Global TEC forecasting based on deep learning techniques: a comparative study and perspectives for a Space Weather operational service

María Graciela Molina [1,2,3,4], Jorge H. Namour[1,2] , Claudio Cesaroni[4,1], Luca Spogli[4,5], Noelia B. Argüelles [1,2,3] , Eric N. Asamoah [4,6]

[1]Tucumán Space Weather Center (TSWC), Facultad de Ciencias Exactas y Tecnología (FACET), Universidad Nacional de Tucumán (UNT), Av. Independencia 1800, Tucumán, Argentina.
[2]Laboratorio de Computación Científica, FACET- UNT.
[3]Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina.
[4]Istituto Nazionale di Geofisica e Vulcanologia, Via di Vigna Murata 605, 00143, Rome (Italy)
[5]SpacEarth Technology, Via di Vigna Murata 605, 00143, Rome (Italy)
[6]University of Salento, Via Piazza Tancredi, n7, 73100, Lecce (Italy)

Key Points:

- Three different deep learning techniques are compared to perform 24 hs forecasting of global TEC
- We show results for quiet and perturbed conditions in three longitudinal sectors of the globe covering different environmental conditions.
- We present a strategy to implement the best model in real time with the perspective of an operative Space Weather product.

Plain Language Summary

Operational space weather ionospheric forecasting requires real-time measurements, efficient computation and, fast and accurate models. In particular, data-driven models such as machine learning models can help train smart models able to catch most of the global ionospheric variability. Nevertheless, machine learning models struggle with generalization and with unbalanced datasets. Space weather is an intrinsically unbalanced problem with few extreme cases within a given solar cycle. We present a machine learning strategy to learn the changes

in the ionosphere (e.g. during the evolution of a geomagnetic storm) as new data arrives while few computation resources are required. We also compared three deep learning techniques using global TEC as input plus a geomagnetic index (Kp) as the external forcing to forecast 24 hs in advance. The results showed that the best method is CNN so we implemented it using incremental training toward an operational version of the forecasting model.

Abstract

The aim of this work is to present a global ionospheric prediction model based on deep learning (DL) to forecast Total Electron Content 24 hours in advance under different space weather conditions. Three different DL techniques have been compared to select the most suitable for the purpose of an operational service: Long Short Term Memory (LSTM), Gated Recurrent Units (GRU) and Convolutional Neural Networks (CNN). The modeling approach inherits and extends what has been proposed by Cesaroni and co-authors (2020). We use TEC on 18 selected grid points of Global Ionospheric Maps (GIMs) as the target parameter and Kp index as the external input. We use a dataset from 2005-2016 for training and testing, we also analyze case studies from 2017 under different geomagnetic conditions. Results show that CNN models have better predictive capabilities than the other two DL models, even under geomagnetically disturbed conditions. Considering the first 24 hours of forecasting, CNN exhibits errors between 0.5 and 2 TECu, while LSTM and GRU errors can reach 3 TECu. We also show how all the proposed models outperform the two naive models: the so-called "frozen ionosphere" and a 27 days averaged model.

Moreover, we implemented the models using incremental training to update them as new data arrives and thus the trained model is able to adapt to rapid changes within the previous 24 hs to the forecasting. Thus, the proposed model can be implemented in an operative manner for Space Weather applications and services.

Introduction

Space Weather is a complex domain composed of highly coupled subdomains involving very different time and spatial scales with events triggered within the Sun, evolving through the interplanetary medium that can reach the Earth's environment. These events may have a severe impact on the ionosphere because of the complex solar wind-magnetosphere-ionosphere coupling with different consequences in different regions of the globe.

Space Weather thrives on the increase of available data due to the new space mission launched in the last few years (e.g. Parker Solar Probe, Solar Orbiter, etc), the enhancement of instruments deployment and networks (e.g. new Global Navigation Satellite System, GNSS, networks, magnetometers, etc).

In the last decade, the need to move from research to forecast and mitigation services, like those supporting GNSS-based positioning and telecommunications, dramatically increased (see, e.g., Lilensten et al., 2021). Nowadays, the possibility to properly customize those services to the requirements of specific user communities and stakeholders resulted in several operational services, capable of providing decision tools to tackle the ionospheric disturbances in critical domains (Veettil, 2019). This is the case of the Space Weather Service provided to the International Civil Aviation Organization (ICAO), providing advisories according to its Standards and Recommended Practices based on scientific products covering HF communication and GNSS navigation domains (Kauristie et al., 2021). Reliable nowcasting and forecasting algorithms are the core tools of these services.

Recently many efforts have been made to enhance Space Weather monitoring and forecasting capabilities of ionospheric effects due to Space Weather phenomena by using different approaches, models and instrumentations (Zolesi & Cander, 2014; Bilitza et al., 2017; McGranaghan et al., 2018; Perez, 2019; Galkin et al., 2020, among many more).

Concerning forecasting, currently, two different approaches have been developed: (i) physical or semi-empirical models with a data assimilation scheme (see., e.g, Galkin et al., 2012); (ii) completely data-driven models (Bilitza et al., 2022).

The latter includes models based on the use of Artificial Intelligence (AI) that are becoming more and more important thanks to the development of more accurate and faster Machine Learning algorithms (Camporeale et al 2018, Camporeale 2019, among others). In the ionospheric forecasting domain, this class of models have been introduced since the end of the ’90s, mainly leveraging on the use of Neural Network (NN) approaches, and have been dedicated to forecasting ionospheric features on different spatial/temporal scales (e.g., Zolesi & Cander, 2014). After an AI winter in the first decade of this century, the ionospheric community started exploiting recently-introduced Deep Learning (DL) techniques.

Yi Han et. al. 2021, devoted their time to investigate how to forecast ionospheric TEC values in three IGS GNSS monitoring stations at the low-latitude region in high solar activity and magnetic storm periods. In their work four different machine learning models including artificial neural network, long short-term memory networks, adaptive neuro-fuzzy inference system based on subtractive clustering, and gradient boosting decision tree (GBDT) were

used during high solar activity and magnetic storm periods. The numerical results indicate that the ionosphere prediction accuracy in high solar activity is not as good as in the low solar activity period, but the machine learning based approach still significantly outperforms the GIM and IRI-2016 models.

According to Zhou Chen et. al., 2022 several different LSTM-based algorithms were tested and a multi-step auxiliary algorithm was developed which performed best. The proposed algorithm predicted the global ionospheric IGS-TEC in the next 6 days (the MAD and RMSE are 2.485 and 3.511 TECU, respectively).

In the works of Liu et. al., 2020 LSTM neural network (NN) was employed to forecast the 256 spherical harmonic (SH) coefficients that are used to construct global ionospheric maps (GIM). In order to train the LSTM NN, multiple input data including historical time series of the SH coefficients, solar extreme ultraviolet (EUV) flux, disturbance storm time (Dst) index, and hour of the day are used. Then the best performing LSTM model is used to forecast the SH coefficients, and the global hourly TEC maps are reproduced using
the 256 predicted SH coefficients. The results show that the first/second hour TEC root mean square error (RMSE) during storm time is 1.27/2.20 TECU and 0.86/1.51 TECU during quiet time, indicating that the developed model performs well during both quiet and storm times and common ionospheric structures such as equatorial ionization anomaly (EIA) and storm-enhanced density (SED) are accurately represented in the anticipated TEC maps.

In the devoted work of Zewdie et al. (2021), they presented a data-driven forecasting of ionospheric total electron content (TEC) using Long-Short Time Memory (LSTM)method. In the process of selecting the input parameters to train the algorithm, they made use of random forest to perform regression analysis and estimate the importance of input parameters. Relative importance of 34 different parameters including the solar flux, solar wind density, and speed the three components of interplanetary magnetic field, Lyman-alpha, the Kp, Dst and Polar Cap (PC) indices were analyzed. The LSTM method was applied to forecast the TEC up to 5 h ahead, with 30-min cadence.

Mallika, L. et al in 2020, devoted their time to build ML algorithm to forecast ionospheric time delays using GNSS observations. The implement Gaussian Process Regression (GPR) utilizing its kernel based-approach and Bayesian rules. This GPR model performed better than univariate linear model (ARMA) and the non-linear ANN during both geomagnetic quiet and geomagnetic disturbed days.

Our work aims to compare different ML techniques for global ionospheric forecasting from a space weather perspective. The main focus is on tree DL techniques namely Long-short term

memory (LSTM), Gated Recurrent Unit (GRU) and Convolutional Neural Networks (CNN), having in mind a Neural Networks Non-Autoregressive with Exogenous values (NNARX) scheme (Cesaroni et al, 2020 and references within). Several specific domain characteristics are considered to implement loose physics-informed DL forecasting. The forecasting horizon is 24 hours and the external input is given by the Kp index, which is able to depict different conditions of the geomagnetic disturbance at the global level. As a future operative AI-based tool, we propose a method to update the models as new data arrives bearing in mind the computational time to run the model in a real-time scenario.

This work also aims at improving the TEC forecasting approach introduced by Cesaroni et al. (2020), which works in two main stages: 1) single-point forecasting based on ML, and 2) extending the forecasting for any point in the globe. Thus, the objective is to improve the first stage of the proposed technique by Cesaroni et al. (2020) and the best-performing DL technique will be used later in the second stage.

The paper is structured as follows: after the introduction, the section "Data and Methods" presents a data science-based methodology (including details of each stage), and describes the datasets used for the modelling and the software tools. The section "Modeling using deep learning" describes the proposed DL architectures or models. Then the "results" section includes results for the training process, the validation using actual data in the test set and in cases of study, and the results using incremental training. Finally, there is a discussion and conclusion section.

Data and Methods

Due to the ubiquity of GNSS receivers, GNSS-derived vertical Total Electron Content (vTEC) (hereafter referred simply as TEC), is used as the parameter to monitor the ionosphere on a global scale, by assuming the ionosphere as a single shell located at the peak of the F2-layer (see e.g. Ciraolo et al., 2007). The Kp index is assumed as the geomagnetic index driving the external forcing (Menvielle and Berthelier, 1991). These two features will be the input to the DL models.

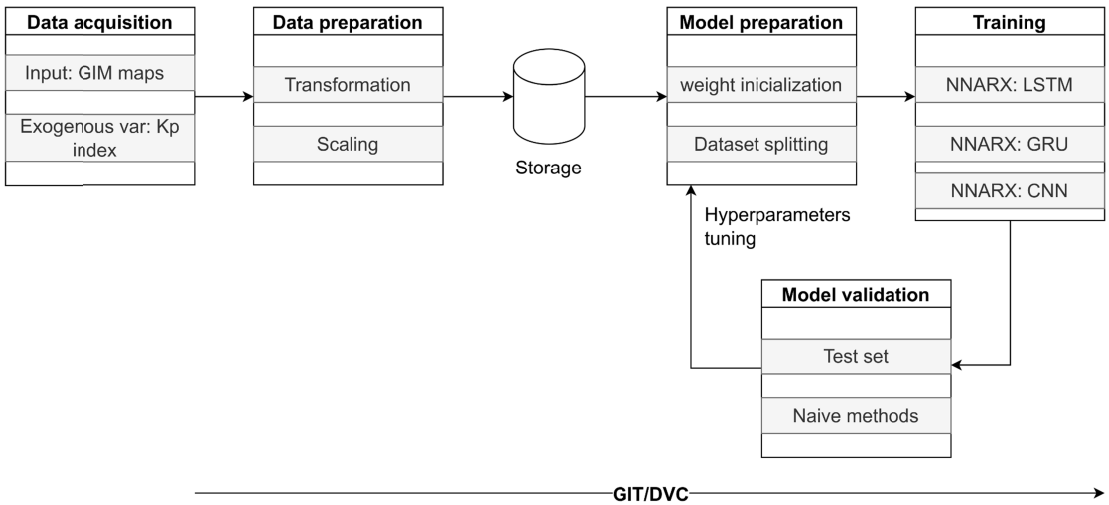


Figure 1. Scheme of the data pipeline for the proposed deep learning models.

We follow a data pipeline flow as in Figure 1 and we implemented the models using Python and Keras.

A data pipeline flow aims at providing a tool for the experiment to be easily scalable, reproducible and able to tackle the increasing amount and complexity of data. In our case, the data and code flow is tracked and versioned by using two tools: GIT for the code and DVC (Data Version Control) for the data, models and metrics. These 2 tools provide consistency and traceability during the modeling and training process.

