International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 04 | April - 2023 Impact Factor: 8.176 ISSN: 2582-3930

Prediction of Total Electron Content (TEC) variations using space weather data with AI

Tharun Racharla, Saurabh Sakharkar, Sakshi Rajurkar, Aman Mishra
Information Technology Department Tulsiramji Gaikwad Patil College of engineering and Technology Mohgaon, Maharashtra, India

Abstract—The effect of Space weather is usually linked to disturbances in the ionosphere (gradients in the Total electron content (TEC)). This has significant effect especially for GPS users causing degradation in range measurements, loss of lock by the receiver of the GPS signal. The ionospheric electron density is sensitive to the space weather parameters like sunspot numbers, F10.7, interplanetary magnetic field etc. and shows variation with respect to them along with other parameters like longitude, latitude, hour of day, etc. The variation in the TEC cannot be estimated or calculated, but can be predicted based on previous space weather data. In this study, we propose a novel ionospheric TEC forecasting model developed using Random Forest Regression algorithm. We have used observation data from the International reference Ionosphere (IRI) to model and forecast ionospheric TEC. The prediction model will be hosted on flask serer-web application.

Key words: Total electron content, Random Forest Regression, Global Positioning System, Artificial Intelligence, Machine learning, web application component.

I. Introduction

The ionospheric electron density is sensitive to the space weather parameters like sunspot numbers, F10.7, interplanetary magnetic field etc. and shows variation with respect to them as well as other factors like latitude, longitude, height, hour of day, etc. These parameters are detected and measured at L1 point (currently by ESA / NASA satellites) and are transmitted to earth. The abovementioned parameters affect the ionospheric TEC, which in turn shows abrupt variations and deviations from the nominal characteristics.

Therefore, these unconventionalities, which are typically observed during solar and geomagnetic storms, are necessary to be identified. This is also important, because the performance of the GNSS system, which is currently driving millions of applications, is also affected by the ionospheric variabilities.

The aim of the project is the prediction of the TEC variations providing a handle to take necessary actions to avoid or mitigate the resultant impairments arising out of the stated facts. The objective is to develop the proper AI algorithm to use the available space weather and other data to predict the TEC variations over Bengaluru, India.

In addition, the prediction model will be hosted on a website application to make the features of the model accessible and interactive for any user, especially other students. The goal of the website is to provide a user interaction with the prediction model as well as the awareness towards the domain of study.

II. Need for predicting the ionospheric TEC variations

A. *Global Positioning System (GPS)*

It is currently one of the most popular global satellite positioning systems due to global availability of signal as well as performance. GPS is a satellite-based navigation radio system which is used to verify the position and time in space and on the Earth. GPS nowadays allows to measure positions in real time with an accuracy of few centimeters (Warnant et al., 2007). The advent of GPS has led to technical revolutions in navigation as well as in fields related to surveying. The GPS system – an all-weather satellite-based radio navigation system – can provide users on a world-wide basis with navigation, positioning, and time information which is not possible with conventional navigation and surveying methods.

B. *Total Electron Content*

Total Electron Content: Total electron content (TEC) is an important descriptive quantity for the ionosphere of the Earth. TEC is the total number of electrons integrated between two points, along a tube of one meter squared cross section, i.e., the electron columnar number density. It is often reported in multiples of the so-called TEC unit, defined as TEC is significant in determining the scintillation and group and phase delays of a radio wave through a medium.

Ionospheric TEC is characterized by observing carrier phase delays of received radio signals transmitted from satellites located above the ionosphere, often using Global Positioning System satellites. TEC is strongly affected by solar activity [2].

C. F2 layer of Ionosphere

It is the highest permanently observable layer of the ionosphere. It exhibits a distinct maximum of free-electron density occurring at a height that ranges from about 225 km in the polar winter to over 400 km in daytime near the magnetic equator. Unlike the other ionospheric layers, the F2-layer tends to rise during the middle of the day, except at middle to high latitudes in winter. Its maximum electron density occurs during the day, its minimum usually just before sunrise. It is the layer that is most useful for long-range radio transmission.

