Comparison of the Forecast Accuracy of Total Electron Content for Bidirectional and Temporal Convolutional Neural Networks in European Region

Artem Kharakhashyan and Olga Maltseva *

Institute for Physics, Southern Federal University, Rostov-on-Don 344090, Russia; aharahashyan@sfedu.ru
* Correspondence: oamaltseva@sfedu.ru; Tel.: +7-9-185772533

Abstract: Machine learning can play a significant role in bringing new insights in GNSS remote sensing for ionosphere monitoring and modeling to service. In this paper, a set of multilayer architectures of neural networks is proposed and considered, including both neural networks based on LSTM and GRU, and temporal convolutional networks. The set of methods included 10 architectures: TCN, modified LSTM-/GRU-based deep networks, including bidirectional ones, and BiTCN. The comparison of TEC forecasting accuracy is performed between individual architectures, as well as their bidirectional modifications, by means of MAE, MAPE, and RMSE estimates. The F10.7, 10 Kp, Np, Vsw, and Dst indices are used as predictors. The results are presented for the reference station Juliusruh, three stations along the meridian 30°E (Murmansk, Moscow, and Nicosia), and three years of different levels of solar activity (2015, 2020, and 2022). The MAE and RMSE values depend on the station latitude, following the solar activity. The conventional LSTM and GRU networks with the proposed modifications and the TCN provide results at the same level of accuracy. The use of bidirectional neural networks significantly improves forecast accuracy for all the architectures and all stations. The best results are provided by the BiTCN architecture, with MAE values less than 0.3 TECU, RMSE less than 0.6 TECU, and MAPE less than 5%.

Keywords: ionosphere; total electron content; forecasting; BiGRU; BiLSTM; BiTCN; GNSS; temporal convolution; space weather; machine learning

1. Introduction

The study of the effects of space weather on the environment is one of the prioritized areas of geophysics [1]. One of the components of the environment is the near outer space of the Earth. Among the parts of this space is the ionosphere, which affects the operation of technological systems such as global navigation satellite systems (GNSS), satellite communications, and other space communications applications [2]. The state of the ionosphere is described by such parameters as critical frequency foF2, maximum height hmF2 of the layer F2, total electron content TEC, and others. The study and prediction of TEC attracts special attention, since its values determine the accuracy of positioning [3,4]. Currently, among the methods of forecasting the ionospheric parameters, methods using neural networks are distinguished as one of the most diverse and undergoing active development [5]. However, since each region of the globe has different properties related to the influence of space weather, it is necessary to choose a method that can provide the most accurate TEC forecast for each region. This paper attempts to select from a variety of methods the most appropriate one for the European region. For this purpose, both the conventional methods', long short-term memory LSTM [6], gated recurrent unit GRU [7], and new ones including bidirectional [8] and temporal convolutional networks TCN [9], architectures are used. From the extensive literature, we selected those papers that include quantitative estimates of forecast accuracy in different regions.

An example of using the LSTM architecture for TEC prediction is a paper which proposed a method with an advance time of 24 h [10]. The authors used CODE (Center for Orbit Determination in Europe) map data with a 1 h step for Beijing station, as well as an index solar radio flux at 10.7 cm (F10.7) of solar activity and geomagnetic activity ap as input parameters. The dataset consists of 1 January 1999 to 1 September 2016. The root mean square error (RMSE) was 3.5 TECU. It is interesting to note that the test years were 2001 and 2015, i.e., years of high solar activity. Sun et al. [11] developed and applied the bidirectional long-term memory (BiLSTM) model to TEC prediction on the same database and using F10.7 and ap indices as in [10]. The results obtained were compared with other methods. The RMSE for BiLSTM was 3.35 TECU; 3.48 TECU was obtained for single-layer LSTM and 3.67 TECU was obtained for double-layer LSTM. A feature that improved the LSTM result is the ability to utilize both the past and future data to make prediction.

Sivakrishna et al. [12] applied the BiLSTM algorithm to one-hour-ahead forecast the Indian regional TEC map. The training data included 261 days (from November 2015 to September 2016). The validation dataset included 24 days in September 2016, and the testing data were 31 days on October 20. The maps were plotted against data from 26 Global Positioning System stations and compared with Artificial Neural Network (ANN) and LSTM model results during both geomagnetic quiet (20 October 2016) and disturbed (14 October 2016) periods. For quiet conditions, the difference between the experimental and predicted values was from 1 TECU up 2.5 TECU for BiLSTM, and from 1 TECU to 3.5 TECU for disturbed conditions and were better than the results for ANN and LSTM. The LSTM model performed better than the ANN model. A comparison between the Indian regional ionospheric forecast of TEC maps with and without solar and geomagnetic indices (F10.7, Kp) as input to BiLSTM during quiet and disturbed periods showed little improvement.

The LSTM architecture has become the reference model against which the results of new approaches are compared, and most often, the results of this comparison are given when they are better than for LSTM. However, the LSTM method continues to be modified. In Chen et al. [13], several architectures, single-step self-prediction model, single-step auxiliary prediction model, multistep self-prediction model, and multistep auxiliary prediction (MSAP) model, were used for TEC prediction based on International GNSS Service (IGS) maps with an advance time of 24 h, 48 h, and 144 h. Their algorithms are presented in the appendix of the paper [13]. The database covered a set of TEC values from January 2011 to December 2019 and indices included the following: the sunspot number R, F10.7, ap, and Dst. In contrast to traditional data processing methods, the cycle consisted of 90 days. Among these, the first 30 days are considered as the training set, the middle 30 days are considered as the validation set, and the last 30 days are considered as the test set.

Comparing the interval within 24 h, the averaged MAE (RMSE) values calculated from the single-step self-prediction model within 48 h increases from 2.721 (3.806) TECU to 3.259 (4.483) TECU. For the MSAP model, these values were 2.116 (3.033) TECU for 24 h and 2.225 (3.175) TECU for 48 h. As the prediction interval increases, the differences become larger: 5.53 (7.427) TECU and 2.485 (3.511) TECU for MSAP and 144 h.

A TEC forecasting model based on deep learning was proposed by Tang et al. [14], which consists of a convolutional neural network (CNN), a long short-term memory (LSTM) neural network, and an attention mechanism. The attention mechanism is added to the pooling layer and the fully connected layer to assign weights to improve the model. The dataset included TEC from 24 GNSS stations of China Network over the 9-year period (from 2010 to 2018) to predict the value of the TEC 24 h ahead. The data from 2010 to 2017 is used as the training set and the data from 2018 is used as the test set. Bz, Kp, Dst, and F10.7 indices are used to characterize solar and geomagnetic activity. The results were compared to the accuracy of models such as NeQuick, LSTM, and CNN-LSTM. The statistical results were obtained from the data of 24 GNSS stations, 24 h a day and 365 days a year in the test set. For all stations, the MAE was 2.6 TECU for the NeQuick, 1.53 TECU

for the LSTM, 1.36 TECU for the CNN-LSTM, and 1.17 TECU for the CNN-LSTM-Attention model. The RMSE for the NeQuick was 3.59 TECU, 2.25 TECU for the LSTM, 2.07 TECU for the CNN-LSTM, and 1.87 TECU for the CNN-LSTM-Attention model. The latitudinal and longitudinal dependencies of forecast accuracy were also determined from the data of six stations with similar latitudes and longitudes. RMSE decreases with the increase in latitude near the same longitude, which is about 1 TECU in the mid-latitude region and between 1.6 and 3 TECU in six stations around 30°E. Accuracy estimates depending on the level of disturbance were performed separately for days with Kp < 3 and Kp ≥ 3. For a quiet day on 21 August 2018, the statistics showed RMSE = 4.14 TECU, MAE = 3.16 TECU, and the mean error ME = −0.12 TECU for the NeQuick model. For the LSTM model, statistical estimations have provided the following results: RMSE = 4.29 TECU, MAE = 3.0 TECU, and ME = −1.5 TECU. For the other architecture, CNN-LSTM, the following error values were obtained: RMSE = 4.14 TECU, MAE = 2.98 TECU, and ME = −1.62 TECU. For the proposed CNN-LSTM-Attention architecture, the results were better: RMSE = 3.99 TECU, MAE = 2.81 TECU, and ME = −1.46 TECU. During the perturbed day of 26 August 2018, the magnetic storm led to big changes of TEC, and the accuracy of all methods decreased. For the NeQuick model, RMSE = 4.99 TECU, MAE = 3.67 TECU, and ME = 1.71 TECU. The forecast for the LSTM method provided the following values: RMSE = 5.19 TECU, MAE = 3.76 TECU, and ME = −2.07 TECU. These values decreased to RMSE = 4.43 TECU, MAE = 3.06 TECU, and ME = 0.21 TECU for CNN-LSTM and to RMSE = 4.11 TECU, MAE = 2.81 TECU, and ME = −0.64 TECU for the CNN-LSTM-Attention architecture.

Boulch et al. [4] proposed the ConvRNN method and compared the results with the data presented in the papers [15–17] for specific Chinese stations and on a global scale, with a lead time of 2 and 48 h. In the range 22°N–39°N, an increase in the forecast accuracy with increasing latitude was obtained. On a global scale, the method in [17] provided RMSE = 3.1 TECU and the method in [4] gave 0.89 TECU for a lead time of 48 h and 0.38 TECU for 2 h ahead.

Iluore and Lu [18] analyzed the data from the equatorial station MAL2 (Kenya) and determined that GRU was more accurate than LSTM, Multilayer Perceptron (MLP), GIM_TEC, and the IRI-Plas 2017. The data covered 9 years from 1 January 2010 to 31 December 2018. The data from the year 2010 to 2016 were used for the training, while data from the year 2017 were used for validation and the data in the year 2018 were used to estimate the performance of the models. The GRU unit showed a prediction error of 2.004 TECU while the LSTM, MLP, GIM_TEC, and IRI-Plas 2017 models showed a prediction error of 2.055 TECU, 2.336 TECU, 5.913, and 16.183 TECU, respectively.

Kaselimi et al. [19] proposed a spatiotemporal deep learning architecture, CNN-GRU, combining a convolutional neural network to capture the spatial variability of TEC and a gated recurrent unit for temporal variability modeling. Real TEC measurements were used by determining the slant STEC and its conversion to TEC for six stations in different regions of the globe, with F10.7, solar sunspot number SSN, Kp, ap, and Dst as the input parameters. The TEC values obtained by the proposed method were compared not only with the results of the neural network methods, LSTM [20], BiLSTM [11], and RNN [21], but also with the GIM map CODE and IRI model values. The dataset included data from 2014 to 2018 with the test period being during the second half of 2018. Specific estimates of MAE ranged from 0.7–1.8 TECU for CNN-GRU, 0.9–1.8 TECU for Recurrent Neural Network (RNN), and 0.8–2.2 TECU for LSTM and BiLSTM.