The methodology involves two main parts: (a) Data acquisition, preparation and storage, and (b) modelling which involves the preparation, training and validation as an iterative process.

At first (a), the pipeline starts at the acquisition process which, in this case, is straightforward since the data is obtained offline from different sources. We used TEC obtained from Global Ionospheric Maps (GIM) maps provided by the International GNSS Service (IGS) with a spatial-temporal resolution of 2.5° (lat) - 5° (lon) - 2 hs in IONEX format (see Schaer et al., 1998; Hernández-Pajares et al., 2009). We selected 18 grid points globe (Fig. 2) to coverdifferent latitudes and local time sectors. This follows from Cesaroni et al. (2020), for comparison purposes. It is worth mentioning that, in a real-time scenario, the acquisition step may change but the rest of the stages should be kept the same.

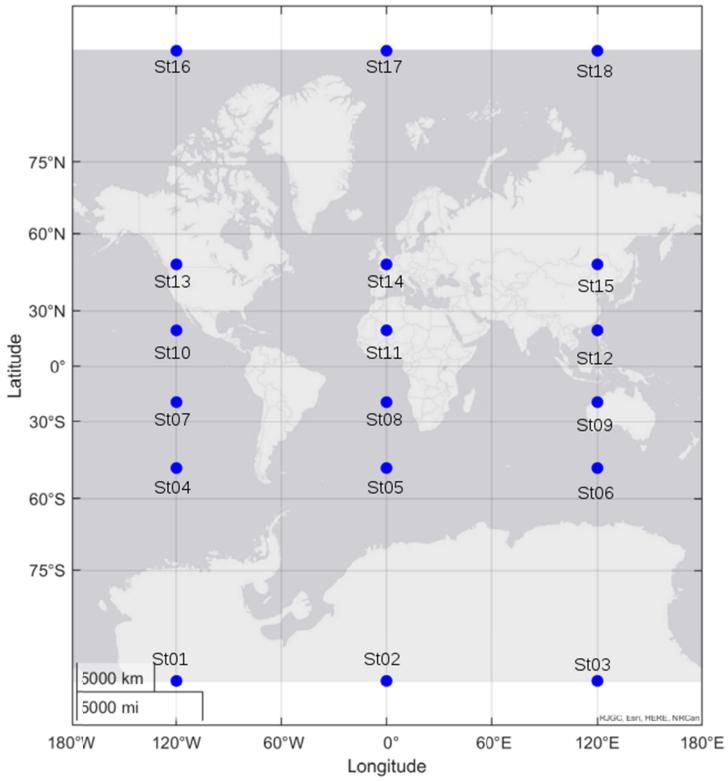


Fig 2. Bluedots in the map show the selected virtual station in three meridional sectors covering different latitudinal regions (adapted from Cesaroni et al., 2020).

As reported in Figure 2, the grid points (or virtual stations,dots) are distributed to cover high-, mid- and low-latitudes in three longitudinal sectors. Additionally, different environmental conditions are considered, such as lands, deserts or oceans. Geographic coordinates are reported in table 2 of Cesaroni et al. (2020).

As already mentioned above, we also use the Kp index to provide the deep-learning models with information about the geomagnetic conditions. The Kp index is a time series of 3 hours time resolution provided by the World Data Center for Geomagnetism, Kyoto (wdc.kugi.kyoto-u.ac.jp).

The model scheme follows the so-called NNARX scheme or nonlinear autoregressive neural network with external input (Nørgård et al., 2000). Figure 3 shows the NNARX scheme where the input is the TEC time series (for each grid point) and the exogenous variable is the Kp index. We use the information from the last 24 hours (12 TEC and 8 Kp time steps) to forecast TEC 24 hours ahead. That is a multivariable multi-step scheme (many steps to many steps) (also in Figure 3).

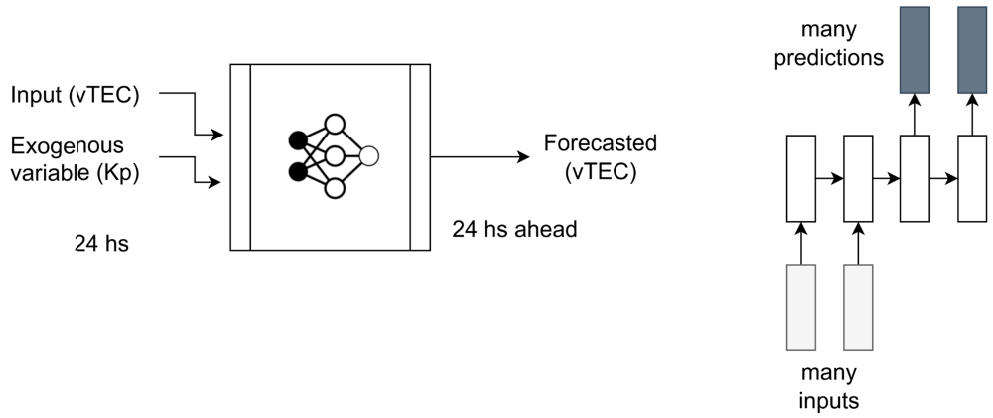


Figure 3. NNARX scheme and many-to-many steps prediction using deep learning.

Once the data is acquired from two kinds of file sources, IONEX format for GIM and CSV format for Kp index, we proceed with the data preparation stage. This includes two parts: (i) managing different data resolutions and (ii) integrating the data in a single format (JSON). We interpolated the Kp index using K-nearest neighbours (Abu Alfeilat et al., 2019) (sensitive to the local structure of the data) to upsample to 2 hours resolution. Thus, after the data transformation process, we obtain a dataset composed of 18 TEC independent time series (one for each grid point) and a Kp index time series, each of them with a time resolution of 2 hours. Another important operation over the dataset during the preparation is the data scaling. In our particular case, we normalized the datasets since they do not have a strictly Gaussian distribution which is a standard procedure in machine learning (ML). Then, the datasets are stored in a database to be used in the modelling. We used data ranging from 01/Jan/2005 to 31/Dec/2016. Also, to study the performance under different geomagnetic conditions we used selected cases from 2017.

The data-driven modelling stage (b) starts by splitting the data and defining a specific weights initialization, both as part of the model preparation. In our proposed models, we used a data splitting of 99% for the train set (which in turn is also divided into 99% train and 1% validation), and 1% for the test set. We have experimented with other splitting ratios for train, validation and testing but we decide on the mentioned ratios due to the intrinsic characteristics of the datasets. As observed in Figure 4, the TEC time series for the

considered years, and for the grid point named St3 (station 3), has a well-known modulation following the solar cycle (Mendillo, 2006). During years of minor solar activity, TEC maximum values reached ~20 TECu. While during years of high solar activity, TEC can easily reach values above 40 TECu. This behaviour is observed in any of the TEC time series (at every grid point) (not shown here). This is almost a structural break in the time series which makes the learning process difficult. This is especially hard when using a data splitting ratio such as 70/30 because few samples corresponding to the high activity will be used during the training (affecting the data balancing in the training set), and thus this high activity TEC will be hard to learn by the neural network. Another consideration in the splitting process is that because the dataset is a time series, it can not be randomly split but needs to maintain the order of the samples (there is a time dependency among the samples).

Another way to tackle this problem can be: using fewer (last) years of data (study of short-term predictions) or adding at least another complete solar cycle of data. We have trained with fewer samples (e.g., 1 year) (not shown in this work) with poor performance. The low performance was expected mainly because of two reasons: a) relatively few samples per year (4380) for a deep learning technique and b) a low number of representative cases (e.g. few space weather events, solar cycle is not considered, etc). On the other hand, the option of adding more than one solar cycle of data is not possible since we have had data available since 1999.

For the above reasons, we considered the data splitting as 99% training (%99 for train, 1% validation) and 1% for the test set.

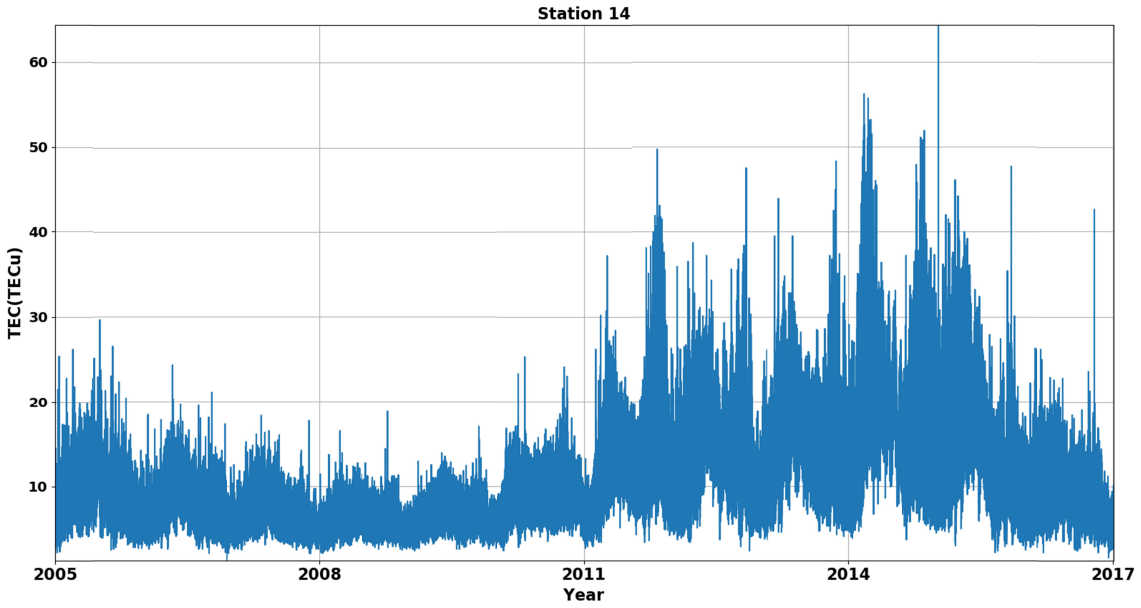


Figure 4. TEC time series for grid point St14 (50° N, 0° E) from 2005 to 2017.

We also considered that the datasets do not follow a normal distribution and, hence, the initialization of the weights should be set following a more adequate probability distribution function. An intelligent weight initialization can enhance the learning capabilities of deep learning models. Figure 5 shows a histogram of the TEC values at every grid point from 01/Jan/2005 to 31/Dec/2016. The distribution follows a left-lagged truncated gaussian distribution. In such cases, a Glorot normal distribution can be selected for the weight initialization as it was shown by Glorot and Bengio (2010) that this distribution in combination with a proper activation function can enhance the learning capability by preventing vanishing gradient problems. The main idea is to find a good variance for the distribution from which the initial parameters are drawn, to ensure the gradients have significant effects on the weights of the initial layers. This variance is adapted to the activation function used and is derived without explicitly considering the type of distribution. We will address this concept again in the following section corresponding to the different architectures when we explain the activation function used in each case. The Glorot normal distribution is implemented in Keras and can be easily set as one of the hyperparameters before the training.

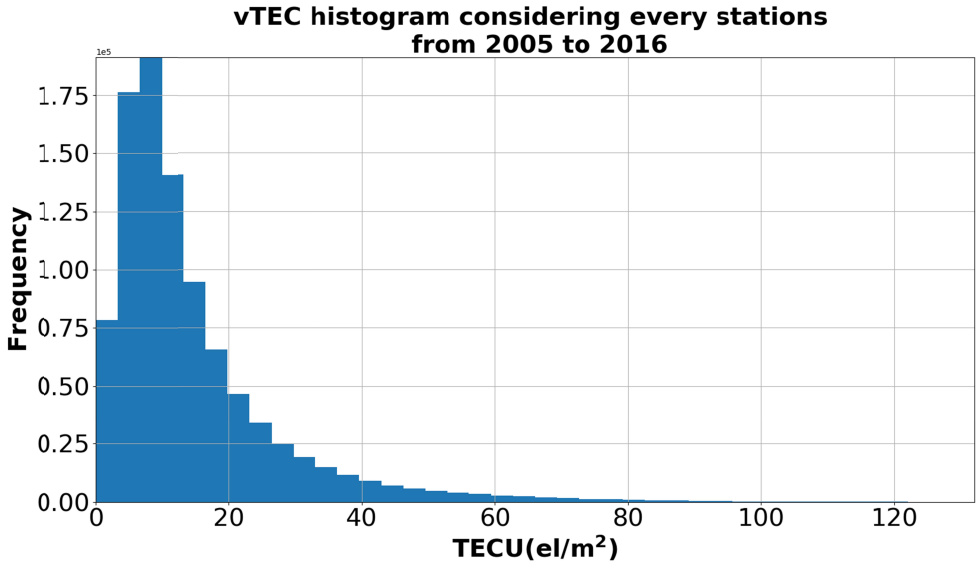


Figure 5. Histogram of the total TEC samples for all the grid points ranging from 01/Jan/2005 to 31/Dec/2016

Regarding the models, we performed 18 separate models using three different DL techniques at each grid point: Long short term memory (LSTM), Gated recurrent unit (GRU) and Convolutional Neural Networks (CNN). Both LSTM and GRU neural networks are Recurrent Neural Networks (RNNs). In the following section, we discuss and explain the used architectures. Concerning the metrics for both the learning process and the performance evaluation of the models, we use the basic matrics Root Mean Square Error (RMSE) or Mean Absolute Error (MAE), as suggested for fit performance by Liemohn et al. (2018)

We also compare the performance against the so-called "frozen ionosphere" which is considered a naif forecasting model, also referred as "recurrence test". It is based on the assumption that the forecasted ionospheric condition is equal to the actual measured value one day before at the same hour (Cesaroni et. al., 2020). Additionally, we considered another naive method based on the last averaged 27 days which often is used as a reference curve in ionospheric studies (Forbes et al., 2000).

