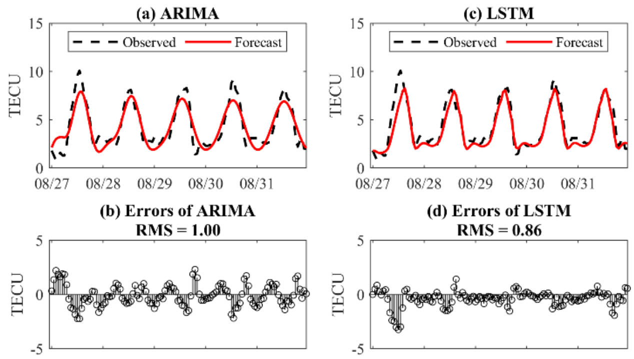



**Figure 5.** (a-c) Observed TEC at $45^0$ S and Forecast TEC calculated by proposed methods (b-d) Errors from Aug 27 to Aug 31, 2020.

In Fig. 4, we showed the time series of TEC, Dst, and F10.7 for the geomagnetic storm period. The Dst values range between -174 nT and 21 nT and F10.7 values range between 67.9-75.8 sfu. The F10.7 values indicate quiet space weather for ionospheric variation in training data but the Dst value of -174 nT indicates a severe geomagnetic storm on August 26, 2018 (in test data).

In Fig. 5, we showed the observed TEC, forecast TEC, and RMS errors of proposed methods for the geomagnetic storm period. We forecasted TEC values with an accuracy of 1.00 and 0.86 TECU for ARIMA and LSTM methods, respectively.

We also showed the RMS and MAE values of proposed methods for other locations in Table 1.

**Table 1.** RMS and MAE values (in TECU) of proposed methods for quiet space weather and geomagnetic storm

| Locations | Quiet Space Weather | | | | Geomagnetic Storm | | | |
|---|---|---|---|---|---|---|---|---|
| | ARIMA | | LSTM | | ARIMA | | LSTM | |
| | RMS | MAE | RMS | MAE | RMS | MAE | RMS | MAE |
| $75^0$ N | 0.61 | 1.25 | 0.88 | 2.95 | 0.96 | 2.95 | 0.95 | 2.60 |
| $45^0$ N | 1.28 | 2.67 | 1.05 | 3.15 | 1.12 | 2.56 | 1.13 | 2.66 |
| $15^0$ N | 2.24 | 6.22 | 1.43 | 4.66 | 3.77 | 11.24 | 2.90 | 9.42 |
| $15^0$ S | 2.28 | 5.56 | 1.46 | 4.03 | 2.48 | 5.46 | 2.76 | 9.34 |
| $45^0$ S | 1.71 | 6.60 | 0.80 | 2.57 | 1.00 | 2.30 | 0.86 | 3.23 |
| $75^0$ S | 0.85 | 1.88 | 0.83 | 1.84 | 1.29 | 3.41 | 1.38 | 3.79 |

## 4. CONCLUSION

In this study, the TEC data of six different locations were used to analyze the prediction performance of ARIMA and LSTM methods in different locations and space weather conditions.

We showed that both ARIMA and LSTM are successful for forecasting ionospheric TEC, but LSTM is more accurate than the ARIMA model. While LSTM produced better results especially in the low and middle latitudes, there was no significant difference between both methods in the high latitudes. Also, both methods generally predicted TEC values with lower RMS in the quiet space weather period than the geomagnetic storm period.

In the study, we have seen the capabilities and abilities of the AI-based models in forecasting the TEC time series. We showed that deep learning methods provide more accurate forecasting TEC data.

## ACKNOWLEDGEMENT

The author is thankful to the CODE for providing the GIMs, and NASA for providing the space weather indices.

## REFERENCES

Athieno, R., Jayachandran, P. T., & Themens, D. R. (2017). A neural network-based foF2 model for a single station in the polar cap. Radio Science, 52(6), 784-796.

Bagiya, M. S., Joshi, H. P., Iyer, K. N., Aggarwal, M., Ravindran, S., & Pathan, B. M. (2009). TEC variations during low solar activity period (2005–2007) near the equatorial ionospheric anomaly crest region in India. Annales Geophysicae, 27(3), 1047-1057.

Bartholomew D. J. (2020). Time Series Analysis Forecasting, Operational Research Quarterly, 22(2), 199-201.

Chen, J., Zhang, X., Ren, X., Zhang, J., Freeshah, M., & Zhao, Z. (2020). Ionospheric disturbances detected during a typhoon based on GNSS phase observations: A case study for typhoon Mangkhut over Hong Kong. Advances in Space Research, 66(7), 1743-1753.

Dautermann, T., Calais, E., Lognonné, P., & Mattioli, G. S. (2009). Lithosphere-atmosphere-ionosphere coupling after the 2003 explosive eruption of the Soufriere Hills Volcano, Montserrat. Geophysical Journal International, 179(3), 1537-1546.

Kaselimi, M., Voulodimos, A., Doulamis, N., Doulamis, A., & Delikaraoglou, D. (2020). A Causal Long Short-Term Memory Sequence to Sequence Model for TEC Prediction Using GNSS Observations. Remote Sensing, 12(9), 1354.

Lin, C. C., Shen, M. H., Chou, M. Y., Chen, C. H., Yue, J., Chen, P. C., & Matsumura, M. (2017). Concentric traveling ionospheric disturbances triggered by the launch of a SpaceX Falcon 9 rocket. Geophysical Research Letters, 44(15), 7578-7586.

McKinnell, L. A., & Poole, A. W. (2004). Predicting the ionospheric F layer using neural networks. Journal of Geophysical Research: Space Physics, 109(A8).

Occhipinti, G., Rolland, L., Lognonné, P., & Watada, S. (2013). From Sumatra 2004 to Tohoku-Oki 2011: The systematic GPS detection of the ionospheric signature induced by tsunamigenic earthquakes. Journal of Geophysical Research: Space Physics, 118(6), 3626-3636.

Ruwali, A., Kumar, A. S., Prakash, K. B., Sivavaraprasad, G., & Ratnam, D. V. (2020). Implementation of Hybrid Deep Learning Model (LSTM-CNN) for Ionospheric TEC Forecasting Using GPS Data. IEEE Geoscience and Remote Sensing Letters. DOI: 10.1109/LGRS.2020.2992633

Sai Gowtam, V., & Tulasi Ram, S. (2017). An Artificial Neural Network-Based Ionospheric Model to Predict NmF2 and hmF2 Using Long-Term Data Set of FORMOSAT-3/COSMIC Radio Occultation Observations: Preliminary Results. Journal of Geophysical Research: Space Physics, 122(11), 11-743.

Srivani, I., Prasad, G. S. V., & Ratnam, D. V. (2019). A deep learning-based approach to forecast ionospheric delays for GPS signals. IEEE Geoscience and Remote Sensing Letters, 16(8), 1180-1184.

Şentürk, E., Livaoğlu, H., & Çepni, M. S. (2019). A Comprehensive Analysis of Ionospheric Anomalies before the Mw 7.1 Van Earthquake on 23 October 2011. The Journal of Navigation, 72(3), 702-720.

Vaishnav, R., Jacobi, C., & Berdermann, J. (2019). Long-term trends in the ionospheric response to solar extreme-ultraviolet variations. In Annales Geophysicae, 37(6), 1141-1159.