Unfortunately, many papers do not provide Mean Absolute Percentage Error (MAPE) values, while a comparison of MAE and RMSE does not always give the full picture. As shown in Table III in Kaselimi et al. [19], different methods can give the best results for different stations. As for the dependence of the results on latitude, they were better for the middle-latitude station Graz than for the high-latitude station Tixi. This paper also compares the processing time in minutes and the number of the required trainable parameters per method. The most efficient was the RNN architecture, which requires 23 min (for 200 training epochs), while the heaviest is the BiLSTM model which requires

280 min (for the same 200 epochs). The proposed CNN-GRU architecture requires 67 min for its training.

However, other approaches are also possible. Using the advantages of neural networks combined with the Global Ionosphere Map GIM, an empirical model to predict TEC, 24 h in advance at the global scale was developed in the work of Cesaroni et al. [22]. A nonlinear autoregressive neural network with eXternal input NARX was used as a neural network. The Polytechnic University of Catalonia's (UPC) products for single point forecasting were used as GIM. To extend the forecasting at a global scale, a NeQuick2 model adapted to an effective sunspot number R12eff was used. TEC and Kp data for 11 years (2005–2015) divided into training (70%), validation (15%), and testing (15%) datasets with a time resolution of 2 h were used. Examples for four points and two quiet days (30 May 2016 and 21 August 2016) yielded an RMSE of 1.5 TECU, 0.32 TECU, 2.3 TECU, and 3.0 TECU. Accumulated statistics showed that the RMSE was 3.5 TECU for the period June 2017 to May 2018. For the disturbed period of 7–11 September 2017, the RMSE = 3.8 TECU. The RMSE for six per disturbed periods of different intensities (from 1 to -142 nT) was in the range 3.4–5.1 TECU and showed no dependence on Dst. This 24 h empirical approach was implemented on the Ionosphere Prediction Service (IPS), a prototype platform to support different classes of GNSS users.

The work of Natras et al. [23] provides an overview of most of the methods used for TEC forecasting: ANN, LSTM, LSTM-CNN, Encoder-Decoder LSTM, Extended ED-LSTME, NARX, conditional Generative Adversial Network cGAN, and others, and gives statistical characteristics of these methods. In particular, for 1 h forecasting and up to 1-day and 2-day forecasting, the RMSE for a 1 h TEC forecast in low latitudes ranges from 2 to 5 TECU for different learning algorithms and different levels of solar activity. For the mid-latitude 1 h VTEC forecast, the RMSE is about 1.5 TECU. For a 1-day TEC forecast, the RMSE was 4 TECU in high solar activity and 2 TECU in low solar activity. However, noting the large complexity of using deep learning methods as a drawback of these methods, the authors used three algorithms: Decision Tree and ensemble learning of Random Forest, Adaptive Boosting (AdaBoost), and eXtreme Gradient Boosting (XGBoost) for 1 h and 24 h VTEC forecasts. According to the authors, the advantages of these methods are that they are simple, non-parametric, fast to optimize, computationally efficient, and can be used on a limited dataset. The dataset for training and cross-validation included the period January 2015–December 2016. The test dataset included the period January–December 2017. Test results of six methods are given for three latitudes 70°N, 40°N, and 10°N (longitude not shown) and two periods: year 2017 and 7–10 September 2017. The results for the method showing the highest accuracy (Random Forest) are as follows. For 1 h ahead and latitude 70°N, the RMSE was 0.54 TECU, for 40°N the RMSE was 0.92 TECU, and for 10°N the RMSE was 1.2 TECU. For 24 h ahead and a latitude of 70°N, the RMSE was 1.06 TECU, for 40°N, the RMSE was 1.86 TECU, and for 10°N, the RMSE was 2.2 TECU. For the disturbed period 7–10 September 2017 and 1 h ahead, the RMSE were 0.73 TECU, 1.31 TECU, and 1.29 TECU for 70°N, 40°N, and 10°N, respectively. For 24 h ahead, these values were 1.77 TECU, 3.95 TECU, and 3.95 TECU.

Thus, an analysis of the literature data shows that there is a certain sequence of using neural network methods LSTM to GRU to bidirectional to TCN, which allows gradually increasing the accuracy of the TEC forecast. However, the results can be highly dependent on the combination of architectures, region, and space weather conditions.

In a previous work [24], the following results for single hidden layer LSTM and gated recurrent unit neural networks were obtained for the Juliusruh station in 2015: MAE = 1.5 TECU, RMSE = 1.9 TECU, MAPE = 17% for GRU, MAE = 1.39 TECU, RMSE = 1.85 TECU, and MAPE = 14% for LSTM.

In this paper, a set of multilayer neural network architectures, including both LSTM and GRU neural networks and temporal convolutional networks, are proposed and considered. The main aim of this paper is to complement them with bidirectional architectures

and to determine the method that provides the highest prediction accuracy in the European region.

In Section 2, experimental data is described and the behavior of the basic indices of solar and geomagnetic activity and the total electron content is illustrated for the chosen years. Additionally, Section 2 presents the neural networks-based forecast methods. The results of the proposed methods are given in Section 3. Section 4 contains a discussion of the results. Section 5 provides the conclusions.

2. Materials and Methods

In this section, information about Materials is given in Section 2.1. Information about Methods is presented in Section 2.2, where a detailed description, including diagrams and formulas, for each method is provided in the corresponding partitions.

2.1. Experimental Data and TEC Behavior

The comparison of the forecasting accuracy of TEC using recurrent and convolutional neural networks, including conventional and bidirectional implementations, was carried out for 3 selected years with significantly different geomagnetic conditions and solar activity levels. The values of global JPL GIM-TEC maps were calculated from IONEX files with a time step of 2 h (https://urs.earthdata.nasa.gov (accessed on 10 February 2023)) for Juliusruh (54.6°N, 14.6°E), Murmansk (69°N, 33°E), Moscow (55.6°N, 37.2°E), and Nicosia (35.1°N, 33.2°E) for three years 2015, 2020, and 2022. The data on the indices of solar and geomagnetic activity, Dst, F10.7, proton density Np, planetary 10 Kp index, and speed of solar wind Vsw, was taken from SPDF OMNIWeb Service (http://omniweb.gsfc.nasa.gov/form/dx1.html (accessed on 10 February 2023)).

The state of the ionosphere and the dynamics of the total electron content depend on the level of solar and geomagnetic activity. The behavior of the basic indices influencing the TEC changes is given in Figure 1 for the years chosen for the analysis: 2015 (near to a maximum of 24 solar cycle), 2020 (a minimum of 24 cycle and transition to 25 cycle), and 2022 (an ascending branch of 25 cycle).
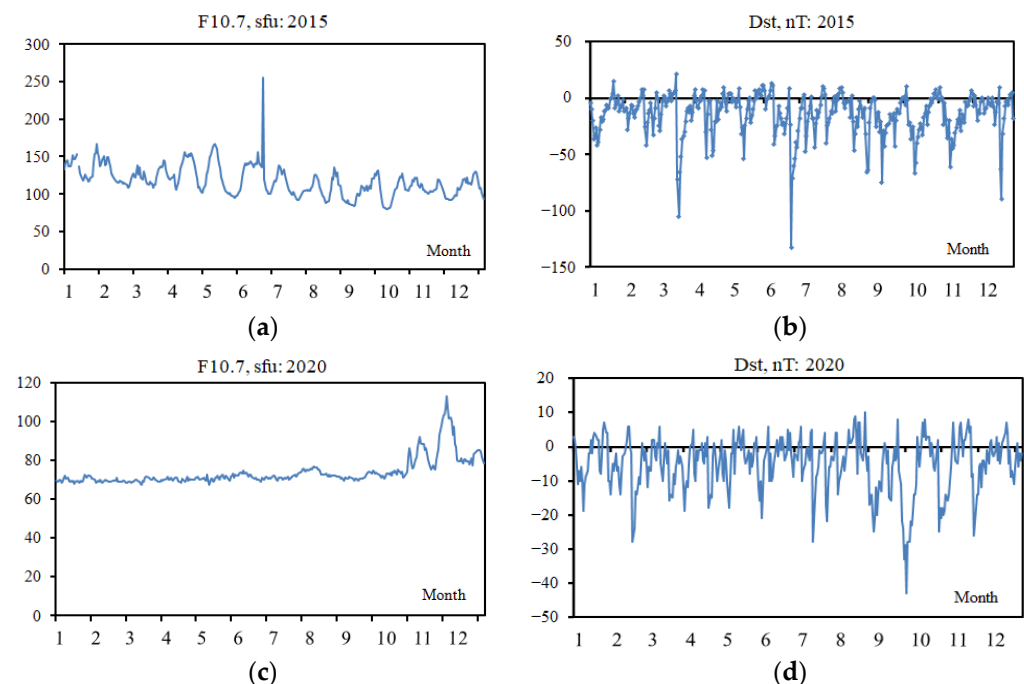
Figure 1. *Cont.*
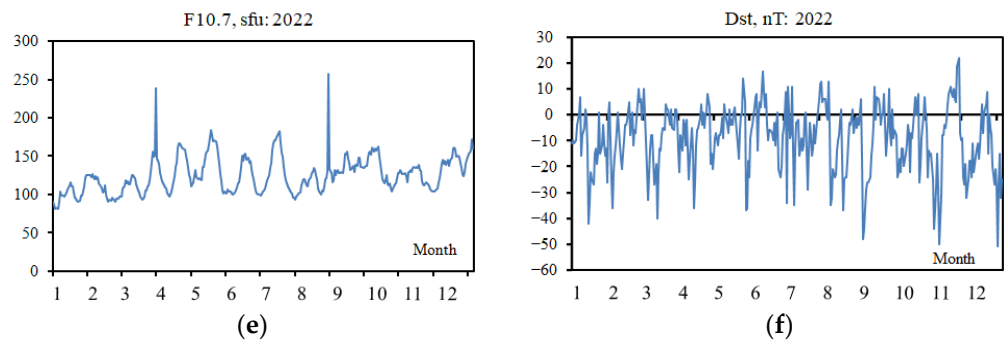
Figure 1. Average daily values of F10.7 and Dst indices for different years: (**a**,**b**) 2015; (**c**,**d**) 2020; (**e**,**f**) 2022.

There is a clear downward trend in F10.7 in 2015, a persistence in 2020 with a rather sharp increase at the end of the year, and an increase in 2022. The maximum disturbance was observed in 2015 and the minimum in 2020.