A final remark on the general data pipeline is related to hyperparameter tuning. The parameters within the ML model that are not trained but need to be selected (e.g. number of layers, number of cells, activation function, etc.) are referred to as hyperparameters. Their selection can be done by simple trial/error procedure (sometimes aided by the domain knowledge) or by different techniques. In this work, we report only the best architectures resulting from the hyperparameter tuning by doing a grid search. The hyperparameters selected for the grid search are the number of cells, number of layers, number of epochs and batch size. The reason behind this choice is a trade-off between the tuning computational time and the accuracy obtained. In future developments, we will consider other hyperparameter tuning strategies (e.g. evolutive algorithms, memetic algorithms, among others).

The datasets do not have significant outliers and have not values either. This is due to the fact that TEC obtained from GIM and Kp index are values resulting from modelling and not actual measurements (Orus et al.,2003;Matzka et al., 2021n).

It is worth mentioning that the TEC and Kp data sets are intrinsically unbalanced due to the low probabilities of extreme space weather events within a solar cycle (see, e.g., Nikitina et al., 2022).

Nevertheless, we have almost a complete solar cycle of data to train our models with a good overall performance as it will be shown in the following sections

Data and code used in this work is publicly available in https://doi.org/10.5281/zenodo.7817642 and https://github.com/Laboratorio-Computacion-Cientifica/Global-TEC-forecasting-for-space-weather-application-based-on-deep-learning-techniques

Modeling using deep learning

Our models for global ionospheric forecasting are based on three different deep learning networks (LSTM, GRU and CNN). We use 24 hours TEC samples as input and 24 hours Kp values (as the external forcing) for each of the 18 selected grid points. Thus for each technique, we obtain 18 independent models .

RNNs are a class of artificial neural networks that enables the passing and sharing of information along a temporal sequence. RNNs can learn temporal dynamic behaviour and thus they are one of the most used DL techniques for forecasting time series. RNNs use a *state* or *memory* (h_t) that is updated at each time as a sequence is processed using the same parameters each time step (see Figure 6). RNNs can manage variable-length sequences as input, track long-term dependencies (especially in some architectures such as LSTM), and maintain information about the order. The main drawback of RNNs is the vanishing and exploding gradient problems caused by the long-term dependencies within the network. Both LSTM and GRU are RNNs

The LSTM (Hochreiter and Schmidhuber,1997) more significant characteristic is its ability to decide which is the most relevant information from the past that should be propagating ahead in the sequence. This is done by maintaining a state cell and a group of gates to control the flow of the information, they are: (a) forget gate in charge of getting rid of irrelevant information, (b) input gate, where the current relevant information is stored, (c) the update gate that selectively performs the update of the cell state, and (d) the output gate that returns a filtered version of the cell state. Each gate output is passing through different activation functions. The capability of maintaining and selecting significant long-term memory (and short-term dependencies) makes LSTM one of the most used DL techniques in forecasting time series. However, LSTM may exhibit some disadvantages or pitfalls in certain cases such as the tendency to overfit. Also, LSTM, typically, is able to catch (learn) smooth variations within the time series and may experience some difficulties to learn fast changes and reach maximum or minimum time series values (peaks). As with any deep learning technique, its performance can be diminished when using small datasets, less complex problems or

imbalanced datasets.

GRU (Cho et.al, 2014) can be seen as a simplified variant of LSTM. The network also has gates to pass information through the sequence. GRU also uses gating units that modulate the flow of information inside the unit, however, without having separate memory cells. It has 2 gates: (a) the reset gate, which provides a mechanism to forget less relevant information from the past; and (b) the update gate, to decide how much information from the past needs to be passed to future states. GRU, however, does not have any mechanism to control the degree to which its state is exposed (as it is in LSTM) but exposes the whole state each time. Complete detail on how LSTM and GRU internally manage time dependencies and the vanishing gradient problem is out of the scope of this work but for more details on this topic see e.g. Chung et.al., 2014.

The third DL technique used in this work is CNN in one dimension (1D). CNNs are a class of artificial neural networks able to learn from spatial information and thus they are widely used in computer vision problems. CNNs work by applying a cascade of operations that include: (a) convolutional operations (filters trained using the train set) that generate the so-called feature map, followed by a non-linearity operation (often using ReLU activation function); (b) a pooling layer is applied to the feature map to downsample it but keeping the spatial invariance (spatial information) while reducing the dimensionality; (c) stages (a) and (b) can be repeated depending on the problem and dataset; (d) after the last pooling layer, a dense layer is in charge, finally, of the prediction. Typically, CNN inputs are images and this is why the spatial information is important to be transmitted through the CNN, they are not able (in principle) to learn time dependencies. Nevertheless, CNN can be adapted to time series and a CNN 1D version is used in this work (Kiranyaz et. al., 2021). Figure 7 shows a simple scheme on the CNN 1 D applied for TEC forecasting.

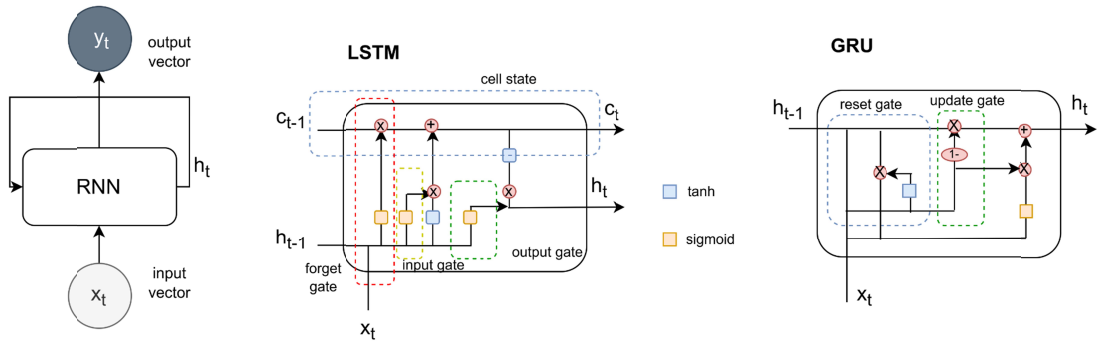


Figure 6. Simple scheme of RNN, LSTM and GRU

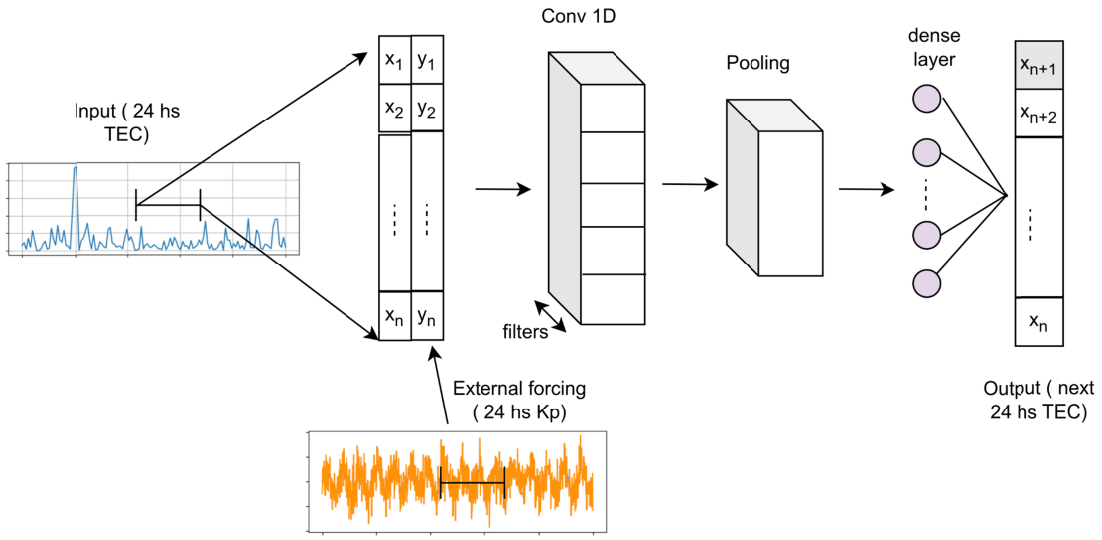


Figure 7. A simple scheme of CNN 1D. The input is a 24 hs TEC time series plus the external forcing (Kp) also for 24 hs, while the output is a 1D time series corresponding to the forecasted 24 TEC.

In summary, we propose three different DL techniques applied in an NNARX scheme (TEC as input and Kp as external forcing) to forecast TEC 24 hs in advance at each of 18 grid points in the globe. The proposed architectures for LSTM, GRU and CNN 1D, are reported in table 1. The presented architecture corresponds to the best performance after the grid search.

	LSTM and GRU	CNN
complete dataset	52560 samples	52560 samples
train set (99%)	52035 samples (4336 days)	52035 samples (4336 days)
test set (1%)	525 samples (43 days / 0.01)	525 samples (43 days / 0.01)
details	activation 1 = tanh activation 2 = tanh	Kernel size =2 conv layer = 1 pooling layer = 1 flatten layer = 1 dense =1 filters = 32 activation = ReLU
step in (24 hs)	12	12
step out (24 hs)	12	12
optimizer	adam	adam
cells	5	-
loss	MSE	MSE
epochs	20	20
batch	32	32

Table 1. Architectures for the LSTM, GRU and CNN models

Results

In this section we show the results obtained for the three methods for the test set, selected

cases of study, and we also analyze the training process.

Training

Figures 8a, 8b and 9 show the training process for each of the used methods (LSTM, GRU, CNN) and for every single virtual station (St 01, St 02, …, St 18). This grid of figures shows the loss function for the validation (orange line) and training (blue line) over 20 epochs for each of the 18 virtual stations. The panels are placed from west to east (left to right), and north to south (up to down) in such a manner that each station is placed in agreement with the map of virtual stations (Figure 2)

Typically, when the loss function for the validation curve is approaching or is close to the loss function for the training curve, it means that the training is successful. Here this behaviour can be observed for up to 20 epochs or iterations (Fig 8a, 8b, and 9). If at some point the validation curve starts to grow and turning away from the training curve (that keeps decreasing) is an indication of overfitting (not happening here).