D. Factors affecting TEC variations

In this work, the TEC is observed at the F layer because this region has the highest variability of free electrons, causing the greatest effect on GPS received signal compared to other layers. More than two-third of electron concentration are located at F2 layer. This method is conducted by going through several processes

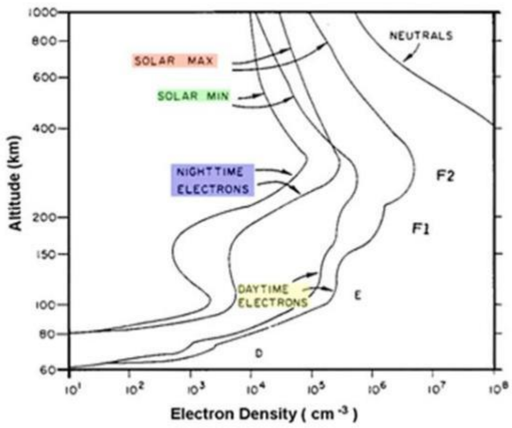
1) Latitude and longitude : The ionospheric TEC values change with respect to the location on the earth' s surface. The effect of solar activities changes as we move from equator to polar region. It also varies with high altitude areas like mountains or lower altitude areas like oceans. The locations can be specified using latitude and longitude values.

2) Time of observation: In the ionosphere, radiation from the sun (primarily X-rays, ultraviolet and particles) bombard gas molecules and cause them to release electrons. These are called "free electrons" and they are negatively charged. The molecules that lose the electrons become postively charged. The name for charged molecules or atoms is "ions" and these positive ions are what the ionosphere is named after. The free electrons and ions cause HF radiation that is moving upward to bend back toward the earth, an effect called refraction. This bending allows the signals to travel back down to the surface at ranges well over the horizon, giving HF radiation its unique long range characteristics. When radiation from the sun is no longer present during nighttime, many of the electrons join back with the ions, but some remain free throughout the night. An important parameter for EM propagation is the amount of free electrons that are present; this is called the "electron density". The higher the electron density, the more HF radiation is bent toward the surface.

The F layer is the highest significant layer in the ionosphere and the central part has the greatest electron density in the earth's atmosphere. This is the main region where HF radiation from below is bent back down. We use the term "reflection" in the table because the bending is like a reflection. Technically, this bending process is called refraction, but we can think of it as a reflection from the F layer. During the day the F layer separates into two layers, called F1 and F2. The F2 is the stronger and more important layer.

Layer	**Approximate Elevation**	**Importance**	**When Present**
F	140km – 400km	Main " reflection" region	Always– stronger during daytime

During the daylight hours, the electron densities in the E and F layers are about 100 times larger than at night. In addition to the changes in solar radiation due to the time of day, there are also changes in solar radiation due to changes in the sun itself. The sun undergoes an approximately 11 year cycle of changing radiation. The number of sunspots follows the same 11 year cycle; therefore the solar cycle is sometimes referred to as the sunspot cycle. During periods of solar maximum (most sunspots) the ionosphere has a higher electron density than other periods.



3) *Interplanetary magnetic field (IMF):* The interplanetary magnetic field (IMF) is a magnetic field that originates from the Sun and extends throughout the solar system. It is created by the solar wind, a stream of charged particles that flows from the Sun and carries the IMF with it. The IMF affects the total electron content (TEC) of the ionosphere, which is the total number of free electrons in a column of unit cross-sectional area from the Earth's surface to the top of the ionosphere. The IMF can affect TEC by altering the ionospheric electric field, which is responsible for the vertical drift of the plasma in the ionosphere. When the IMF is oriented in a southward direction, it can connect with the Earth's magnetic field and create a "magnetic reconnection" event, which can generate a large amount of electric field in the ionosphere. This electric field accelerates the plasma, causing it to move upwards and increase the TEC. This effect is most pronounced in the Polar Regions, where the IMF is most likely to be oriented in a southward direction. Conversely, when the IMF is oriented in a northward direction, it does not connect with the Earth's magnetic field as easily, and the electric field in the ionosphere is weaker. This causes the plasma to move downwards, decreasing the TEC. Overall, the IMF is an important factor in determining the behavior of the

ionosphere and the TEC, and is an important consideration for space weather forecasting and satellite communication systems.

4) *Ap and Kp indices*, The Kp and Ap indices are two important measures of geomagnetic activity that are used to monitor the Earth's ionosphere and its impact on radio wave propagation. They are both derived from measurements of the Earth's magnetic field taken at various geomagnetic observatories around the world.

The Kp index is a measure of the disturbance of the Earth's magnetic field caused by solar activity. It is a 3-hourly average of the standardized range of variation of the magnetic field measured at 13 different observatories. The Kp index ranges from 0 to 9, with higher values indicating greater magnetic activity.

The Ap index is a similar measure of geomagnetic activity, but is based on the daily average of the magnetic field measurements from a larger number of observatories. The Ap index ranges from 0 to over 400, with higher values indicating greater geomagnetic activity.