Figure 2 shows the median values, TEC(med), for the Juliusruh station (left panels) and stations along the 30°E meridian (right panels).
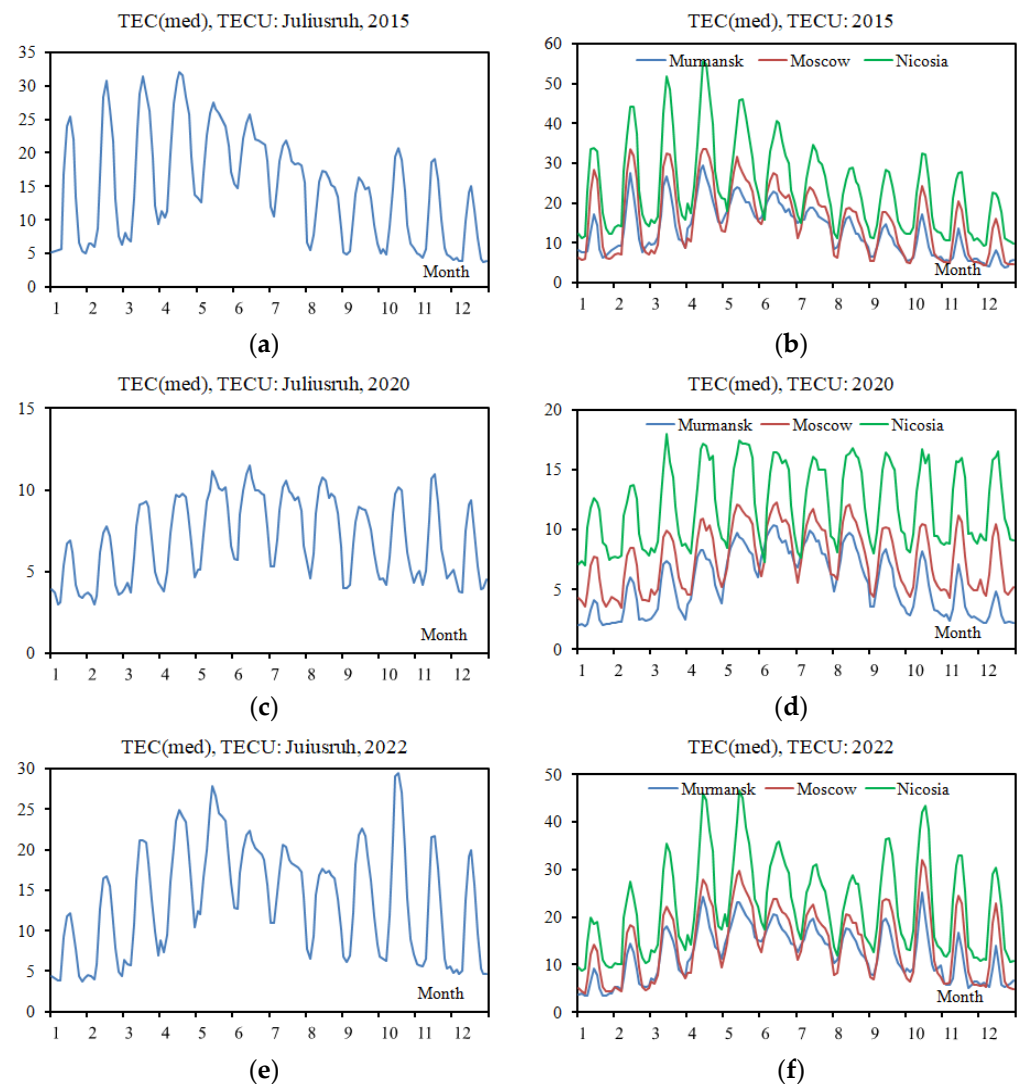
Figure 2. Variations of medians, TEC(med), for the reference station, Juliusruh, and stations along the meridian 30°N: (**a**,**b**) 2015; (**c**,**d**) 2020; (**e**,**f**) 2022.

The year 2015 is on the descending branch of the 24th cycle, which determines higher TEC values in the first half of the year compared to the second half. The year 2020 falls on the minimum of solar activity and is characterized by a more uniform seasonal course and 2–3 times lower maximum values compared to the year 2015. The year 2022 is on the rising branch of cycle 25, and the seasonal dependence differs from the first two cases both in the dynamics of change and in the maximum values of TEC. For stations along the meridian 30°E, this trend is preserved with the corresponding ratio between the maximum values, and this will be reflected in the MAE and RMSE values for different forecast architectures. The dynamics of TEC(med) for the Juliusruh station shows trends determined by the course of the F10.7 index. A comparison of the dynamics of TEC(med) along the meridian shows that for any considered level of solar activity, TEC predominantly increases with decreasing latitude. As noted above, many papers show that taking indices into account allows an increase of the accuracy of TEC prediction. Calculation of the correlation coefficients of TEC with these indices showed that in this case, the most significant were the Dst and solar wind speed Vsw indices. The correlation coefficients are shown in Figure 3.
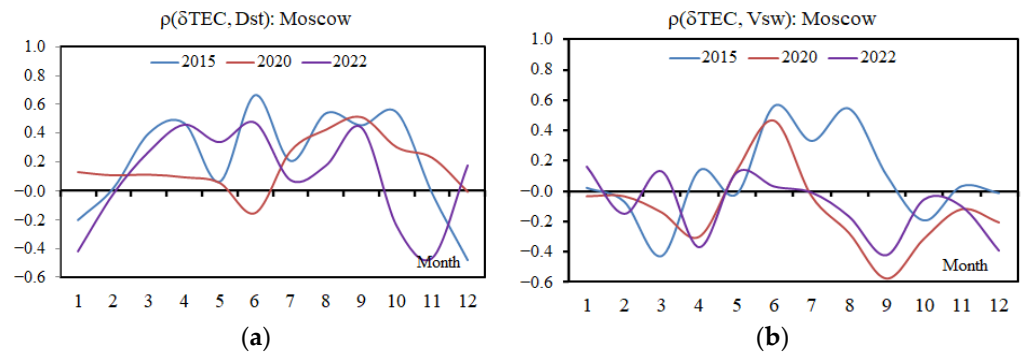
Figure 3. Behavior of TEC correlation coefficients with indices which make the main contribution to TEC change for the Juliusruh station: (**a**) Dst, (**b**) Vsw.

The most commonly used indices are F10.7 and Dst. In this work, indices F10.7, Dst, 10 Kp, Np, and Vsw were used separately and together. It can be seen that these two factors can influence in phase or in antiphase. Figures 4 and 5 show that the influence of Dst, which led to a negative response, was stronger than that of Vsw for June 2015.
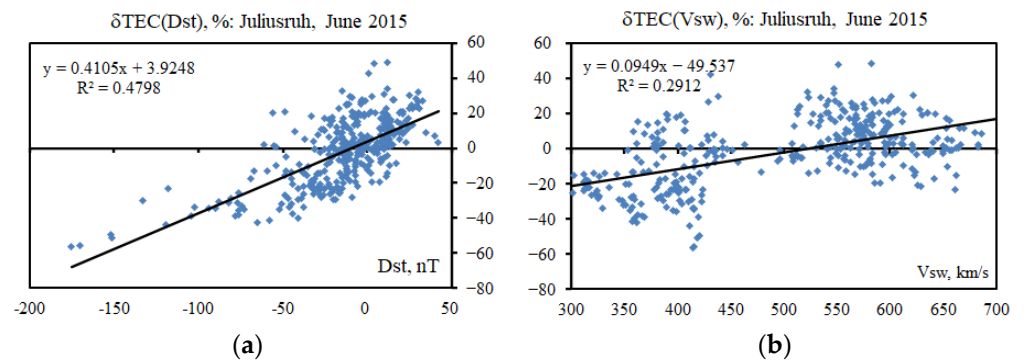
Figure 4. The scatter plot of changes in the relative deviations of TEC from the median (δTEC) with indices: (**a**) Dst, (**b**) Vsw. Regression equations and squares of correlation coefficients are shown inside the figures.
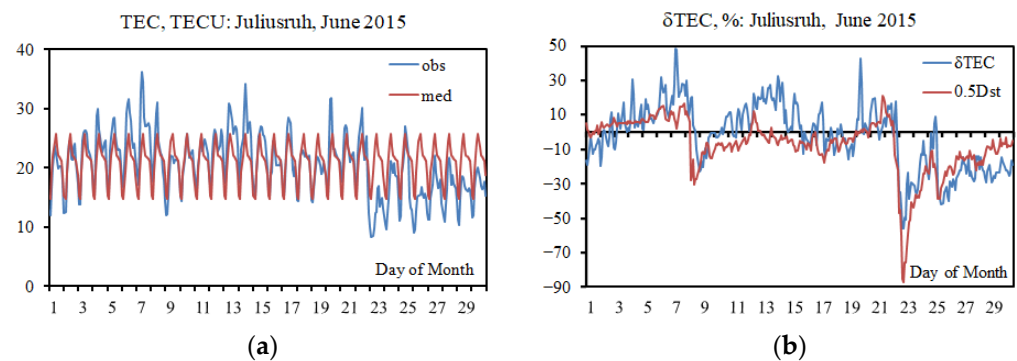
Figure 5. Illustration of negative disturbances during periods of large negative values of the Dst index: (**a**) influence on behavior of TEC, (**b**) response δTEC.

2.2. Neural Networks Architectures for TEC Forecasting

In this paper, a set of architectures of multilayer neural networks based on GRU, LSTM, and temporal convolution are proposed and considered. The proposed non-bidirectional architectures are first considered in comparison with each other and are then subjected to further modification using bidirectional processing. After that, a comparison is made of both bidirectional architectures with each other and with the original options. This section describes these modifications in detail.

2.2.1. Data Preprocessing

For each year and each station, the data preparation procedure was as follows. Datasets for each year considered consisted of samples with a time step of 2 h for both TEC values and solar and geomagnetic activity indices. For indices that are provided only on a daily basis, daily values were used in 2 h increments throughout the respective day.

Each dataset was split into training, validation, and test subsets. The first 40% of the samples from yearly datasets were used as training data, the next 10% were used for validation, and the last 50% of the samples corresponding to the second half of the year were used for testing. TEC samples are organized into subsequences formed by the applied sliding window with a width equal to 12 samples. All the architectures in this paper are considered for two sets of indices of solar and geomagnetic activity. One set included only Np, 10 Kp, Dst, and F10.7; the second set included Np, 10 Kp, Dst, F10.7, and Vsw. One of the tasks of this work was to determine the possibility of improving the accuracy of TEC values' prediction by including Vsw in the training set due to a high correlation with TEC in this case (Figure 3). The indices are organized into individual sequence features, but no sliding window is applied. Each index array is shifted so the TEC value at time step t was predicted using the index at time step $t-1$. Thus, the TEC value at the time step t is predicted using 12 previous TEC values and a single previous value of the respective included indices. Data arrays are truncated appropriately to maintain the same array lengths for all features.