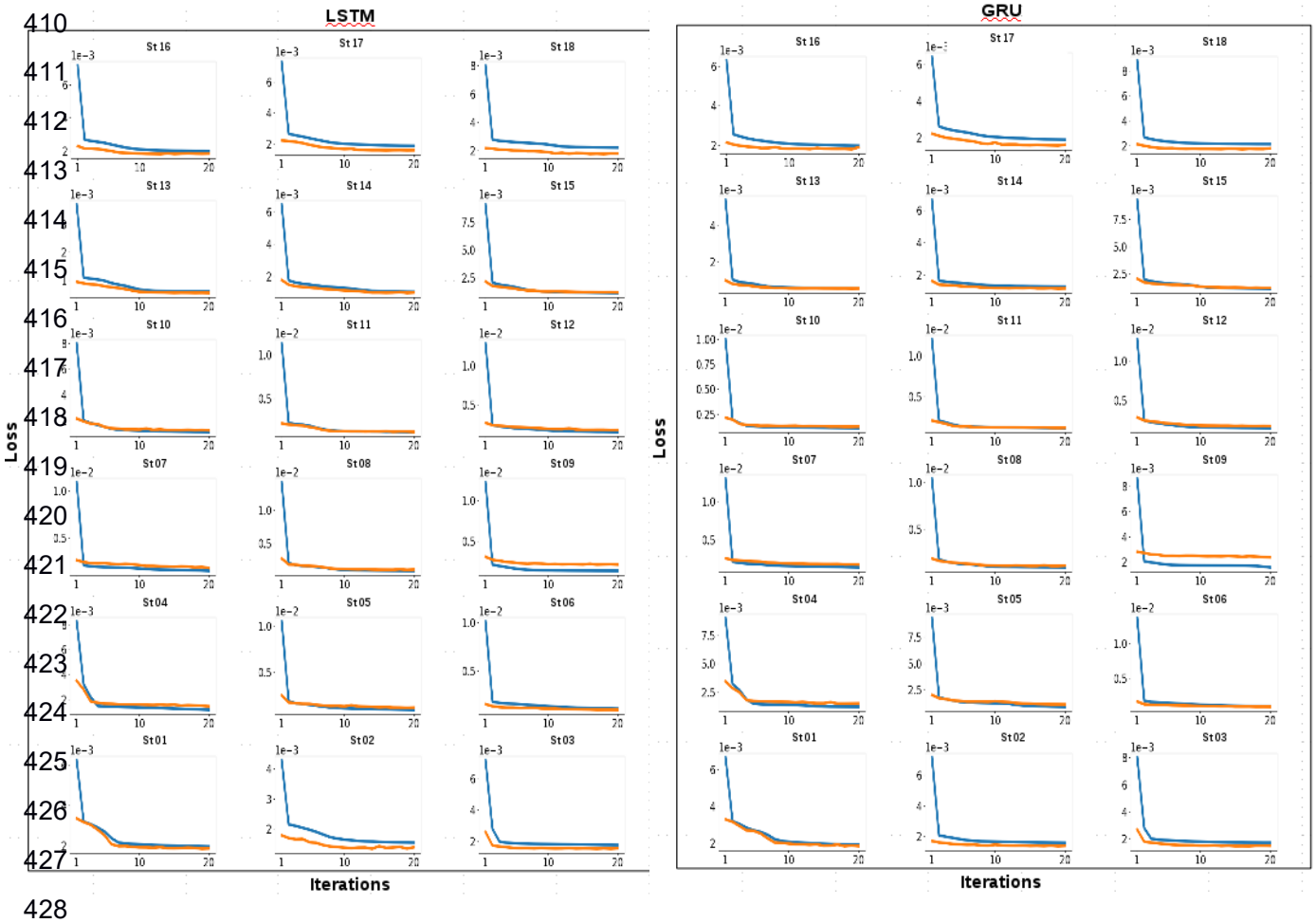


Figure 8 . Loss function during training over 20 epochs for (a)LSTM and (b) GRU models and for each virtual station. Those stations are plotted according to their position within the globe, e. g. the first upper row corresponds to North high latitude virtual stations. While plots in the same column correspond to the same longitude virtual station

When comparing Figures 8 (a) and (b) it can be observed that training results for LSTM and GRU are very similar which is an expected result (as mentioned before GRU can be seen as a simplified version of LSTM). The training process is similar but with slight differences in favor of GRU. See as an example the loss function for St 09, using LSTM the errors start around 10^{-2} while in GRU it is better with $8x10^{-3}$. Similar behaviour is observed over virtual station St 05. Only in virtual station St 10, the opposite behaviour is observed, LSTM error is $\sim 8x10^{-3}$ and GRU is $\sim 10^{-2}$. Nevertheless, none of these two methods can be considered better than the other in terms of the training process. As mentioned before this is an expected result.

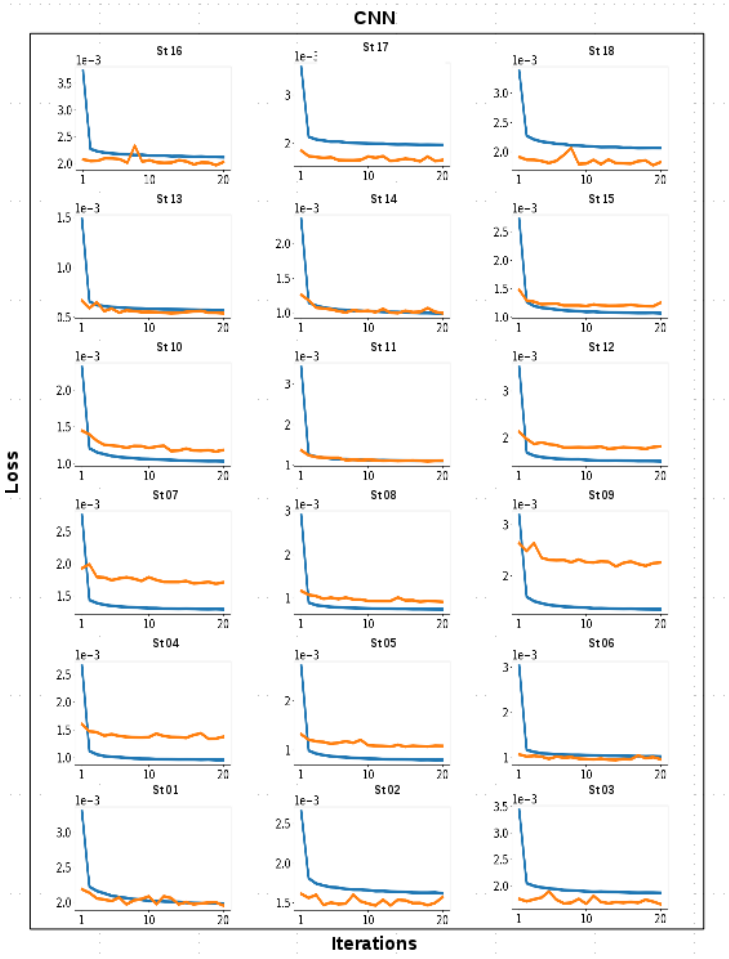


Figure 9. Loss function during training over 20 epochs for CNN models and for each of the virtual stations.

On the other hand, the training in CNN (Figure 9) shows, at every virtual station, a more oscillating behaviour of the validation curve around the training curve. This is considered a really good training process. Another important characteristic observed for the CNN training at St 04, St 07 and St 09 (in particular), the validation curve is oscillation over the training curve which is considered as an underfitting of the model. There are several strategies that, potentially, can be followed to improve the training at these stations. However, when the errors are compared against LSTM or GRU, it can be seen that CNN has better performance (less error in the worst case) in each of the stations (even at St 04, St 07 and St 09). Thus, we considered sufficient accuracy for our models.

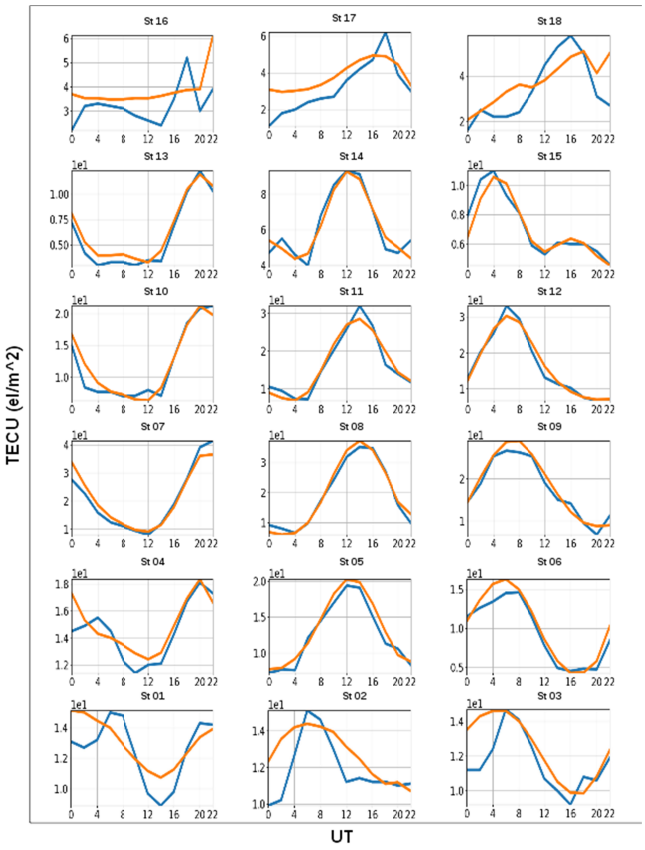
Validation using the test set

When analyzing the performance of each method using the test set (data not used during the training), we carried out two analyses: (a) the performance of the next day (18th November 2016), which means the analysis of the 24 hours after the last day in the training set. It should be mentioned that this corresponds to a geomagnetically quiet day (maximum Kp = 2), and (b) the performance of the next 43 days corresponding to the complete test set.

The main difference between (a) and (b) is that in the case of (b) we can analyse how robust the method is without re-training the model while in (a) we may expect a better performance.

For the purpose of giving a qualitative assessment of the performance, Figure 10a shows the results for the LSTM models at each grid point for the TEC forecasted (orange lines) for the first 24 hours in terms of absolute TEC comparison against the actual TEC values (blue lines). Overall, the predictions follow the tendencies and the actual values almost in concordance.

GRU method has similar results as LSTM as expected (Figure 10b). Nevertheless, some differences can be observed. For example, St13 has better accuracy using GRU than LSTM, LSTM forecasting slightly overestimated TEC between 0 and 8 UT. In contrast for St 14, LSTM has better performance than GRU.


St 16
St 17
St 18
St 13
St 14
St 15
St 10
St 11
St 12
St 07
St 08
St 09
St 04
St 05
St 06
St 01
St 02
St 03
TECU (el/m^2)
UT

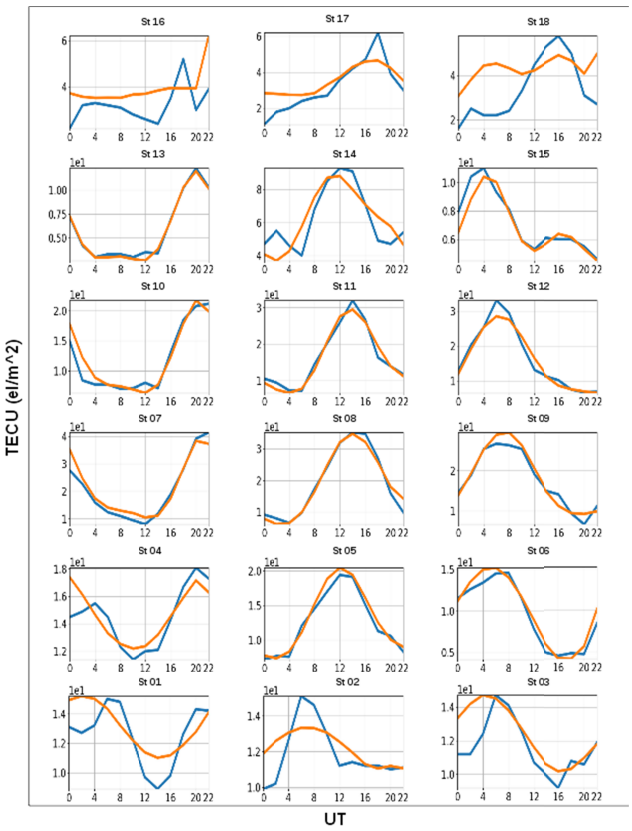

St 16
St 17
St 18
St 13
St 14
St 15
St 10
St 11
St 12
St 07
St 08
St 09
St 04
St 05
St 06
St 01
St 02
St 03
TECU (el/m^2)
UT

Figure 10. LSTM (a) and GRU (b) 24 hours forecast (orange line) at each virtual station compared with the actual value (blue line).

On the other hand, Figure 11 and Figure 12 shows that the CNN models outperform both LSTM and GRU. In the CNN models, even at high latitudes, the predictions are quite accurate showing an overall better performance than LSTM and GRU.