Both the Kp and Ap indices are used to monitor the impact of solar activity on the Earth's ionosphere, including the Total Electron Content (TEC). TEC is the total number of free electrons in a column of unit cross-sectional area from the Earth's surface to the top of the ionosphere. The ionosphere is an important layer of the Earth's atmosphere that is ionized by solar radiation, and plays a key role in radio wave propagation. When the Kp and Ap indices are high, the Earth's magnetic field is disturbed by solar activity, which can result in changes in the TEC. High levels of geomagnetic activity can cause ionospheric irregularities, which can affect the propagation of radio waves through the ionosphere. This can lead to disruptions in satellite communication systems, GPS navigation, and other technologies that rely on radio wave propagation. Overall, the Kp and Ap indices are important measures of geomagnetic activity that are closely monitored by scientists and engineers to better understand and predict the impact of solar activity on the Earth's ionosphere and its effect on radio wave propagation.

III. Data, Model and Methodology

A. Data Collection

Collecting data for training the ML model is the basic step in the machine learning pipeline. The predictions made by ML systems can only be as good as the data on which they have been trained.

The International Reference Ionosphere (IRI) is an international project sponsored by the Committee on Space Research (COSPAR) and the International Union of Radio Science (URSI). These organizations formed a Working Group (members list) in the late sixties to produce an empirical standard model of the ionosphere, based on all available data sources (Charter). Several steadily improved editions of the model have been released. For given location, time and date, IRI provides monthly averages of the electron density, electron temperature, ion temperature, and ion composition in the ionospheric altitude range. The template is designed so that author affiliations are not repeated each time for multiple authors of the same affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization). This template was designed for two affiliations.

The dataset will be prepared using the above website containing features: - year, month, hour of day, height in kilometers,

1) *Web Scraping:*

Web scraping, web harvesting, or web data extraction is data scraping used for extracting data from websites. Web scraping software may directly access the World Wide Web using the Hypertext Transfer Protocol or a web browser. While web scraping can be done manually by a software user, the term typically refers to automated processes implemented using a bot or web crawler. It is a form of copying in which specific data is gathered and copied from the web, typically into a central local database or spreadsheet, for later retrieval or analysis [3].

2) *Scrapy framework:* - Scrapy is a fast high-level web crawling and web scraping framework, used to crawl websites and extract structured data from their pages. It can be used for a wide range of purposes, from data mining to monitoring and automated testing.[1].

B. Machine Learning methodologies

1) *Machine Learning:* Machine learning (ML) is the study of computer algorithms that allows computer programs to automatically improve through experience. This makes it possible to generate predictions without any pre-defined rules or calculation instructions. The basic principle of machine learning is to create an ML model out of example data by using learning algorithms. The ML model represents the machine learning artifact that encodes the decision or prediction logic. The example data are also referred to as training data. The more training data the learning algorithm receives, the more it can improve the ML model and reduce its error rate. A property of the training data, such as a column name is called feature.

2) *Random Forest Regression algorithm*, Random Forest Regression is a machine learning algorithm used for regression tasks, where the goal is to predict a continuous numerical output. It is an extension of the decision tree algorithm, where multiple decision trees are created and combined to make more accurate predictions. The algorithm works by creating a large number of decision trees, each trained on a random subset of the data, and making a prediction by aggregating the predictions of each individual tree. During the training process, each decision tree in the forest is grown by recursively splitting the data into subsets based on the most significant features, until a stopping criterion is met. This process is repeated for each decision tree, resulting in a diverse set of trees that capture different aspects of the underlying relationships between the input and output variables. To make a prediction, the algorithm passes the input data through each decision tree in the forest, and each tree

produces a prediction. The final prediction is then obtained by averaging or taking the median of the predictions made by all the trees in the forest. Random Forest Regression has several advantages over other regression algorithms. First, it can handle a large number of input features and can identify the most significant features for making predictions. Second, it is less prone to overfitting than a single decision tree, as the aggregation of multiple trees reduces the impact of outliers and noise. Finally, it is computationally efficient and can handle large datasets.

IV. Website and model hosting

A Python Flask is employed to deploy the machine learning models in a website framework. The user first inputs the parameters on the GUI built using HTML, CSS to be sent as a query to the Flask framework. Flask will then use the trained model to predict values or labels on the input parameters and output the predicted results on the website page.