2.2.2. Recurrent Neural Networks

Among the variety of different types of recurrent neural networks, neural networks based on the employment of LSTM and GRU cells are most widely used in solving problems of forecasting and modeling time series. Long short-term memory neural networks were developed to solve the long-term dependency problem and make it possible for neural networks to effectively learn information over long periods of time. This architecture was originally introduced by Hochreiter and Schmidhuber [6] and became one of the most widespread architectures that has found an application in almost every area. LSTM networks store the information about learned temporal dependencies between each element of the input sequences during its operation in the internal memory state cell. The

main advantage of LSTM is the ability to handle longer time sequences compared to a conventional recurrent neural network, which faces the vanishing gradient problem.

The set of equations defining the LSTM cell is given as follows:

$$i_t = \sigma(w_i \cdot [h_{t-1}, x_t] + b_i), \tag{1}$$

$$f_t = \sigma\left(w_f \cdot [h_{t-1}, x_t] + b_f\right), \tag{2}$$

$$o_t = \sigma(w_o \cdot [h_{t-1}, x_t] + b_o), \tag{3}$$

$$\widetilde{C}_t = tanh(w_C \cdot [h_{t-1}, x_t] + b_C), \tag{4}$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \widetilde{C}_t, \tag{5}$$

$$h_t = o_t \odot tanh(C_t), \tag{6}$$

where i_t represents the input gate, f_t is the forget gate, o_t is the output gate, σ and $tanh$ are the sigmoid and hyperbolic tangent activation functions, respectively, C_t is the internal memory state, $\widetilde{C}_t$ represents the cell state candidate, C_{t-1} is the state at a timestamp $t-1$, h_{t-1} is the output of the previous LSTM cell at a timestamp $t-1$, h_t is the output, x_t is the input, w_C, w_i, w_f, and w_o are weights for the respective gates, and b_C, b_i, b_f, and b_o are biases. The $\odot$ symbol denotes the Hadamard product.

An alternative solution to the problem of learning long-term dependencies is gated recurrent unit-based neural networks, originally introduced by Cho et al. [7]. The gated recurrent unit is a simplified modification of the LSTM cell that does not store the internal state of the cell but instead uses an update gate and a reset gate to control the progress of the data transformation.

Data processing using GRU cells is determined by the following equations:

$$z_t = \sigma(w_z \cdot [h_{t-1}, x_t] + b_z) \tag{7}$$

$$r_t = \sigma(w_r \cdot [h_{t-1}, x_t] + b_r) \tag{8}$$

$$\widetilde{h}_t = tanh(w_h \cdot [r_t \odot h_{t-1}, x_t] + b_h) \tag{9}$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \widetilde{h}_t \tag{10}$$

where z_t represents the update gate, r_t is the reset gate, $\widetilde{h}_t$ is the output candidate, w_z, w_r, and w_h are weights for the respective gates, and b_z, b_r, and b_h are biases.

The structures of individual cells that make up the neural networks of the corresponding type are shown in Figure 6.
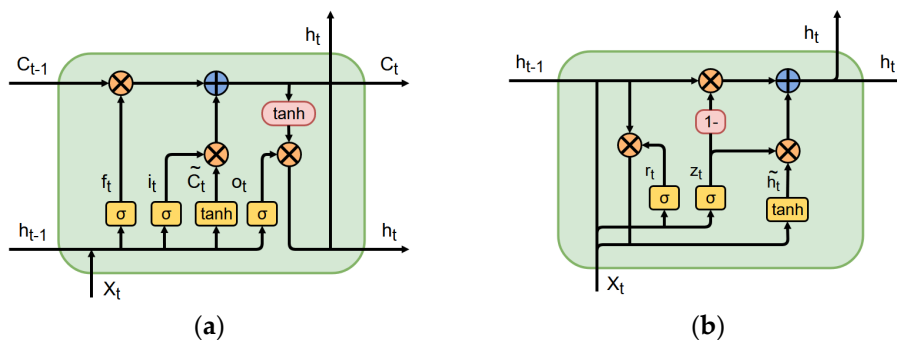
Figure 6. Schematic diagrams of the internal structure of cells: (**a**) LSTM, (**b**) GRU.

2.2.3. Temporal Convolutional Neural Networks

Convolutional neural networks have become widespread in solving image processing problems and have also been successfully applied to time series modeling. It has been shown that convolutional neural networks are able to give results comparable to recurrent neural networks or even surpass them. At the same time, they have more flexibility in meta-parameters management and also have a higher performance in terms of computation time.

The temporal convolutional network consists of a set of dilated causal convolution layers that are processing time steps of each input sequence.

The common approach for 1 d temporal convolution implementation is using a stack of consecutive residual blocks including several dilated convolutional layers and an arbitrary selection of utility, normalization, and activation layers, supplemented by an optional convolution skip connection.

Assuming that d is the dilation factor, x is the input sequence, s is the element of the input sequence, f is the applied filter, k is the filter size, and $s - d \cdot i$ represents the dilation in the backward direction, the 1D dilated convolution can be represented as follows:

$$F(s) \; = \; (x *_d f)(s) \; = \; \sum_{i=0}^{k-1} f(i) \cdot x_{s-d \cdot i}. \tag{11}$$

Depending on the dilation factor, this equation can be reduced either to the usual convolution when $d = 1$ or can be used to increase the receptive field of a neural network by introducing different time steps between filter passes.

2.2.4. Bidirectional Neural Networks

Bidirectional recurrent neural networks are a modification of conventional recurrent neural networks designed to improve their learning capabilities. This architecture is widely used in tasks where the input data has a hard-to-determine or complex time dependence as well as in solving context-sensitive problems or when establishing a connection in a series of consequential events.

While conventional LSTM and GRU neural networks are processing data in one direction, learning the temporal dependencies according to the input sequence order, their learning capabilities can be further enhanced by introducing the bidirectional architecture. Bidirectional RNN consists of cells that are arranged into two layers, one of which processes the original sample in the forward direction and the second one, in the backward direction. The outputs of each of the intermediate layers are concatenated to form a combined output. This allows the network to determine the temporal dependencies in the input information from both the past and the future during the learning process, rather than simply relying on inputs corresponding to the current time step or previous steps. A similar data processing technique can be applied to convolutional neural networks. The schematic diagram of data processing in a bidirectional recurrent neural network is shown in Figure 7.
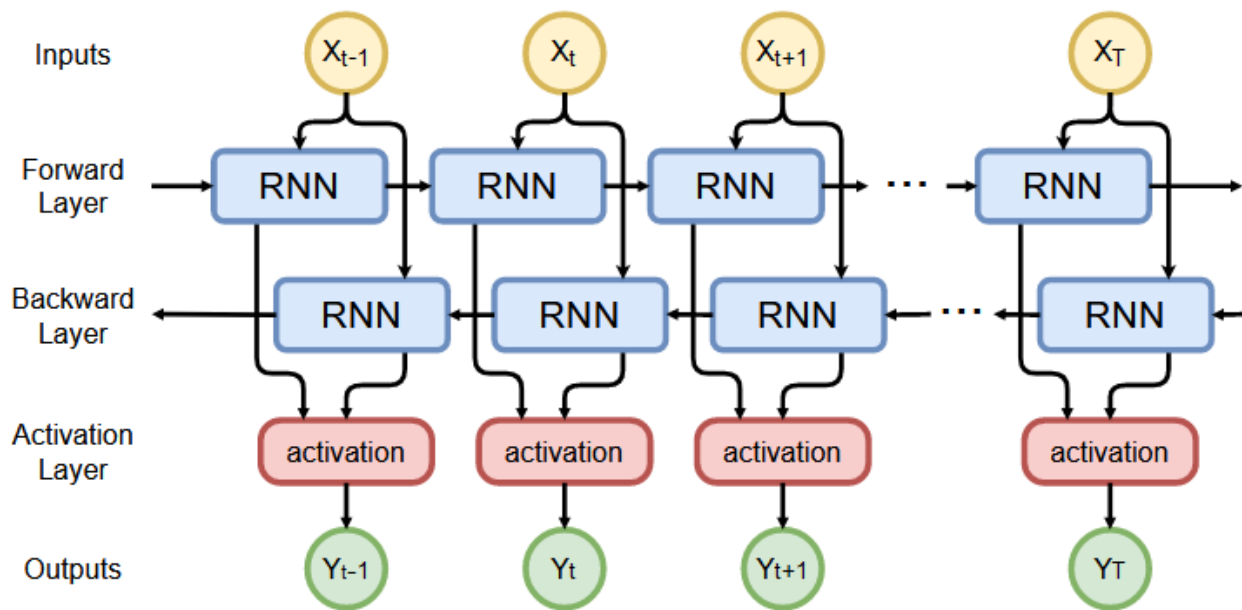
Figure 7. Schematic diagram of the bidirectional recurrent neural network.

2.2.5. The Proposed Deep Neural Networks for TEC Forecasting

The architectures considered in the work can be divided into two main categories—recurrent neural networks and temporal convolutional neural networks. Each of these categories, in turn, is further subdivided into conventional and bidirectional. Hereinafter, the following notion will be employed: the prefix "Bi" denotes bidirectional architecture; for recurrent neural networks, the postfix "RFF" or "FRF" is used to indicate the recurrent layer position in the layer stack, defining the line of architectures. The postfix of the abbreviation "RFF" denotes a recurrent layer preceding two fully connected layers, and the ending "FRF" corresponds to a recurrent layer enclosed between two fully connected layers. Activation functions and normalization layers are not noted in the abbreviations. The "LSTM" or "GRU" sub-prefix specifies whether the long short-term memory or gated recurrent unit layer is used as a recurrent layer; temporal convolutional networks are marked using the "TCN" postfix.