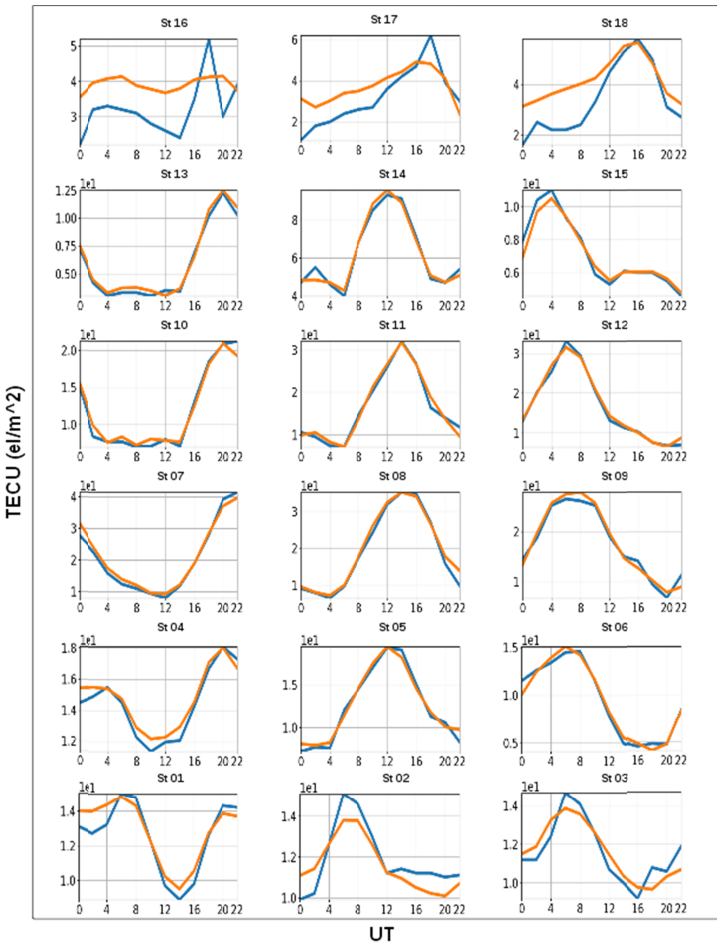


Figure 11. CNN 24 hs forecast (orange line) at each virtual station compared with the actual measured value (blue line).

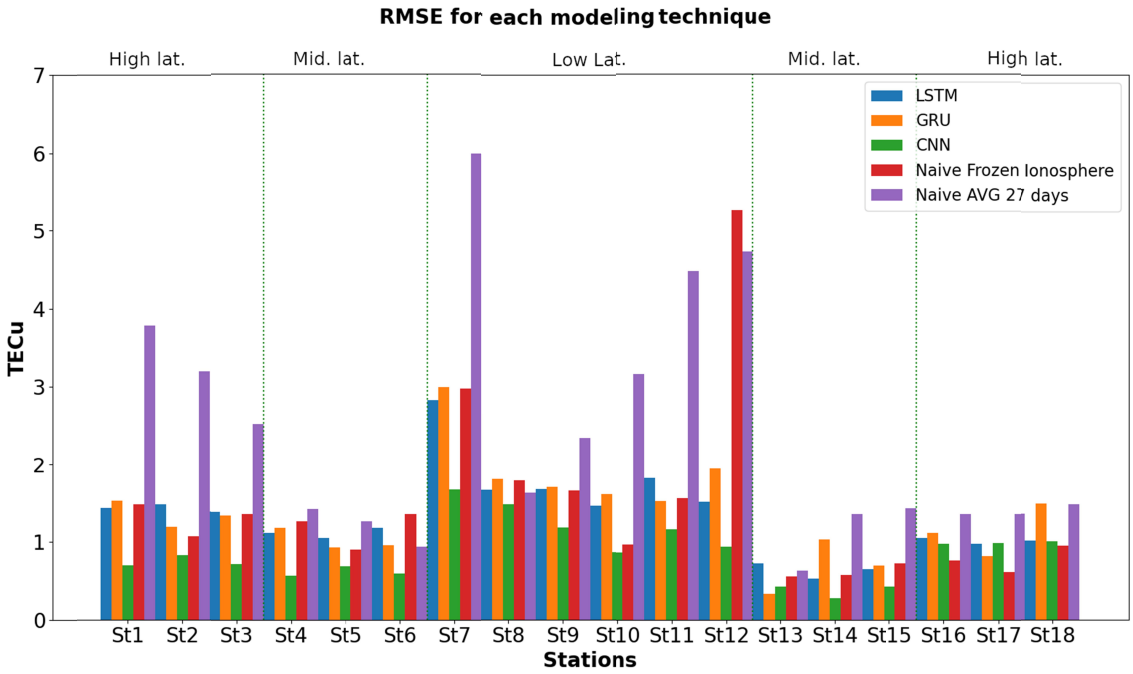


Figure 12. RMSE for the first 24 hours of forecasting. Vertical dash lines divide each region according to latitude.

Figure 12 shows the comparison of RMSE for each considered method: LSTM, GRU, CNN, naive frozen ionosphere, and naive averaged over the last 27 days.

When observing the RMSE for the 24 hs prediction, the 3 proposed models show good accuracy in general with RMSE less than 3 TECu. CNN models' worst performance has an RMSE of less than 2 TECu, being the best model at any location, except St17 where it is slightly higher than GRU and equal to LSTM. GRU models have the least accurate performance in any virtual station except for St13. Nevertheless, at St13 any of the models have a significantly good performance (RSME < ~0.8 TECu). Moreover, GRU at St7 has the worst performance reaching ~3 TECu. In general, LSTM has a stable performance between CNN and GRU, but it has a considerably high RMSE at St7 (~2.8 TECu). When compared to the naive methods, any DL method outperforms the frozen ionosphere or the averaged 27 days methods, in most cases. An exception can be observed in St 17, but none of the methods exceeds 2 TECu of RMSE. The naives are similar to a low boundary and they behave well when there is not much variability.

Figure 12 is divided in latitudinal sectors (vertical dash lines) showing a clear difference in

the accuracy of any model depending on the region (low, mid or high latitude). The complex morphology and dynamics of the ionosphere at low latitudes both during calm ((see e.g. Fejer et al., 2011; Cesaroni et al., 2015) and disturbed periods (see e.g. Astafyeva et al., 2015; Nava et al., 2016; Olwendo et al., 2017; Molina et al., 2020; Alfonsi et al., 2021; Spogli et al., 2021) with strong local phenomena such as equatorial plasma bubbles causing ionospheric scintillation and spread-F, electric fields originating from the action of the dynamo, atmospheric gravity waves propagating from below; plus the space weather effects due to the solar wind-magnetosphere-ionosphere coupling, among others (see., e.g, Li et al., 2021 and references therein). This complexity is one of the reasons that forecasting at low latitude virtual stations (close to the crests of the EIA) is especially challenging (as expected), which is clear from Figure 12 that any of the methods have less performance at low latitude virtual stations, being GRU the poorest. Additionally, among low latitude virtual stations, St7 (20° S,120° W) has the least accuracy for any method. This is due to the fact that St7 is located in the ocean where the GIM is less accurate due to the lack of ground based GNSS stations being the main source of data for GIM.

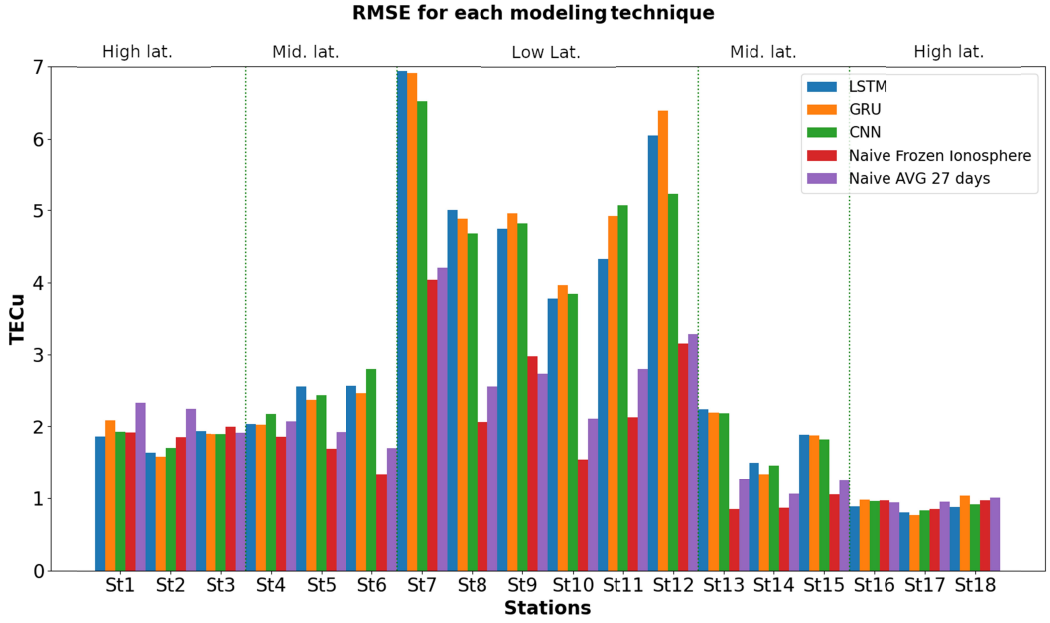


Figure 13. RMSE for the complete test set and for each method. Vertical dash lines divide each region according to latitude.

In order to analyze the behavior of the proposed models in terms of robustness when there is no retraining, online or incremental training, we calculated the RMSE for each virtual station and each method (LSTM, GRU, CNN, naive frozen ionosphere and naive 27-avg) for the complete test set (43 days). We used 24 hs forecasting steps to walk through the test set. After the first forecasted 24 hs (one step) the models are fed with the actual TEC and Kp values to forecast the next 24 hs (48 hs) without re-training, online or incremental training (the weights on the NN are the same as the output of the training). Thus it is expected that the models' performance deteriorates especially in the last days of the test set and consequently, the overall RMSE is higher than in the first 24 hs forecasting.

Figure 13 shows the RSME values calculated for the complete test set for all the virtual stations and for the three proposed (LSTM, GRU and CNN) and the naive methods. It can be seen that the RMSE for low-latitude virtual stations is considerably larger than in the other cases (mid or high-latitude regions). In general, the errors follow a similar pattern to those in Figure 12. Hence, the behavior of each model is consistent when compared with the observed 24 hs test set where CNN models have the best performance. Moreover, St07 exhibits the worst performance, which is expected due to the above-mentioned characteristic of being an oceanic low-latitude virtual station. The only exception is St14 which shows, in this case, a similar RMSE for any of the DL methods, but the RSME is quite low (~1.5 TECu). When compared with the naive methods, errors in DL methods are worse. This is because a) the above-mentioned problem of not training again using the new data as the prediction "walks" through the test set, and b) the test set (43 days) contains few perturbed days (2 days geomagnetically disturbed with Kp=5), which means, for example, that the actual data is closed to the reference curve (e.g. 27-avg days), in general terms the lower the activity the higher the chance of naive methods to have a better performance.

Case studies

It is well known that the ionosphere exhibits regular variability and morphology at different spatial and time scales, as well as non-regular variability (depending on different forcing) (Zolesi & Cander, 2014; Mendillo, 2006). In particular, we are interested in space weather effects on the ionosphere, thus we analyzed different cases of study under geomagnetic storm conditions. It is important to stress that each of the cases of study corresponds to time periods not considered in the training of the models. Moreover, we analyzed cases from 2017.

We have analyzed four geomagnetic storms (see Table 2 to better highlight the behaviour of the TEC forecasting under different levels of geomagnetic disturbance (intensity from G1 to G4, according to the NOAA scale). Table 2 summarizes the dates of the selected events (column 1), the maximum Kp (column 2), the specification of the class of the storm and the solar event generating the disturbance (CH = Coronal Hole; CME = Coronal Mass Ejection; column 3). Column 4 contains the mean value (for each DL method) of the global ΔTEC and column 5 is the standard deviation of global ΔTEC.

The global ΔTEC for any of the DL models is calculated as follows,

$$Global\Delta TEC = \frac{1}{st}\sum^{st} \Delta TEC$$

Where st is the number of the virtual station stations (grid points) and ΔTEC is the TEC difference between the forecasted TEC and the corresponding GIM TEC value.

We consider as "storm events" the days between the Sudden Storm Commencement (SSC) or sudden variation of Dst due to the arrival of the disturbance at the magnetopause until the recovery to the pre-disturbance conditions.