Hosting a web application using the Python Flask library is a relatively straightforward process. Flask is a lightweight web framework that allows developers to quickly build and deploy web applications using Python. In a research context, Flask can be used to build and deploy web applications that allow researchers to share and collaborate on their work.

Here are the basic steps to host a web application using Flask:

a) Build the web application: The first step is to develop the web application using Flask. This involves writing the code for the application and defining the routes and endpoints that users will interact with.

b) Test the web application: Once the application is built, it should be tested to ensure that it is functioning properly. This can be done by running the application on a local server and testing it using a web browser.

Overall, hosting a web application using the Python Flask library can be a powerful way to share and collaborate on research projects. With its ease of use and flexibility, Flask is a popular choice for building and deploying web applications.

Number of max leaf nodes	Mean_Squared_Error	R^2 score
3000	1.1688765448786531	0.9990046905587582
3500	1.112044336276327	0.9990490600317836
4000	1.0649238349366144	0.9990805890490252
7000	0.9271443318739999	0.9991565253774405
8000	0.9048686193386523	0.9991662647332936
11000	0.8600974409766134	0.9991830853362879
14000	0.8341831606984393	0.9991908751794039
17000	0.8185083334923904	0.9991952006940032
20000	0.8076370726431922	0.9991977025996497

The results represent the mean absolute error and R^2 scores for different number of maximum leaf nodes in a machine learning model trained using random forest regression. As the number of maximum leaf nodes increases from 3000 to 20000, the mean absolute error decreases from 1.16 to 0.81, and the R^2 score increases from 0.999 to 0.999. The lowest mean absolute error and highest R^2 score are observed at 20000 maximum leaf nodes.

This suggests that increasing the number of maximum leaf nodes improves the performance of the model. However, it's important to note that there may be a trade-off between model performance and computational complexity, as increasing the number of maximum leaf nodes may increase the model's training time and memory requirements. Therefore, it's important to carefully choose the appropriate number of maximum leaf nodes that balances model performance and computational efficiency.

Acknowledgment

In this study, an ionospheric forecast model in India is established based on 14 Indian city locations through a website application. Compared with empirical models, deep learning model have higher prediction accuracy, especially during magnetic storms. In this paper, the attention mechanism is added to the LSTM neural network, which inherits the high precision of the deep learning model and effectively improves stability. Currently, it should be noted that our experiments only discuss a relatively quiet solar period. We acknowledge the increasing solar activity levels and suggest that measurements in the coming years, where activity is expected to be much stronger, will provide a good opportunity for continued model development.

References

[1] Scrapy documentation, website- https://docs.scrapy.org/en/latest/.

[2] SPACE WEATHER PREDICTION CENTER NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION , https://www.swpc.noaa.gov/phenomena/total-electron-content#:~:text=The%20Total%20Electron%20Content%20(TEC,radio%20signal%20will%20be%20affected.

[3] Web Scraping, https://en.wikipedia.org/wiki/Web_scraping.

[4] Neural networks, https://www.ibm.com/cloud/learn/neural-networks.

[5] Tang, J.; Li, Y.; Ding, M.; Liu, H.; Yang, D.;Wu, X. An Ionospheric TEC Forecasting Model Based on a CNN-LSTM-Attention Mechanism Neural Network. Remote Sens. 2022, 14, 2433. https://doi.org/10.3390/rs14102433

[6] Zettergren, M. D., & Snively, J. B. (2019). Latitude and longitude dependence of ionospheric TEC and magnetic perturbations from infrasonic-acoustic waves generated by strong seismic events. Geophysical Research Letters, 46, 1132– 1140. https://doi.org/10.1029/2018GL081569.

[7] Prediction of Global Ionospheric TEC Based on Deep Learning https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2021SW002854

[8] TEC by se weather prediction center https://www.swpc.noaa.gov/phenomena/total-electron-content#:~:text=The%20Total%20Electron%20Content%20(TEC,radio%20signal%20will%20be%20affected.

[9] HF and Lower Frequency Radiation - The Ionosphere and the Sun https://www.met.nps.edu/~psguest/EMEO_online/module3/module_3_2.html.

[10] F10.7 CM RADIO EMISSIONS https://www.swpc.noaa.gov/phenomena/f107-cm-radio-emissions.

[11] M Themens, D.R.; Reid, B.; Jayachandran, P.T.; Larson, B.; Koustov, A.V.; Elvidge, S.; McCaffrey, A.M.;Watson, C. E-CHAIM as a model of total electron content: Performance and diagnostics. Space Weather 2021, 19, e2021SW002872. [CrossRef]