During the first stage of this research, the development of RFF and FRF multilayer architectures was undertaken. The architectures of the RFF line include the recurrent layer preceding two fully connected layers, while the FRF-line architectures consist of the recurrent layer enclosed between the two fully connected layers. The schematic diagrams of the proposed recurrent architectures are shown in Figure 8. The value of *N* determines the number of cells in each respective layer. The recurrent layer included in the architecture differs in each individual case and is defined either as LSTM, or as GRU, or as BiLSTM, or as BiGRU. These layers were followed by a parametric rectified linear unit (PReLU) activation function with the slope parameter specified using parentheses. PReLU multiplies the negative input values by the given parameter. For RFF architectures, the batch normalization layer was included to perform interlayer input normalization across mini-batches and to stabilize the learning process. The same approach applied to the FRF architecture provided no positive feedback, thus the batch normalization layer is not included.
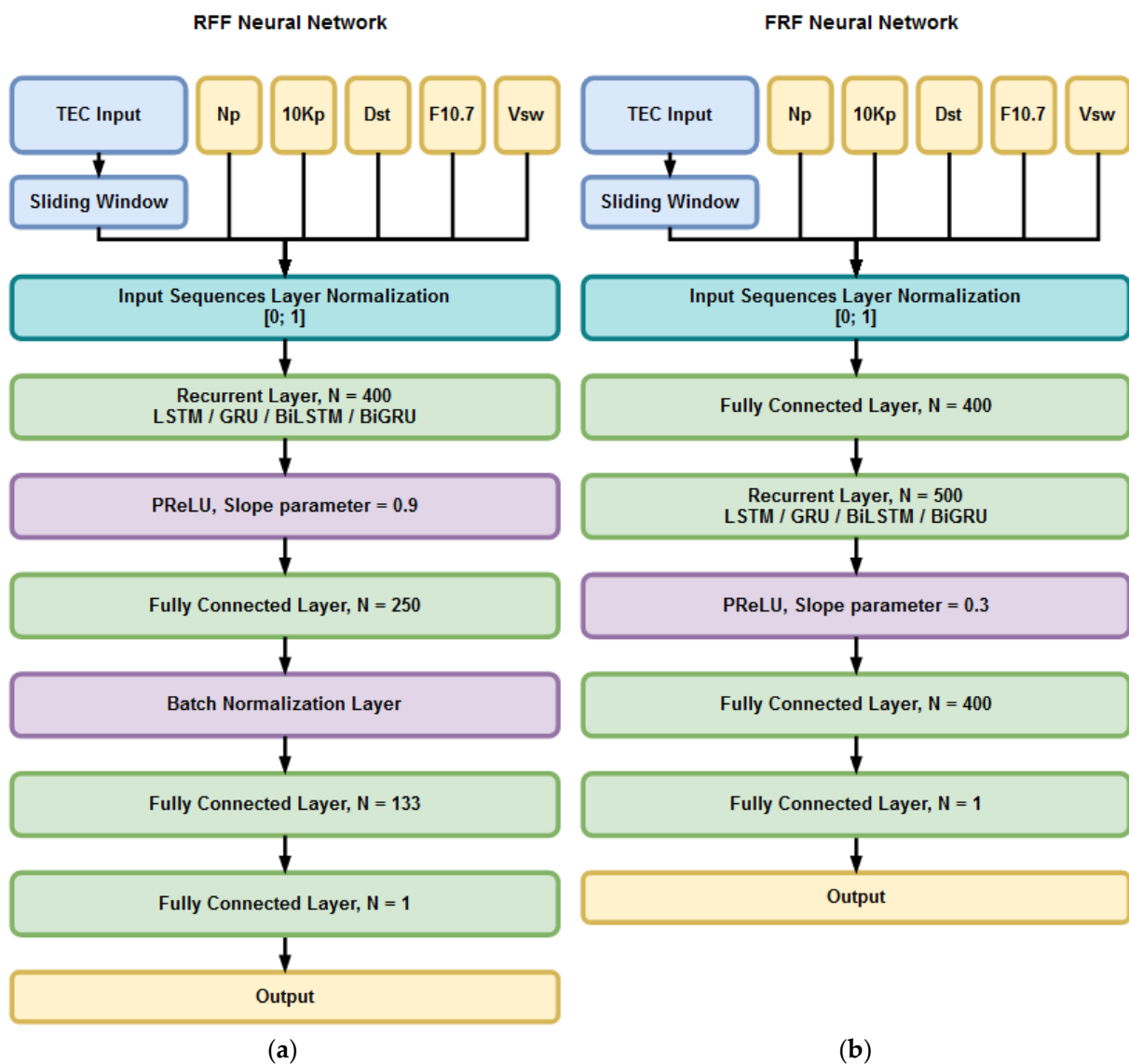
Figure 8. Schematic diagrams of the proposed multilayer recurrent neural network architectures: (**a**) RFF, (**b**) FRF. The selected recurrent layer and indices vary in each specific case.

Weight initialization in fully connected layers is performed using the Glorot initializer [25]. Input and recurrent weights for recurrent layers were initialized using an orthogonal matrix decomposition initializer [26]. For recurrent layers, the tanh activation function is used for state activation and sigmoid is used for gate activation, and each recurrent layer is followed by the PReLU activation function. The schematic diagram of the proposed temporal convolutional network is presented in Figure 9.

Temporal Convolutional Network
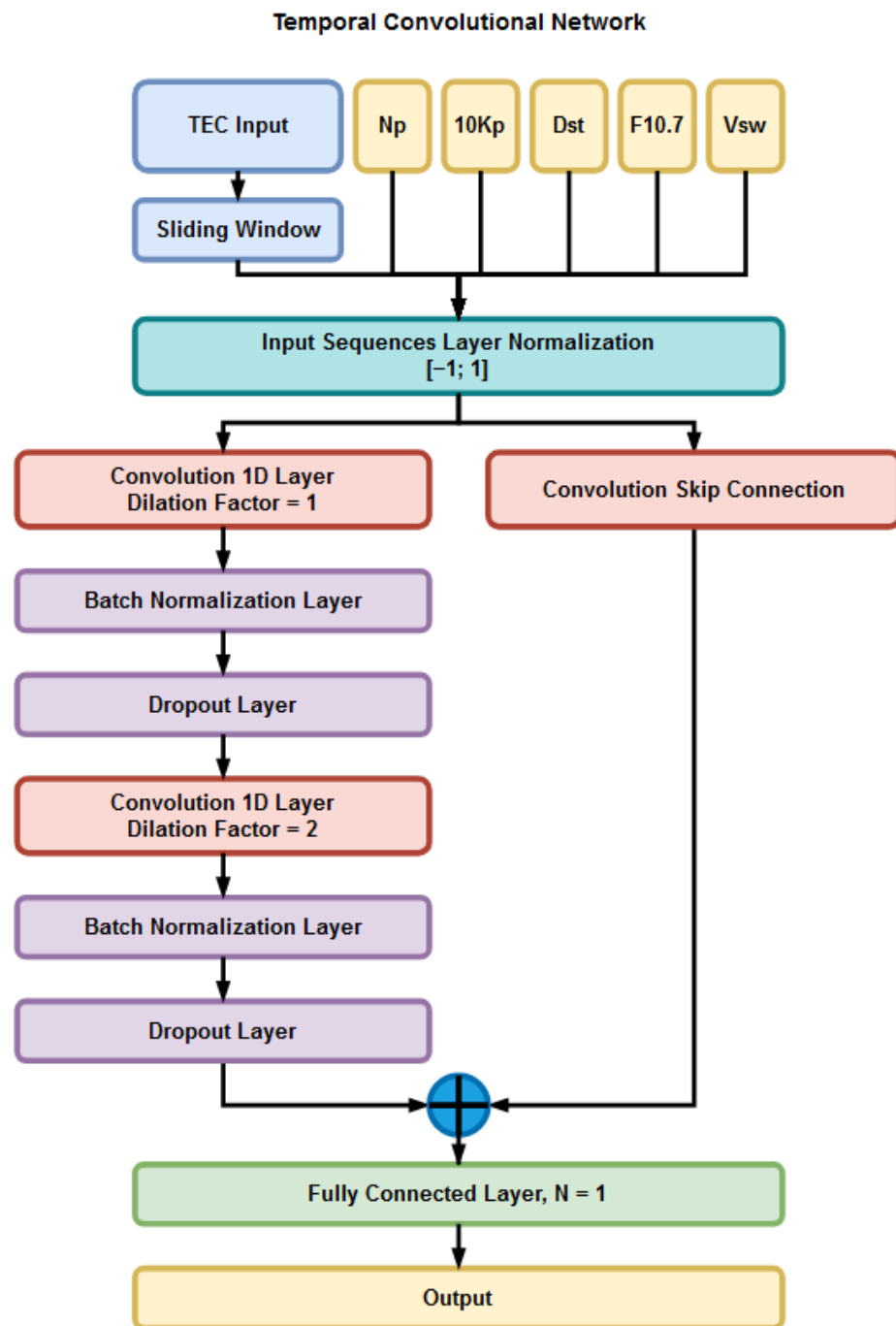
Figure 9. Schematic diagram of the proposed temporal convolutional network.