The table contains the same cases of study selected by Cesaroni et al (2020) but it is worth mentioning that the metrics reported in that paper refer to the comparison of the forecasted global maps obtained by applying both the "single point forecasting" and extending the forecasting for any point in the globe as described in the introduction. Even if a direct comparison with the performance of the NARX used in Cesaroni et al (2020) is not possible,any of the new proposed DL techniques seems to show better performance than the multi-layer perceptron implemented in the mentioned paper.

			μ (TECu)			σ (TECu)		
Date	Max Kp	Scale/origin	LSTM	GRU	CNN	LSTM	GRU	CNN

27-31 May 2017	7	G3 - strong / CME	-0.25	-0.14	-0.18	0.96	0.96	1.01
07-11 Sep 2017	8+	G4 - severe / CME	0.03	0.15	0.06	1.33	1.47	1.19
26Sep - 03 Oct 2017	7-	G3 - strong / CH	0.28	0.4	0.38	1.01	1.02	1.09
07 - 12 Nov 2017	6+	G1 /CH	0.03	0.15	0.06	1.33	1.5	1.19

Table 2. Case studies considering different geomagnetic storm scales, and the mean and standard deviation for LSTM, GRU and CNN models.

The first important point to address is the fact that, for the selected cases of study, the DL models have been tested for a dataset (2017) very far from the training set, with overall good results showing that they are robust enough even when using far information to train (2005 to 2016). Another note is that we are dealing with averaged overall errors (global ΔTEC) and because of that we expect a significantly smoothed behavior consequently there are no punctual details of the forecasting performance when observing at the single grid point. Nevertheless, this allows having a perspective of the performance of the proposed forecasting models in a more general scope.

When observing the global ΔTEC, for any of the DL models (LSTM, GRU, CNN), the μ varies between 0.03 and 0.38 (in absolute value), while σ varies between 0.96 and 1.5 (in the severe geomagnetic storm case). Thus, even in the worst-case scenario any of the proposed methods have an overall good performance compared with the results in Cesaroni et al., 2020. As expected, the worst results originate from the severe (G4) geomagnetic storm in September 2017. When comparing the different DL techniques, GRU models exhibit different behaviour depending on the storm (see the variation of σ, for example) while CNN appears to be consistent (σ between 1.01 and 1.19). The LSTM model is in the middle with more dependency on the storm case (σ between 0.96 and 1.33). A priori, from an overall point of view, CNN methods are more consistent and have better performance.

Figure 14 shows details for the 07-11 September 2017 severe geomagnetic storm. The left upper panel shows the Global ΔTEC (TECu) for each of the 3 methods. Dash horizontal lines mark 2 thresholds of 2 TECu and 4 TECu. In concordance, the left bottom panel shows SymH index (http://wdc.kugi.kyoto-u.ac.jp/wdc/Sec3.html) as a proxy for the geomagnetic activity. As expected, the major errors (ΔTEC > ± 2 TECu) can be found during the development of the storm from the sudden storm commencement (SSC) during late hours on the 8th of September, to the main phase (symH ~ -125 nT) on the 9th of September. After that, it can also be observed that during the brief recovery phase followed by a new perturbation in the geomagnetic field that leads to a new peak (SymH ~ -100 nT) and later on a long recovery phase during the next days, ΔTEC returns to an acceptable value between -2 to 2 TECu (for any of the methods). When observing the CNN model, it can be seen that the global ΔTEC follows the geomagnetic perturbation (slightly better than the others) but the major difference can be observed during the main phase of the geomagnetic storm when the CNN model is considerably better than the other two. This suggests that the CNN model is able to learn better from sudden changes within the data. A similar analysis with similar results was done for the other storms in table 2 (not shown here).

The right upper and lower panels show basic statistics also reported in table 2.

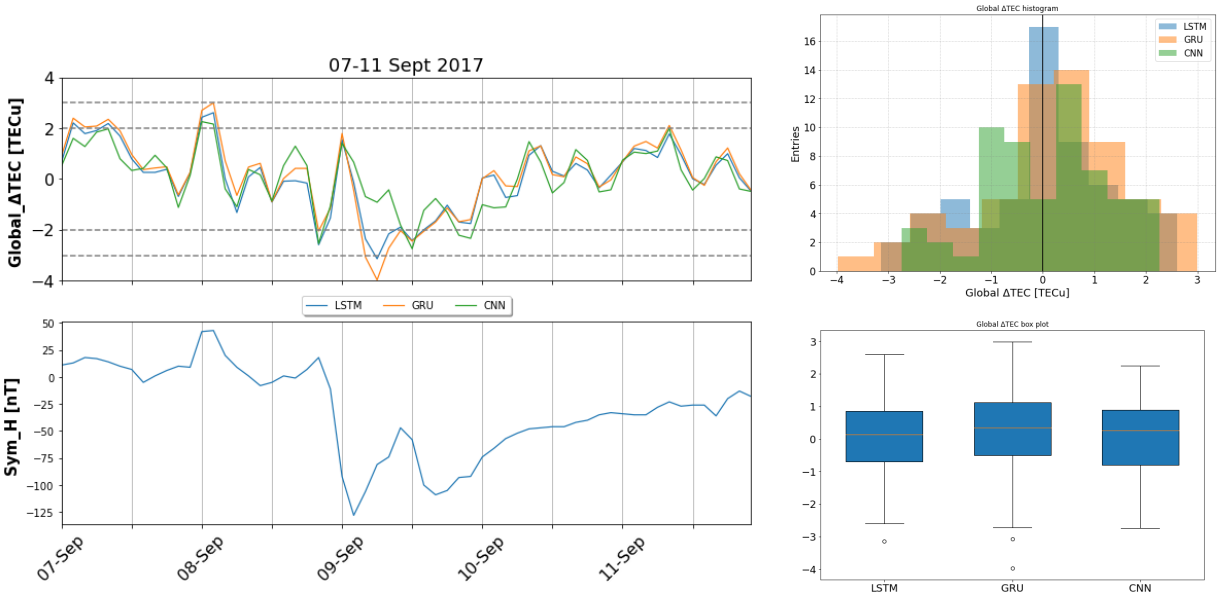


Figure 14. 07- 11 September case of study: left upper panel corresponds to Global ΔTEC for each DL technique, the left lower panel plots the SymH index, the right upper panel presents

677 the histogram for Global ΔTEC for each technique, and the right bottom pane shows a
678 boxplot also for each of the modelling methods.
679
680 ***Pre-operative prototype: incremental training***
681
682 Since the main reason for the forecasting method in space weather is to be implemented for
683 real-time operations, we tested each of the methods proposed in the paper using incremental
684 training (Ade et al., 2013; Losing et al., 2018).
685 The main objective of doing that is to assess the performance of the models when they are
686 constantly updated with new information coming from freshly available data. This is
687 especially important in space weather forecasting because one of its characteristics is being
688 an intrinsically unbalanced problem where fewer extreme events occur for a solar cycle.
689 Thus, there are few samples to learn from and often ML methods are not able to catch such
690 cases.
691 The general procedure in incremental training is the following: the models are trained once
692 (see above in subsection "Training"), then a prediction is made to forecast the first 24 hs after
693 the last training set sequence. The actual data is compared with this prediction and RSME is
694 calculated to evaluate the performance of the model. For the next step (range between 24 yo
695 48 hours after the training set), we use the previous 24 hours (new data) to re-train by
696 adjusting the weights of the already trained model. Thus, the data, as it arrives, is
697 continuously used to extend the model (Ditzler et.al., 2013).
698 Here, We show the implementation of incremental training for the CNN model since, as
699 stated above, it has the best performance of the three DL techniques. Moreover, we analyze in
700 detail the worst case possible (among the virtual stations) which is the forecasting for virtual
701 station 7 (oceanic low latitude station).
702
703
704
705
706
707
708
709
710

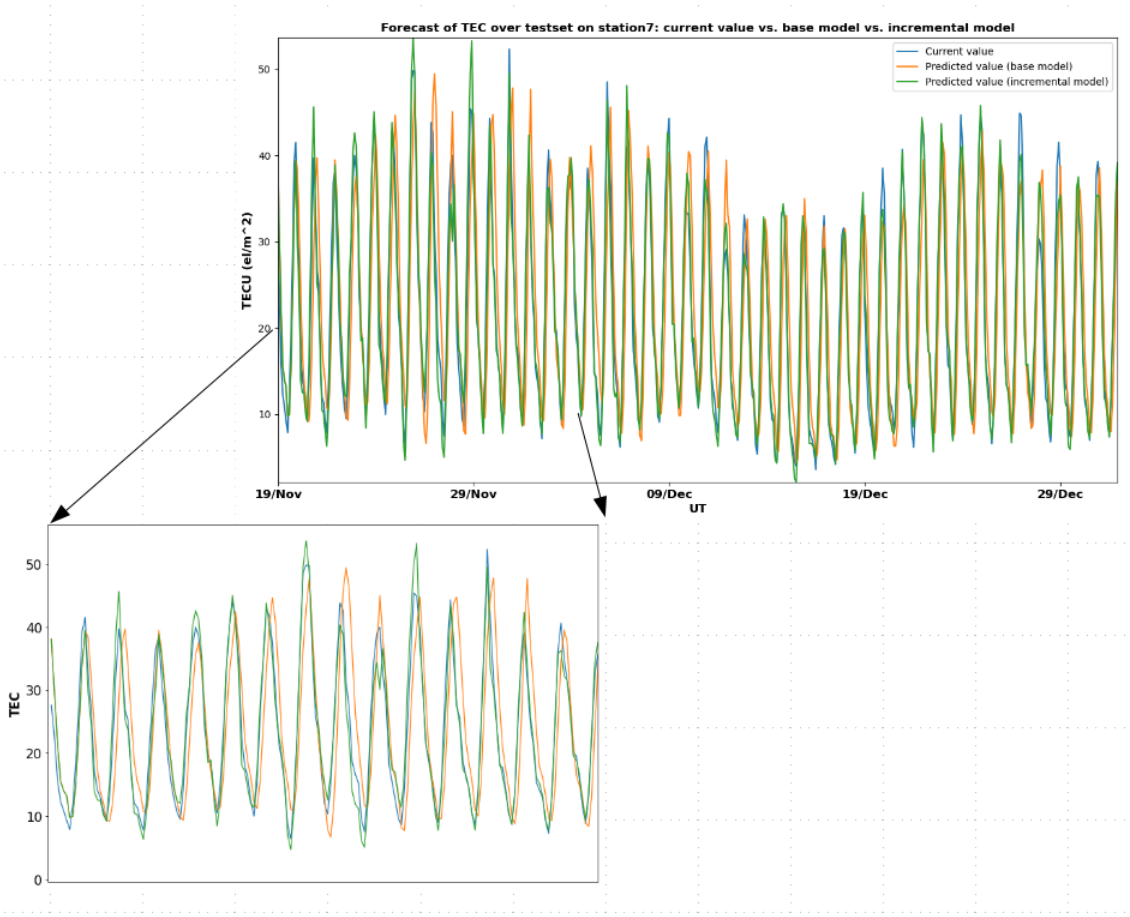


Figure 15. Comparison of actual TEC values (blue), CNN base model (orange) and CNN with incremental training (green) for virtual station 7. The upper plot shows the results for the complete test set (43 days) and the lower plot is a zoom for the first 14 days.

Figure 15 shows for the first forecasted day the same behaviour for the CNN base model (without re-training) as CNN with incremental training. As expected the results are the same because new data is not used yet. For the next step (48 hs), the previous 24 hs TEC and Kp are used to update the model. This means that the weights are adjusted by using the new 24 hs data and thus the result is significantly better. In the zoomed panel, it can be observed that the CNN model with incremental training (green line) is following quite well the actual data (blue), even the peaks, in most of the cases. We apply incremental training to every other virtual station for the test set.

Figure 16 shows the RMSE for all the methods considered in this paper LSTM, GRU and CNN with incremental training, plus the above-mentioned naive methods. As expected, the incremental training allows the model to be updated and the error diminishes. It is worth mentioning that the problem observed in Figure 13 where the models' performance decay over time is now almost solved (within reasonable error of ~2 TECu at each virtual station). This also reinforces the hypothesis that the naive methods performed better in quiet conditions (see also Figure 13). In Figure 16 it can be seen that at almost all the virtual stations the naive methods performance is poor. The exception is virtual station 9 (low latitude) where the CNN error is slightly higher (~3 TECu) than the naive methods and the LSTM and GRU models perform better (slightly over 2 TECu). Many techniques or strategies can tackle this local problem, some of which are discussed in the following section.

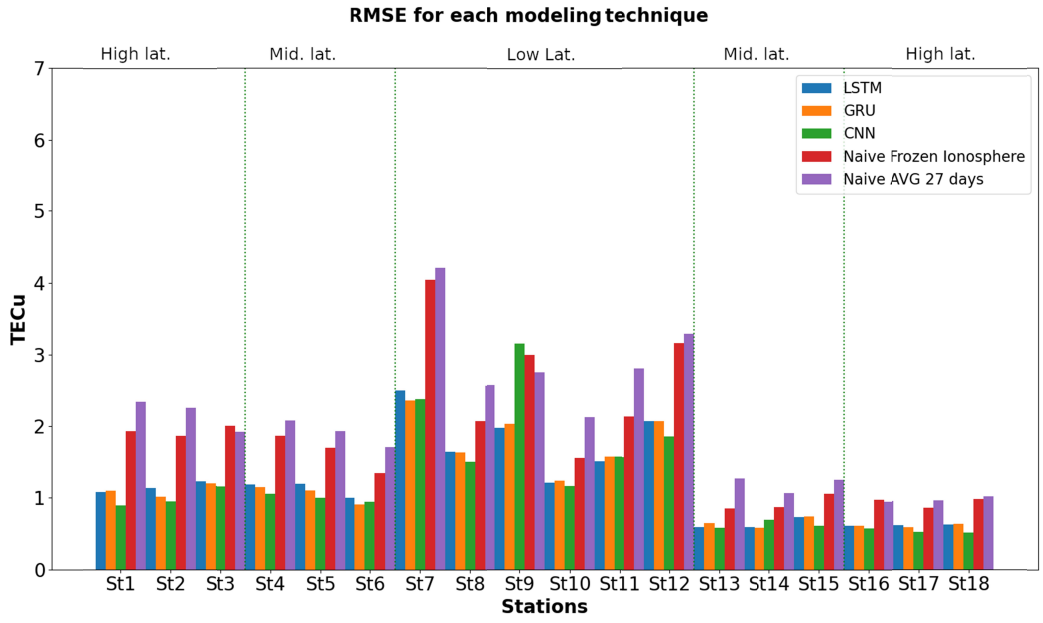


Figure 16. RMSE for each of the DL models with incremental training and RMSE for the naive models. The errors are reported for every virtual station.

Once again, we also consider the selected cases of study (see table 2) under different geomagnetic conditions and we present the results using the incremental training for CNN model.

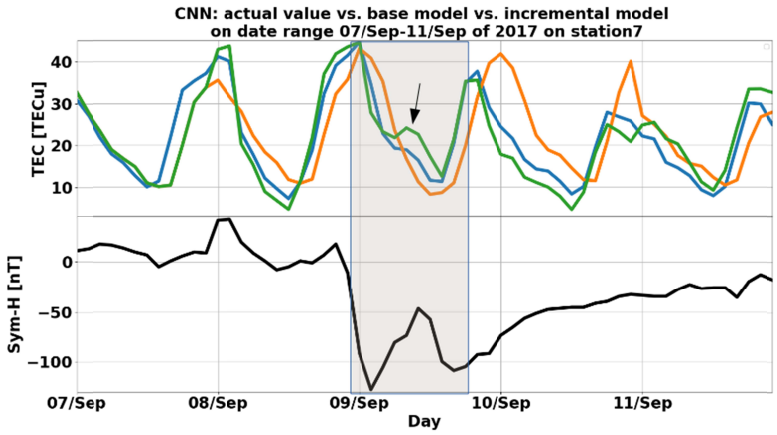


Figure 17. Upper panel: comparison between the base model (orange), the model with incremental training (green) and the actual data (blue). Lower panel Sym-H index for the period.

Figure 17. show results for the 07-Sep to 11-Sep severe (G4) geomagnetic storm case, for virtual station 7. As it can be observed, immediately after the first 24 hs the incremental training corrects the base model using the newest data. The corrected prediction (green line) is able to catch the fast changes such as the marked in shade. An important remark is that the base model is trained with very old data (2005-2016) and the forecasting is in 2017, but with the incremental training after the first forecasted day, the model is able to update fast to the new stream of data.

It is important to mention that the base model training takes approximately 20 min on a desktop PC. Once the base model is trained, the weights update using new data (incremental training) takes less than 10 sec in the same PC. This is interesting if we consider the possibility of an operative DL-based tool for forecasting.

It is worth mentioning that the proposed model update, incremental training, is one of many aspects to consider when implementing ML-based models deployed in an operative manner. Unlike in academia, ML workflow in operations should contemplate the following states: data management, model learning, model verification and model deployment. Additionally, aspects such as ethics, law, security and trustworthiness (XAI methods, end-users experience, etc,) are aspects that should also be considered (Arrieta et al., 2020; Paleyes et al., 2022).

Discussion and conclusion

In this work, we considered a data science approach to perform forecasting of the ionospheric Total Electron Content with a forecasting horizon of 24 hours able to account for the variability induced by space weather events. In order to accomplish this, we present a data pipeline to acquire the data, pre-process and prepare it for the modeling, and we implemented 3 different DL techniques. For the validation, we used RMSE as the metrics within the test set. We also implemented incremental training to update the models as the forecasting advances in time and new data can help correct the weights.

Regarding the data science approach, few discussions have been raised. First, the acquisition stage here is straightforward (since it is offline data) also the data resolution is 2 hours, thus there is enough time to acquire and prepare the data. Under these conditions, the real-time version for this stage is also trivial and easy to implement. Nevertheless, if the data resolution changes (e.g. we use directly GNSS derived TEC at certain locations or 15 minute time resolution GIM maps) it may constitute a challenge (e.g. data quality control, time scale). In general, the targeted time/space scales will steer the acquisition process in real-time scenarios.

Feature selection and engineering (e.g. adding statistic-based features) are also important and are basically led by the knowledge of the domain (and the dataset). A desired trade-off between having few dimensions to train and having enough information regarding the problem is not an easy task. From the data perspective, the aim is to have as few features as possible to lower the amount of data needed (Bellman, 1057). Still, we would like to have as much information as possible about the domain for a smart feature selection such as some scoring techniques (e.g. feature ranking method). In this regard, *a priori* knowledge is important. For example, the selection of the appropriate geomagnetic index for a given location (e.g. using Dst for equatorial virtual stations or AU for high latitude virtual stations, among others). Also, the time scales are important here, for example, a solar activity proxy such as the F10.7 is often used in ionospheric studies for longer scales than a day. Thus, in future works, we propose tackling the feature selection problem. Also, the data integration stage is desired to be smooth (e.g. easy way of adding, removing data, data structure, enhancing automatization of pre-processing operations, etc), with as few modifications as possible within the pipeline.

Regarding the operative implementation, additional technical considerations are needed

including the deployment and the monitoring of the performance. Moreover, it is desirable to update, or even change, the models or techniques without much modification in the other stages of the data pipeline, so more complex software development is needed to ensure the scalability and consistency of the system.

At the modeling stage, it is worth mentioning that any of the proposed DL technologies have good training (see section)

Besides the technical issues mentioned above, from the physical point of view, the proposed models (LSTM, GRU, CNN) have a significantly good performance on a global scale but for certain locations theory can have poor performance. This is the case of virtual station 7 which is a low-latitude station located in the ocean (worst case). In this case, the data is known to have less quality and any of the methods result in reduced forecasting capabilities. Nevertheless, with the possibility of incremental learning (see Figure 16), many issues have been overcomed. An exception is virtual station 9, where CNN could not lower the error in comparison with the other models (especially the naive models). Many strategies can be used here: improve the data quality by adding more data, augmenting the time resolution (when possible), try other hyperparameters tuning or other architectures. Another approach is ensemble learning to seek better predictive performance by combining the prediction of multiple methods. Also, other machine learning techniques can be used. Many of the mentioned techniques and strategies are proposed for future works.

Considering the first 24 hours forecasting the best method is CNN with the error ranging between 0.5 and 2 TECu without re-training. As the model is used in further samples, the performance deteriorates until, in some cases, the naive methods perform better. The proposed solution is to update the model as it “walks” in time (with a 24 hours step). In Figure 16, we show the improvement in the forecasting capabilities of the DL models when incremental training is added. In general, the performance is quite good (not more than 2 TECu) proving to be more reliable and robust models (in particular CNN).

We also analysed different cases of geomagnetic storms, in particular using data very far from the one used in the training set (different storms in 2017). As expected best results are obtained using CNN with incremental learning.

An important characteristic of CNN models in comparison with LSTM ans GRU is that, CNN model seems to have more stability and it is able to better consider fast changes. This is because CNN uses kernels to create the so-called feature maps that usually consider the spatial relationship/dependences/information from the surroundings of a given sample (depending on the kernel size). In our proposed model, the kernel size is 2, which means that

a sliding window of 2x2 is used to construct the feature map. Then, each new element of the feature map considers the information within that window. Each window is composed by 2 samples of each feature (TEC, Kp) and for the method, the feature map is considering the next immediate sample (next 2 hours). Thus CNN is considering the short term relationship which is not considered by the 2 RNN other models (LSTM and GRU). This short scale in time is very important for fast variability often observed in the ionosphere.

It is worth mentioning, that this paper is a continuation of the work started by Cesaroni et al. (2020) and we focused here on the first of two stages of their approach to global forecasting of TEC. Thus, the new improvement in each of the single-point forecasting impacts the second stage which is extending the forecasting for any point in the globe.

Although the results shown in this work are quite good, in the future, we are planning to add more data (adding recent years), study better the feature selection and propose other learning methods. Finally, in the near future, we plan to implement our models in an operative manner.

Acknowledgments

This work is supported by the Istituto Nazionale di Geofisica e Vulacnologia, the Universidad Nacional de Tucumán (grant PIUNT E-734), and CONICET.

The SymH index used in this article was provided by the WDC for Geomagnetism, Kyoto (http://wdc.kugi.kyoto-u.ac.jp/wdc/Sec3.html).

Final Global TEC maps from IGS and developed by the Universitat Politècnica de Catalunya (UPC) are available at ftp://cddis.gsfc.nasa.gov. Kp data are available from the Space Weather Prediction Center of the National Oceanic and Atmospheric Administration (NOAA).

Open Research

Final Global TEC maps from IGS and developed by the Universitat Politècnica de Catalunya (UPC) are available at ftp://cddis.gsfc.nasa.gov. Kp data are available from the Space Weather Prediction Center of the National Oceanic and Atmospheric Administration (NOAA). The SymH data used in this paper was provided by the WDC for Geomagnetism, Kyoto (http://wdc.kugi.kyoto-u.ac.jp/wdc/Sec3.html).

Datasets used in this work is publicly available in https://doi.org/10.5281/zenodo.7817642 and also the code is available at https://github.com/Laboratorio-Computacion-

Cientifica/Global-TEC-forecasting-for-space-weather-application-based-on-deep-learning-techniques

References

Abu Alfeilat, H. A., Hassanat, A. B., Lasassmeh, O., Tarawneh, A. S., Alhasanat, M. B., Eyal Salman, H. S., & Prasath, V. S. (2019). Effects of distance measure choice on k-nearest neighbor classifier performance: a review. Big data, 7(4), 221-248. https://doi.org/10.1089/big.2018.0175

Ade, R.R. & R, Deshmukh. (2013). Methods for Incremental Learning : A Survey. International Journal of Data Mining & Knowledge Management Process. 3. 119-125. 10.5121/ijdkp.2013.3408.

Alfonsi, L., Cesaroni, C., Spogli, L., Regi, M., Paul, A., Ray, S., … Sinha, A. K. (2021). Ionospheric Disturbances Over the Indian Sector During 8 September 2017 Geomagnetic Storm: Plasma Structuring and Propagation. Space Weather, 19(3). https://doi.org/10.1029/2020SW002607

Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., … Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Astafyeva, E., Zakharenkova, I., Hozumi, K., Alken, P., Coïsson, P., Hairston, M. R., & Coley, W. R. (2018). Study of the Equatorial and Low-Latitude Electrodynamic and Ionospheric Disturbances During the 22–23 June 2015 Geomagnetic Storm Using Ground-Based and Spaceborne Techniques. Journal of Geophysical Research: Space Physics, 123(3), 2424–2440. https://doi.org/10.1002/2017JA024981

Bellman, Richard Ernest; Rand Corporation (1957). Dynamic programming. Princeton University Press. p. ix. ISBN 978-0-691-07951-6., Republished: Bellman, Richard Ernest (2003). Dynamic Programming. Courier Dover Publications. ISBN 978-0-486-42809-3.