For temporal convolutional networks, the number of filters was equal to 256 with a filter kernel size of 2. Two dropout layers with a dropout factor of 0.005 were used in the main branch. The outputs are connected using addition. Per-channel symmetric rescaling was used for input data normalization so that each feature was in the range of $[-1; 1]$. For conventional temporal convolutional networks, the optional convolution skip connection was used. For the bidirectional temporal convolutional network, the convolution skip layer was connected in parallel to the layers forming the bidirectional part. The schematic diagram of the bidirectional temporal convolutional network is presented in Figure 10.
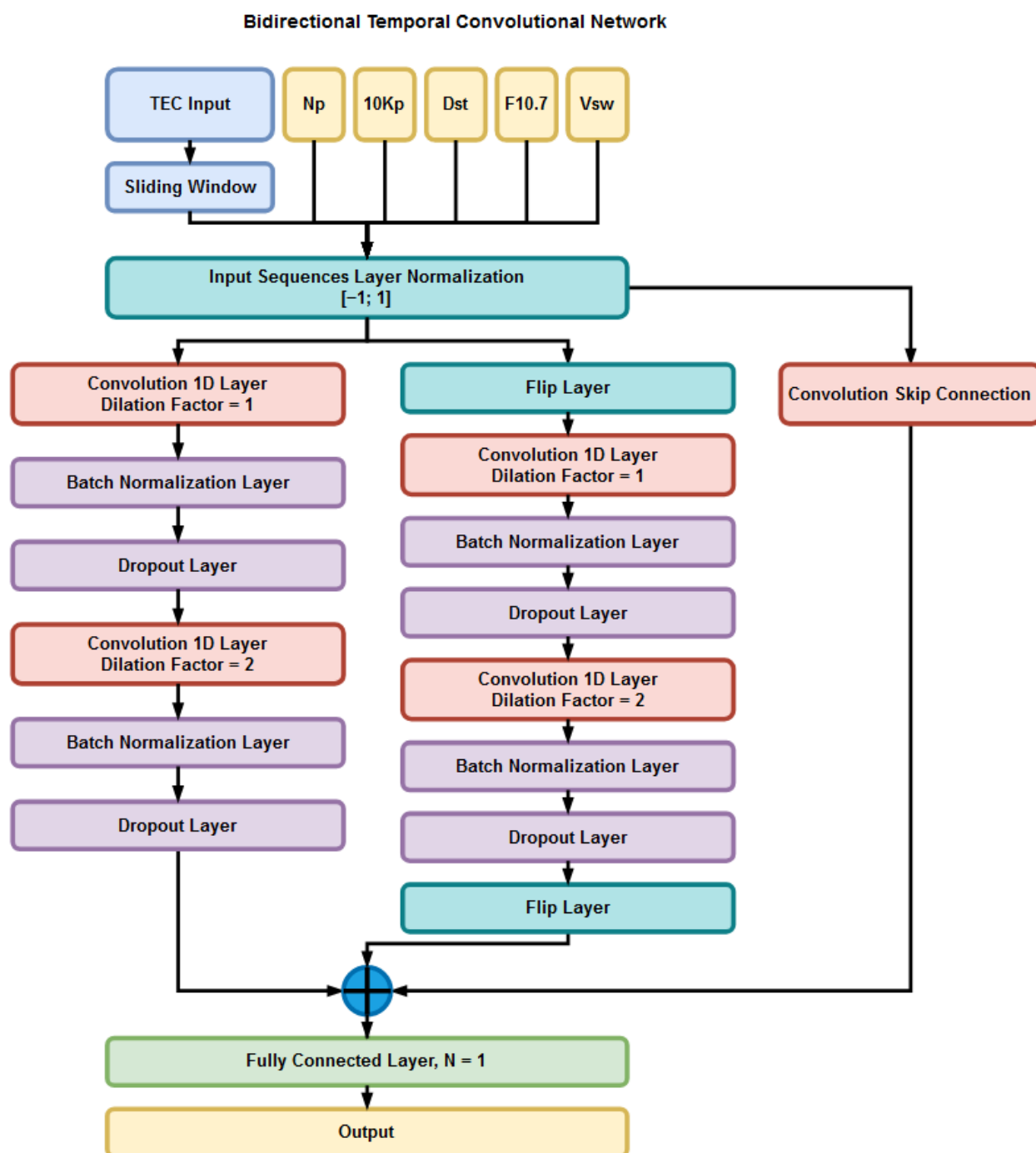
Figure 10. Schematic diagram of the proposed bidirectional temporal convolutional network.

The training was conducted using MATLAB environment and Deep Learning and Parallel Computing Toolboxes. In all the considered cases, both for bidirectional and conventional architectures, the Adam optimizer with the following parameters was used during the training:

GradientDecayFactor: 0.9000;
Squared GradientDecay Factor: 0.9990;
Epsilon: 1×10^{-8};
InitialLearnRate: 0.0030;
Learn RateSchedule: 'none';
LearnRateDropFactor: 0.1000;

LearnRateDropPeriod: 10;
L2Regularization: 0.0001;
GradientThresholdMethod: 'l2norm';
GradientThreshold: Inf;
MaxEpochs: 100;
MiniBatch Size: 1;
ValidationFrequency: 20;
ValidationPatience: Inf;
Shuffle: 'every-epoch';
ExecutionEnvironment: 'gpu';
Sequence Length: 'longest';
SequencePaddingValue: 0;
SequencePaddingDirection: 'left';
ResetInputNormalization: 0.

The mean squared error (MSE) was chosen as a loss function intended for training neural networks, minimizing the difference between real and predicted values during the learning process:

$$loss \; = \; \frac{1}{n}\sum_{i\,=\,1}^{n}(\hat{y}_i - y_i)^2, \tag{12}$$

here and below, $\hat{y}_i$ stands for the predicted values and y_i stands for the observed values, and n is the number of samples.

2.2.6. Statistical Metrics Adopted in the Validation Process

To compare the accuracy of forecasting TEC values using different architectures, a set of statistical metrics was adopted, including Mean Absolute Error, Mean Absolute Percentage Error, root mean square error, and Pearson's correlation coefficient.

The computation is performed using the following equations:

$$MAE \; = \; \frac{1}{n}\sum_{i\,=\,1}^{n}|\hat{y}_i - y_i|, \tag{13}$$

$$MAPE \; = \; \frac{1}{n}\sum_{i\,=\,1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \times 100\%, \tag{14}$$

$$RMSE \; = \; \sqrt{\sum_{i\,=\,1}^{n}\frac{(\hat{y}_i - y_i)^2}{n}}, \tag{15}$$

$$\rho \; = \; \frac{cov(y, \hat{y})}{\sigma_y \cdot \sigma_{\hat{y}}}, \tag{16}$$

where n is the number of samples, $cov(y, \hat{y})$ is the covariance between the predicted and observed values, σ_y is the standard deviation of the observed values, and $\sigma_{\hat{y}}$ is the standard deviation of the predicted values.

3. Results

The statistical results on accuracy of TEC forecasting for a set of 10 considered architectures and the two sets of indices (including and excluding Vsw) are presented in Table 1.

Table 1. Forecasting results for Juliusruh station in 2015.

	Architecture and Indices	MAE (TECU)	MAPE (%)	RMSE (TECU)	ρ
1	LSTM-FRF(10 Kp, Np, F10.7, Dst)	1.141	11.915	1.574	0.967
2	LSTM-FRF(10 Kp, Np, F10.7, Dst, Vsw)	1.150	12.156	1.594	0.967
3	LSTM-RFF(10 Kp, Np, F10.7, Dst)	1.187	12.309	1.623	0.969
4	LSTM-RFF(10 Kp, Np, F10.7, Dst, Vsw)	1.207	12.301	1.669	0.967
5	GRU-FRF(10 Kp, Np, F10.7, Dst)	1.163	11.261	1.586	0.967
6	GRU-FRF(10 Kp, Np, F10.7, Dst, Vsw)	1.273	13.430	1.681	0.965
7	GRU-RFF(10 Kp, Np, F10.7, Dst)	1.152	11.890	1.569	0.971
8	GRU-RFF(10 Kp, Np, F10.7, Dst, Vsw)	1.199	12.521	1.634	0.968
9	TCN(10 Kp, Np, F10.7, Dst)	1.047	11.118	1.426	0.973
10	TCN(10 Kp, Np, F10.7, Dst, Vsw)	1.078	11.649	1.453	0.973
11	BiLSTM-FRF(10 Kp, Np, F10.7, Dst)	1.480	14.371	1.943	0.960
12	BiLSTM-FRF(10 Kp, Np, F10.7, Dst, Vsw)	1.308	14.423	1.686	0.966
13	BiLSTM-RFF(10 Kp, Np, F10.7, Dst)	0.583	6.420	0.797	0.992
14	BiLSTM-RFF(10 Kp, Np, F10.7, Dst, Vsw)	0.570	6.441	0.870	0.990
15	BiGRU-FRF(10 Kp, Np, F10.7, Dst)	0.621	6.952	0.800	0.992
16	BiGRU-FRF(10 Kp, Np, F10.7, Dst, Vsw)	0.835	9.497	1.067	0.986
17	BiGRU-RFF(10 Kp, Np, F10.7, Dst)	0.473	5.231	0.740	0.993
18	BiGRU-RFF(10 Kp, Np, F10.7, Dst, Vsw)	0.544	6.115	0.771	0.992
19	BiTCN(10 Kp, Np, F10.7, Dst)	0.191	2.615	0.555	0.996
20	BiTCN(10 Kp, Np, F10.7, Dst, Vsw)	0.161	2.192	0.507	0.997

The metrics considered were MAE, MAPE, RMSE, and ρ, averaged over half a year during single step 2 h ahead TEC forecasting for three years separately and four selected stations. The results for station Juliusruh are presented in Table 1 for 2015.

The results for every station considered are shown in Figure 11 as a dependence of MAE, MAPE, and RMSE on the architecture's variant number. Thus, the results are presented for the following architectures, indicated by numbers: (1) LSTM-FRF, (2) LSTM-RFF, (3) GRU-FRF, (4) GRU-RFF, (5) TCN, (6) BiLSTM-FRF, (7) BiLSTM-RFF, (8) BiGRU-FRF, (9) BiGRU-RFF, and (10) BiTCN. Due to the fact that there is no unambiguous trend in the influence of the Vsw index on the accuracy of the forecast, only the option with the best result (with and without Vsw) was chosen from the two options.

From Figure 11, it can be seen that among the prediction accuracy metrics, three sections can be distinguished with similar characteristics, depending on the architectures used. The first flat area (numbers 1–5) with the highest values for errors refers to traditional LSTM and GRU architectures with FRF and RFF modifications. The second section (numbers 6–8), which includes results for bidirectional networks, has a minor slope and error values are 1.5–2 times smaller compared to the first section. Finally, the third section, which includes BiGRU and BiTCN architectures, provides the lowest error values, i.e., the highest forecasting accuracy. It should be noted that although there is a certain latitude dependence of the statistical characteristics in the first two groups, for the BiTCN architecture, it practically ceases to exist.
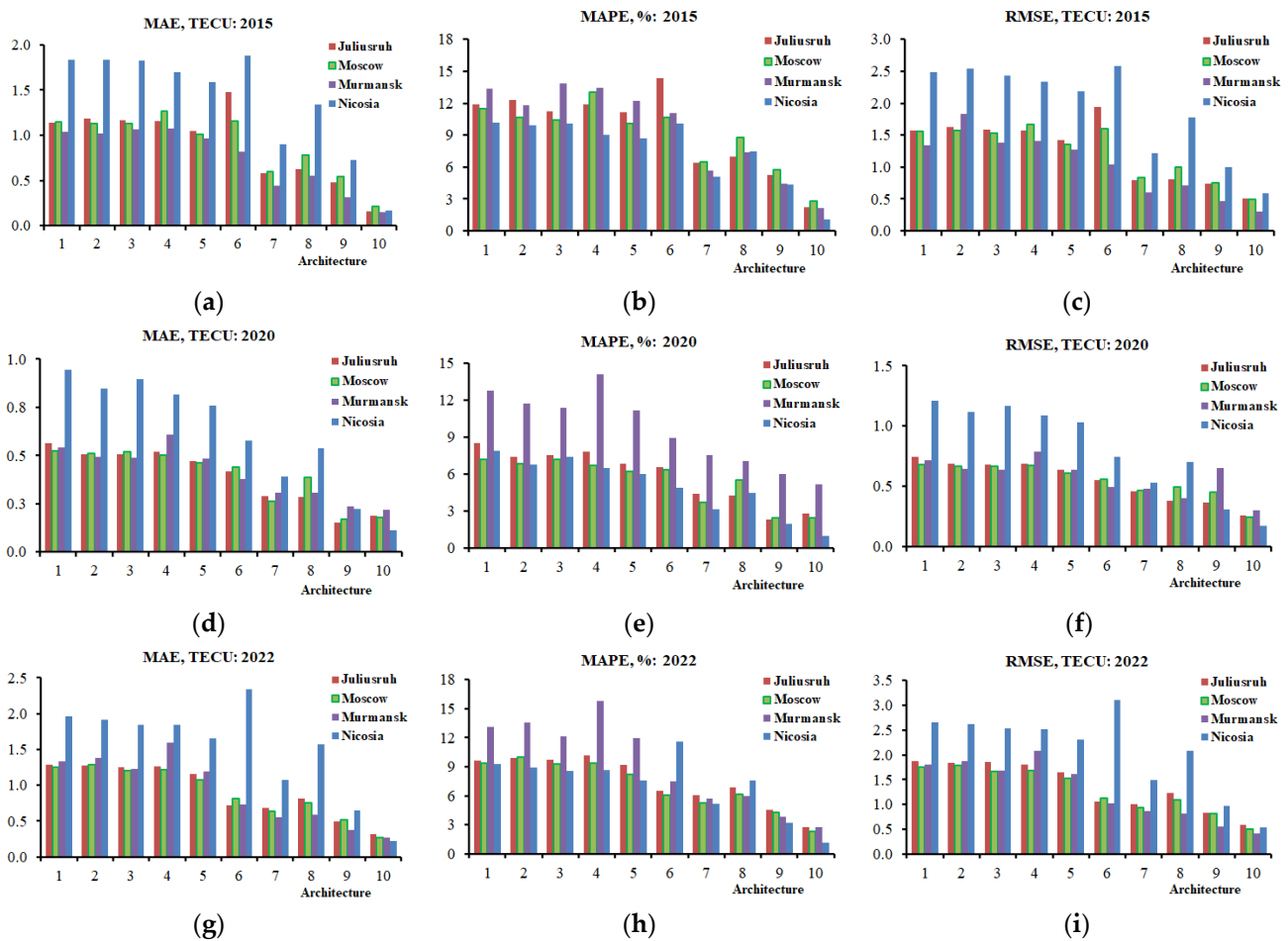
Figure 11. Statistical characteristics of TEC forecast in the European zone depending on a combination of architectures and indices for 2015 (**a–c**), 2020 (**d–f**), and 2022 (**g–i**).

It can be seen that MAPE and RMSE are related to the absolute values of TEC, depending on the level of solar activity. Conventional LSTM and GRU with FRF and RFF modifications and conventional TCN provide results at the same level of accuracy. The usage of bidirectional architecture essentially improves the accuracy of the forecast for almost all architectures and across all stations. It should be noted that during high solar activity, bidirectional architectures of the FRF line (BiLSTM-FRF and BiGRU-FRF) do not improve the accuracy of the forecast for a low-latitude station even compared to conventional methods. The highest accuracy is achieved using BiGRU-RFF and BiTCN architectures: MAPE values are less than 0.6 TECU, RMSE values are less than 0.8 TECU, and MAPE values are less than 7.5%. The best forecasting accuracy is achieved using BiTCN architecture, with less than 0.3 TECU for MAE, less than 0.6 TECU for RMSE, and less than 5% for MAPE.

The forecast errors during the second half of each year for the Moscow station are shown in Figure 12. Each panel for the 2015, 2020, and 2022 years shows the variations of TEC between DoY 185 and 365 along with the Dst index and compares the prediction accuracy for two architectures: LSTM-FRF and BiTCN. All the cases are provided with the Vsw index included in the training data (10 Kp, Np, F10.7, Dst, and Vsw).
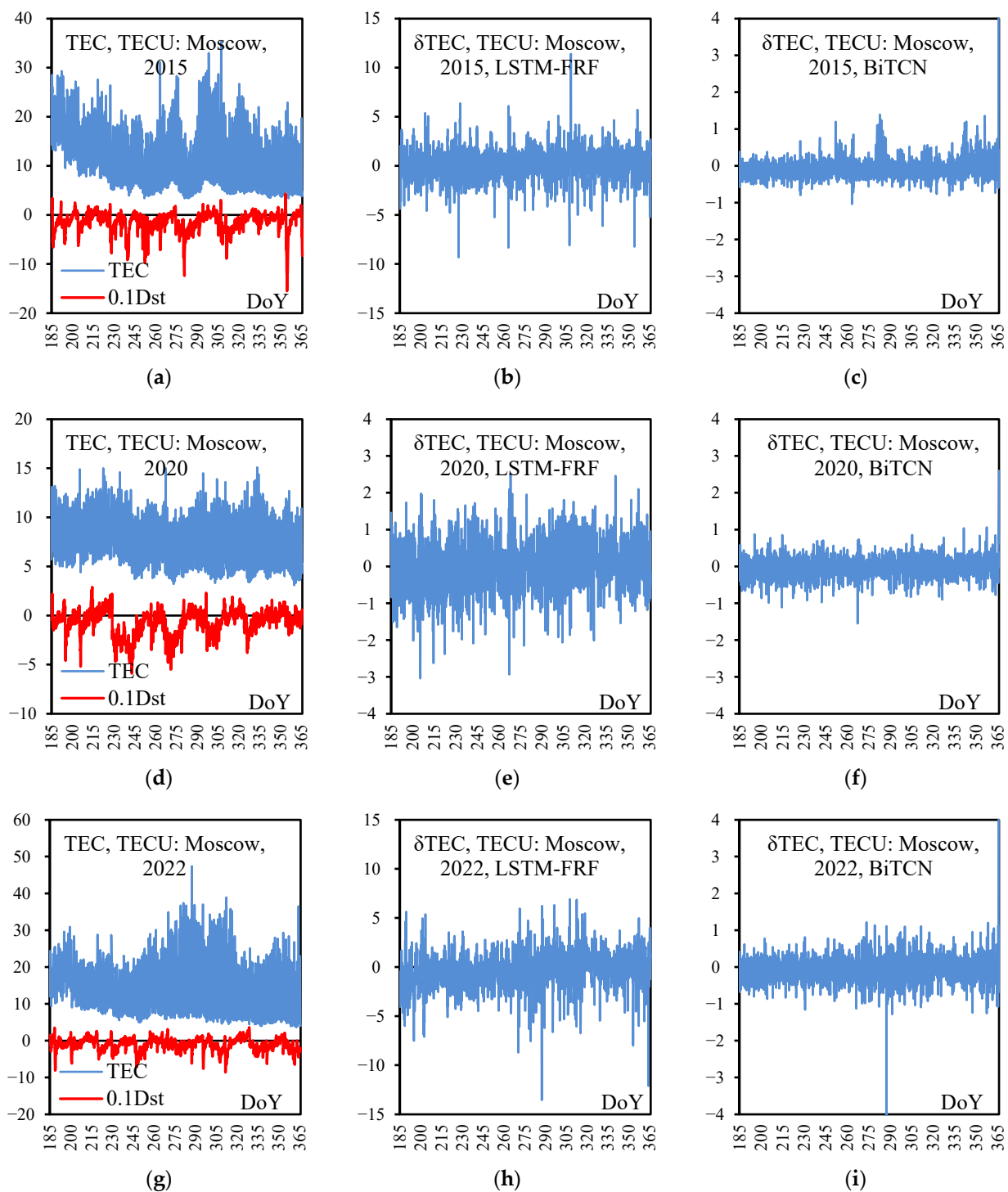
Figure 12. Comparison of the prediction accuracy for two architectures—LSTM-FRF and BiTCN: 2015 (**a–c**), 2020 (**d–f**), and 2022 (**g–i**).

Figure 12 shows the instantaneous TEC and δTEC values for the two selected architectures from among the unidirectional and bidirectional ones, using the Moscow station as an example. Unlike the median values, the TEC values (plots a, d, and g on the left) show the response to disturbed conditions. The average δTEC values for the first and third plots (b,c) are 1.15 TECU and 0.22 TECU in 2015, an improvement of ~5 times can be observed. In 2020 (Figure 12e,f), these values were 0.52 TECU and 0.2 TECU, i.e., an improvement of ~2.5 times, and for the year 2022 (Figure 12h,i), higher values of 1.32 TECU

and 0.27 TECU were obtained and an improvement of ~5 times was achieved. Regardless of the year, the accuracy for BiTCN is 2–5 times higher than for LSTM-FRF: the error is in the range of ±1 TECU. The δTEC outlier in the last panel refers to 14 October 2022, when a positive disturbance occurred during the main phase of the magnetic storm (minimum Dst = −62 nT) due to a sharp increase in Vsw, which was minimal and equal to 300 km/s in UT0 on 14 October and gradually increased to 600 km/s in UT0 on 16 October. It is interesting to note that during the main phase of the second magnetic storm (minimum Dst = 76 nT), there was also a positive deviation δTEC, but of less intensity, since the velocity Vsw was constant, 355 km/s.

4. Discussion

In this paper, a set of architectures of multilayer neural networks based on GRU, LSTM, and temporal convolution is proposed and considered. The implementation of the RFF and FRF architectures improves the results compared to [24]. The further modification of those architectures using bidirectional processing led to a 2-fold improvement in the forecast accuracy. The greatest improvement in accuracy of the forecast was achieved with the employment of a bidirectional temporal convolutional network architecture, though its initial variant provides results at a level comparable to traditional methods. Its bidirectional enhancement allows one to increase the accuracy of the forecast by 5–10 times. The observed dependence of the results on solar activity, a combination of indices, and the latitude shows that to determine the best method in each region, it is necessary to examine not only a set of architectures but also indices, as evidenced by the example with the Vsw index.

It is important to note that the number of cells in bidirectional layers is two times higher than in unidirectional ones, but at the same time, a twofold increase in the number of cells in unidirectional networks did not lead to an increase in prediction accuracy but, on the contrary, reduced it. This result is due to overfitting. The bidirectional neural network includes two separate backpropagation passes: one is for the forward direction of inputs, and the second one is for the reverse direction. During the forward pass, the forward layer processes the input sequence in the original input direction and makes predictions for the output sequence. The backward layer processes the input sequence in reverse order during the backward pass and predicts the output sequence. When both passes are completed, their predictions are compared to the target output sequences and the error is propagated back to each respective layer through the network to update the weights. The forward and backward layer weights are updated based on the errors computed during the forward and backward passes, respectively. This process allows you to optimize the weights on a double amount of data—the original inputs and reverse inputs. For each of the layers of a bidirectional network, optimization is carried out on its own subspace of parameters.