Bilitza, D., Altadill, D., Truhlik, V., Shubin, V., Galkin, I., Reinisch, B., & Huang, X. (2017). International Reference Ionosphere 2016: From ionospheric climate to real-time

weather predictions. *Space weather*, *15*(2), 418-429. https://doi.org/10.1002/2016SW001593

Bilitza, D., Pezzopane, M., Truhlik, V., Altadill, D., Reinisch, B. W., & Pignalberi, A. (2022). The International Reference Ionosphere Model: A Review and Description of an Ionospheric Benchmark. *Reviews of Geophysics*. John Wiley and Sons Inc. https://doi.org/10.1029/2022RG000792

Camporeale, E., Wing, S., & Johnson, J. R. (2018). Machine learning techniques for space weather. Machine Learning Techniques for Space Weather (pp. 1–433). Elsevier. https://doi.org/10.1016/C2016-0-01976-9

Camporeale, E. (2019). The challenge of machine learning in Space Weather: Nowcasting and forecasting. Space Weather, 17, 1166– 1207. https://doi.org/10.1029/2018SW002061

Cesaroni, C., Spogli, L., Alfonsi, L., De Franceschi, G., Ciraolo, L., Galera Monico, J. F., … Bougard, B. (2015). L-band scintillations and calibrated total electron content gradients over Brazil during the last solar maximum. Journal of Space Weather and Space Climate, 5, A36. https://doi.org/10.1051/swsc/2015038

Cesaroni, C., Spogli, L., Aragon-Angel, A., Fiocca, M., Dear, V., De Franceschi, G., & Romano, V. (2020). Neural network based model for global Total Electron Content forecasting. Journal of Space Weather and Space Climate, 10, 11. https://doi.org/10.1051/swsc/2020013

Chen, Z., Jin, M., Deng, Y., Wang, J.S., Huang, H., Deng, X. and Huang, C.M., 2019. Improvement of a deep learning algorithm for total electron content maps: Image completion. Journal of Geophysical Research: Space Physics, 124(1), pp.790-800. https://doi.org/10.1029/2018JA026167

Chen, Z., Liao, W., Li, H., Wang, J., Deng, X. and Hong, S., 2022. Prediction of global ionospheric TEC based on deep learning. Space Weather, 20(4), p.e2021SW002854. https://doi.org/10.1029/2021SW002854

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. In EMNLP 2014 - 2014 Conference on Empirical Methods

in Natural Language Processing, Proceedings of the Conference (pp. 1724–1734). Association for Computational Linguistics (ACL). https://doi.org/10.3115/v1/d14-1179

Chung, J., Gulcehre, C., Cho, K., Bengio, Y. (2014). Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. Presented in NIPS 2014 Deep Learning and Representation Learning Workshop. https://doi.org/10.48550/arXiv.1412.3555

Ditzler, Gregory & Polikar, Robi. (2013). Incremental Learning of Concept Drift from Streaming Imbalanced Data. Knowledge and Data Engineering, IEEE Transactions on. 25. 2283-2301. 10.1109/TKDE.2012.136.

Fejer, B. G. (2011). Low latitude ionospheric electrodynamics. Space Science Reviews, 158(1), 145–166. https://doi.org/10.1007/s11214-010-9690-7

Forbes, J. M., Palo, S. E., & Zhang, X. (2000). Variability of the ionosphere. Journal of Atmospheric and Solar-Terrestrial Physics, 62(8), 685-693. https://doi.org/10.1016/S1364-6826(00)00029-8

Galkin, I. A., Reinisch, B. W., Huang, X., & Bilitza, D. (2012). Assimilation of GIRO data into a real-time IRI. Radio Science, 47(2). https://doi.org/10.1029/2011RS004952

Galkin, I. A., Reinisch, B. W., Vesnin, A. M., Bilitza, D., Fridman, S., Habarulema, J. B., & Veliz, O. (2020). Assimilation of sparse continuous near-earth weather measurements by NECTAR model morphing. Space Weather, 18 (11). https://doi.org/10.1029/2020SW002463

Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. In Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics. Journal of Machine Learning Research, 9, (pp. 249–256). https://proceedings.mlr.press/v9/glorot10a.html.

Han, Y., Wang, L., Fu, W., Zhou, H., Li, T. and Chen, R., 2021. Machine learning-based short-term GPS TEC forecasting during high solar activity and magnetic storm periods. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 15, pp.115-126. https://doi.org/10.1109/JSTARS.2021.3132049

979 Hernández-Pajares, M., Juan, J.M., Sanz, J. et al. (2009) The IGS VTEC maps: a reliable
980 source of ionospheric information since 1998. J Geod 83, 263–275.
981 https://doi.org/10.1007/s00190-008-0266-1

982 Hochreiter, S., & Schmidhuber, J. (1997). LSTM can Solve Hard Long Time Lag
983 Problems. Neural Computation, 9 (8), (pp. 1735–1780).

984 Iluore, K., & Lu, J. (2022). Long Short-Term Memory and Gated Recurrent Neural
985 Networks to Predict the Ionospheric Vertical total electron Content. Advances in Space
986 Research, 70(9), (pp. 652-665). https://doi.org/10.1016/j.asr.2022.04.066

987 Kauristie, K., Andries, J., Beck, P., Berdermann, J., Berghmans, D., Cesaroni, C., ... &
988 Österberg, K. (2021). Space Weather Services for Civil Aviation—Challenges and
989 Solutions. Remote Sensing, 13(18), 3685. https://doi.org/10.3390/rs13183685

990 Kiranyaz, S., Avci, O., Abdeljaber, O., Ince, T., Gabbouj, M., & Inman, D. J. (2021). 1D
991 convolutional neural networks and applications: A survey. Mechanical Systems and
992 Signal Processing, 151. https://doi.org/10.1016/j.ymssp.2020.107398

993 Li, G., Ning, B., Otsuka, Y., Abdu, M. A., Abadi, P., Liu, Z., ... & Wan, W. (2021).
994 Challenges to equatorial plasma bubble and ionospheric scintillation short-term
995 forecasting and future aspects in east and southeast Asia. Surveys in Geophysics, 42(1),
996 201-238. https://doi.org/10.1007/s10712-020-09613-5

997 Liemohn, M. W., McCollough, J. P., Jordanova, V. K., Ngwira, C. M., Morley, S. K.,
998 Cid, C., Vasile, R. (2018). Model evaluation guidelines for geomagnetic index
999 predictions. Space Weather 16(12): 2079–2102.https://doi.org/10.1029/2018SW002067

1000 Lilensten, J., Dumbović, M., Spogli, L., Belehaki, A., Van der Linden, R., Poedts, S., ...
1001 & Zuccarello, F. (2021). Quo vadis, European Space Weather community?. Journal of
1002 space weather and space climate, 11, 26. https://doi.org/10.1051/swsc/2021009

1003 Liu, L., Zou, S., Yao, Y. and Wang, Z., 2020. Forecasting global ionospheric TEC using
1004 deep learning approach. *Space Weather*, *18*(11), p.e2020SW002501.
1005 https://doi.org/10.1029/2020SW002501

1006 Losing, V., Hammer, B., Wersing, H., 2018. Incremental on-line learning: A review and

comparison of state of the art algorithms. Elsevier, 275, p.p 1261-1274.

Mallika, L., Ratnam, D.V., Raman, S. and Sivavaraprasad, G., 2020. Machine learning algorithm to forecast ionospheric time delays using Global Navigation satellite system observations. Acta Astronautica, 173, pp.221-231. https://doi.org/10.1016/j.actaastro.2020.04.048

Matzka, J., Stolle, C., Yamazaki, Y., Bronkalla, O., & Morschhauser, A. (2021). The Geomagnetic Kp Index and Derived Indices of Geomagnetic Activity. Space Weather, 19(5). https://doi.org/10.1029/2020SW002641

McGranaghan, R. M., Mannucci, A. J., Wilson, B., Mattmann, C. A., & Chadwick, R. (2018). New capabilities for prediction of high-latitude ionospheric scintillation: A novel approach with machine learning. Space Weather, 16(11), 1817-1846. https://doi.org/10.1029/2018SW002018

Mendillo, M. (2006). Storms in the ionosphere: Patterns and processes for total electron content. Reviews of Geophysics, 44(4). https://doi.org/10.1029/2005RG000193

Menvielle, M., & Berthelier, A. (1991). The K-derived planetary indices: Description and availability. Reviews of Geophysics, 29(3), 415-432. https://doi.org/10.1029/91RG00994

Molina, M. G., Dasso, S., Mansilla, G., Namour, J. H., Cabrera, M. A., & Zuccheretti, E. (2020). Consequences of a Solar Wind Stream Interaction Region on the Low Latitude Ionosphere: Event of 7 October 2015. Solar Physics, 295(12). https://doi.org/10.1007/s11207-020-01728-7

Nava, B., Rodríguez-Zuluaga, J., Alazo-Cuartas, K., Kashcheyev, A., Migoya-Orué, Y., Radicella, S. M., … Fleury, R. (2016). Middle- and low-latitude ionosphere response to 2015 St. Patrick's Day geomagnetic storm. Journal of Geophysical Research: Space Physics, 121(4), 3421–3438. https://doi.org/10.1002/2015JA022299

Nikitina, L., Fiori, R. A. D., Ghoddousi-Fard, R., & Waddington, G. H. (2022). Statistical analysis of large and extreme global ionospheric total electron content. Journal of Atmospheric and Solar-Terrestrial Physics, 229, 105841. https://doi.org/10.1016/j.jastp.2022.105841

1035 Nørgård PM, Ravn O, Poulsen NK, Hansen LK. 2000. Neural Networks for Modelling
1036 and Control of Dynamic Systems: A Practitioner's Handbook. Springer-Verlag, Berlin,
1037 Heidelberg, pp. 258.

1038 Olwendo, O. J., Cesaroni, C., Yamazaki, Y., & Cilliers, P. (2017). Equatorial ionospheric
1039 disturbances over the East African sector during the 2015 St. Patrick's day storm.
1040 Advances in Space Research, 60(8), 1817–1826. https://doi.org/10.1016/j.asr.2017.06.037

1041 Orús, R., Hernández-Pajares, M., Juan, J. M., Sanz, J., & García-Fernández, M. (2003).
1042 Validation of the GPS TEC Maps with TOPEX data. Advances in Space Research, 31(3),
1043 621–627. https://doi.org/10.1016/S0273-1177(03)00026-7.

1044 Paleyes, A., Urma, R., & Lawrence, N. (2022). Challenges in Deploying Machine
1045 Learning: a Survey of Case Studies. ACM Comput. Surv. 55(6), (pp. 1-29).
1046 https://doi.org/10.1145/3533378

1047 Perez, R. O. (2019). Using TensorFlow-based Neural Network to estimate GNSS single
1048 frequency ionospheric delay (IONONet). Advances in Space Research, 63(5), 1607-1618.
1049 https://doi.org/10.1016/j.asr.2018.11.011

1050 Schaer, S., Beutler, G., & Rothacher, M. (1998). Mapping and predicting the ionosphere.

1051 ·Spogli, L., Sabbagh, D., Regi, M., Cesaroni, C., Perrone, L., Alfonsi, L., Ippolito, A.
1052 (2021). Ionospheric Response Over Brazil to the August 2018 Geomagnetic Storm as
1053 Probed by CSES-01 and Swarm Satellites and by Local Ground-Based Observations.
1054 Journal of Geophysical Research: Space Physics, 126(2).
1055 https://doi.org/10.1029/2020JA028368

1056 Veettil, S. V., Cesaroni, C., Aquino, M., De Franceschi, G., Berrili, F., Rodriguez, F., ...
1057 & Aragon-Angel, A. (2019). The ionosphere prediction service prototype for GNSS
1058 users. Journal of Space Weather and Space Climate, 9, A41.
1059 https://doi.org/10.1051/swsc/2019038

1060 Zewdie, G.K., Valladares, C., Cohen, M.B., Lary, D.J., Ramani, D. and Tsidu, G.M.,
1061 2021. Data-driven forecasting of low-latitude ionospheric total electron content using the
1062 random forest and LSTM machine learning methods. *Space Weather*, *19*(6),
1063 p.e2020SW002639. https://doi.org/10.1029/2020SW00263

1064 Zolesi, B., & Cander, L. R. (2014). Ionospheric prediction and forecasting. Berlin,
1065 Heidelberg: Springer Berlin Heidelberg.