As for the comparison with the results of other works, as noted in the introduction, in many works, only MAE and RSME are mentioned as statistical characteristics, but MAPE is not indicated. The results obtained show that this is not sufficient. This is especially evident from the results for the Nicosia station, which has the highest TEC values. This can be demonstrated by the example of [23], where RMSE ~0.5 TECU were obtained without specifying MAPEs. In this work, even smaller RMSE values were obtained regardless of latitude, indicating that the capabilities of the LSTM, GRU, and TCN methods for TEC prediction have not been exhausted.

Apart from the differences in accuracy, it is important to note that significant discrepancies between the architectures in terms of computational costs are also observed. For bidirectional architectures, the computational costs are significantly higher than for unidirectional ones. In particular, among unidirectional architectures, for the FRF line of architectures, the training time was about 11 s, for the RFF line, the training time was about 9 s, and for the TCN, it was 2 s on average. The difference between the LSTM and GRU variants did not exceed 2 s on average, and the GRU-based architectures were converging at a higher rate. With the transition to bidirectional architectures, the training time for FRF and RFF architectures increased by more than 2 times and amounted to 27 and 22 s on

average. For BiTCN, the training time increased to 4 s. The training was performed using a single RTX4080 and AMD Ryzen 9 7900X.

It is important to note that the training was carried out on a relatively small amount of data that did not contain complete information about the annual behavior of TEC and the indices used; nevertheless, the proposed architectures of bidirectional networks completed the forecasting task with consistently high accuracy.

5. Conclusions

This work was devoted to finding a combination of the latest neural network architectures, which, under specific conditions of TEC variations over three years, according to data from four European stations can provide the most accurate forecast with an advance of 2 h. Of the three selected years, 2015 belonged to the phase of solar activity decline in cycle 24, 2020 had minimal activity during the transition from one cycle to another, and 2022 came to the phase of solar activity increase in cycle 25. According to the annual mean values of TEC and Dst, the year 2022 had higher solar activity than 2015 but lower geomagnetic activity. It is shown that the maximum values of TEC change in accordance with the change in solar activity, which leads to the same trend of changes in the statistical characteristics of the forecast, i.e., the level of solar activity played a greater role than the geomagnetic activity.

The use of solar and geomagnetic activity indices is known to affect the accuracy of the TEC forecast; therefore, when choosing the most effective index, estimates were made for a wide set. However, unlike the results presented in [24], a significant contribution of the Vsw index was observed and preliminary calculations were carried out for all architectures for the two sets of parameters, including and excluding the Vsw index as an input parameter. The results for these cases differed markedly; therefore, out of 20 options, only 10 architectures with the best statistical characteristics (with or without Vsw) were selected for presentation. It is significant that among the 10 presented architectures, there are groups of them that provide the same accuracy. So, the traditional LSTM and GRU methods gave estimates [24] at a level corresponding to the literature data. Modifications of the FRF and RFF architecture lines led to a minor increase in accuracy. The use of bidirectional models led to an additional increase in accuracy by a factor of 1.5–2 as compared to these methods. The next group, including BiGRU and BiTCN architectures, provided the lowest errors, i.e., the highest forecasting accuracy. The error estimates for the BiTCN architecture were the best, with less than 0.3 TECU for MAE, less than 0.6 TECU for RMSE, and less than 5% for MAPE.

Thus, it was found that the combined use of bidirectional and TCN architectures significantly increases the accuracy of the forecast and levels out the latitudinal dependence. It is planned to apply this approach to 24 h ahead forecasting.

Author Contributions: Conceptualization, O.M.; methodology, A.K. and O.M.; software, A.K.; validation, O.M.; formal analysis, A.K. and O.M.; investigation, A.K. and O.M.; resources, A.K.; data curation, O.M.; writing—original draft preparation, A.K. and O.M.; writing—review and editing, A.K. and O.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was financially supported by the Ministry of Science and Higher Education of the Russian Federation (State task in the field of scientific activity 2023).

Data Availability Statement: The OMNI data (F10.7, Np, 10 Kp, Dst, and Vsw indices) were obtained from the GSFC/SPDF OMNIWeb interface at https://omniweb.gsfc.nasa.gov (accessed on 22 January 2022). The data on JPL, GIM, and TEC were obtained from IONEX files located at https://urs.earthdata.nasa.gov (accessed on 10 February 2023).

Acknowledgments: The TEC samples data was obtained from https://urs.earthdata.nasa.gov (accessed on 10 February 2023). Data on solar and geomagnetic activity was taken from SPDF OMNI Web Service http://omniweb.gsfc.nasa.gov/form/dx1.html (accessed on 10 February 2023).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. McGranaghan, R.M.; Camporeale, E.; Georgoulis, M.; Anastasiadis, A. Space Weather research in the Digital Age and across the full data lifecycle: Introduction to the Topical Issue. *J. Space Weather Space Clim.* **2021**, *11*, 50. [CrossRef]
2. Goodman, J. Operational communication systems and relationships to the ionosphere and space weather. *Adv. Space Res.* **2005**, *36*, 2241–2252. [CrossRef]
3. Xie, T.; Dai, Z.; Zhu, X.; Chen, B.; Ran, C. LSTM-based short-term ionospheric TEC forecast model and positioning accuracy analysis. *GPS Solut.* **2023**, *27*, 66. [CrossRef]
4. Boulch, A.; Cherrier, N.; Castaings, T. Ionospheric activity prediction using convolutional recurrent neural networks. *arXiv* **2018**, arXiv:1810.1327312.
5. Yu, S.; Ma, J. Deep learning for geophysics: Current and future trends. *Rev. Geophys.* **2021**, *59*, e2021RG000742. [CrossRef]
6. Hochreiter, S.; Schmidhuber, J. Long Short-term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]
7. Cho, K.; van Merrienboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. *arXiv* **2014**, arXiv:1406.1078v3.
8. Lei, D.; Liu, H.; Le, H.; Huang, J.; Yuan, J.; Li, L.; Wang, Y. Ionospheric TEC Prediction Base on Attentional BiGRU. *Atmosphere* **2022**, *13*, 1039. [CrossRef]
9. Lea, C.; Flynn, M.D.; Vidal, R.; Reiter, A.; Hager, G.D. Temporal Convolutional Networks for Action Segmentation and Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1003–1012. [CrossRef]
10. Sun, W.; Xu, L.; Huang, X.; Zhang, W.; Yuan, T.; Chen, Z.; Yan, Y. Forecasting of ionospheric vertical total electron content (TEC) using LSTM networks. In Proceedings of the 2017 International Conference on Machine Learning and Cybernetics, Ningbo, China, 9–12 July 2017.
11. Sun, W.; Xu, L.; Huang, X.; Zhang, W.; Yuan, T.; Yan, Y. Bidirectional LSTM for ionospheric vertical total electron content (TEC) forecasting. In Proceedings of the IEEE Visual Communications and Image Processing (VCIP), St. Petersburg, FL, USA, 10–13 December 2017; pp. 1–4.
12. Sivakrishna, K.; Venkata Ratnam, D.; Sivavaraprasad, G. A Bidirectional Deep-Learning Algorithm to Forecast Regional Ionospheric TEC Maps. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 4531–4543. [CrossRef]
13. Chen, Z.; Liao, W.; Li, H.; Wang, J.; Deng, X.; Hong, S. Prediction of global ionospheric TEC based on deep learning. *Space Weather* **2022**, *20*, e2021SW002854. [CrossRef]
14. 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. [CrossRef]
15. Chunli, D.; Jinsong, P. Modeling and prediction of TEC in China region for satellite navigation. In Proceedings of the 15th Asia-Pacific Conference on Communications, Shanghai, China, 8–10 October 2009; pp. 310–313.
16. Huang, Z.; Yuan, H. Ionospheric single-station TEC short-term forecast using RBF neural network. *Radio Sci.* **2014**, *49*, 283–292.
17. Niu, R.; Guo, C.; Zhang, Y.; He, L.; Mao, Y. Study of ionospheric TEC short-term forecast model based on combination method. In Proceedings of the 2014 12th International Conference on Signal Processing (ICSP), Hangzhou, China, 19–23 October 2014; pp. 2426–2430.
18. Iluore, K.; Lu, J. Long short-term memory and gated recurrent neural networks to predict the ionospheric vertical total electron content. *Adv. Space Res.* **2022**, *70*, 652–665. [CrossRef]
19. Kaselimi, M.; Voulodimos, A.; Doulamis, N.; Doulamis, A.; Delikaraoglou, D. Deep Recurrent Neural Networks for Ionospheric Variations Estimation Using GNSS Measurements. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–15. [CrossRef]
20. Kaselimi, M.; Voulodimos, A.; Doulamis, N.; Doulamis, A.; Delikaraoglou, D. A causal long short-term memory sequence to sequence model for TEC prediction using GNSS observations. *Remote Sens.* **2020**, *12*, 1354. [CrossRef]
21. Tang, R.; Zeng, F.; Chen, Z.; Wang, J.-S.; Huang, C.-M.; Wu, Z. The comparison of predicting storm-time ionospheric TEC by three methods: ARIMA, LSTM, and Seq2Seq. *Atmosphere* **2020**, *11*, 316. [CrossRef]
22. Cesaroni, C.; Spogli, L.; Aragon-Angel, A.; Fiocca, M.; Dear, V.; De Franceschi, G.; Romano, V. Neural network based model for global Total Electron Content forecasting. *J. Space Weather Space Clim.* **2020**, *10*, 11. [CrossRef]
23. Natras, R.; Soja, B.; Schmidt, M. Ensemble Machine Learning of Random Forest, AdaBoost and XGBoost for Vertical Total Electron Content Forecasting. *Remote Sens.* **2022**, *14*, 3547. [CrossRef]
24. Kharakhashyan, A.; Maltseva, O.; Glebova, G. Forecasting the total electron content TEC of the ionosphere using space weather parameters. In Proceedings of the 2021 IEEE International Conference on Wireless for Space and Extreme Environments (WiSEE), Cleveland, OH, USA, 12–14 October 2021; pp. 31–36. [CrossRef]
25. Glorot, X.; Bengio, Y. Understanding the Difficulty of Training Deep Feedforward Neural Networks. In Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, AISTATS, Sardinia, Italy, 13–15 May 2010; pp. 249–356.
26. Saxe, A.M.; McClelland, J.L.; Ganguli, S. Exact solutions to the nonlinear dynamics of learning in deep linear neural networks. *arXiv* **2013**, arXiv:1312.6120.