

APPLICATION OF CONVOLUTION NEURAL NETWORKS FOR CRITICAL FREQUENCY f_oF2 PREDICTION

B.G. Salimov

*Institute of Solar-Terrestrial Physics SB RAS,
Irkutsk, Russia, salimov@iszf.irk.ru*

O.I. Berngardt 

*Institute of Solar-Terrestrial Physics SB RAS,
Irkutsk, Russia, berng@iszf.irk.ru*

A.E. Khmel'nov 

*Matrosov Institute for System Dynamics and Control Theory
SB RAS,
Irkutsk, Russia, hmel'nov@icc.ru*

K.G. Ratovskiy

*Institute of Solar-Terrestrial Physics SB RAS,
Irkutsk, Russia, ratovskiy@iszf.irk.ru*

O.A. Kusonskiy

*Institute of Geophysics, Ural Branch RAS,
Yekaterinburg, Russia, zavlab@arti.igfuran.ru*

Abstract. Ionosphere has an important impact on the quality of radio communication, radar, and global positioning. One of the essential characteristics describing the state of the ionosphere is its critical frequency f_oF2 . Its prediction provides effective modes of operation of technical radio equipment as well as enables calculation of the corrections needed to improve the accuracy of its functioning. Different physical and empirical models are generally used for f_oF2 prediction. This paper proposes an empirical prediction technique based on machine learning methods and observational history. It relies on a regression approach to the prediction based on the known daily quasi-periodicity of ionospheric parameters related to solar illumination. Algorithmically, this approach is implemented in the form of convolutional neural networks with two-dimensional convolution. The input data for the analysis is presented

as two-dimensional solar time — date matrices. The model was trained on data from the mid-latitude ionosonde in Irkutsk (RF) and tested using data from several mid-latitude ionosondes: Arti (RF), Warsaw (Poland), Mohe (China). It is shown that the main contribution to the prediction value of f_oF2 is made by the data on the nearest few days before the prediction; the contribution of the remaining days strongly decreases. This model has the following forecast quality metrics (Pearson correlation coefficient 0.928, root mean square error 0.598 MHz, mean absolute error in percent 10.45 %, coefficient of determination 0.861) and can be applied to f_oF2 forecast in middle latitudes.

Keywords: ionosphere, machine learning, neural networks, f_oF2 .

INTRODUCTION

The ionosphere is a dynamic medium that, on the one hand, is affected by solar radiation, processes in the underlying atmosphere and the overlying magnetosphere and plasmasphere, and, on the other hand, has a certain inertia associated with ionization and recombination of constituent particles and with mass transfer. Thus, when predicting ionospheric characteristics, it is important to take into account both local dependences related to the ionosphere inertia and quasi-periodic ones associated with diurnal variation of solar radiation. Therefore, to predict ionospheric characteristics, it is effective to use models considering the state of the ionosphere in the past, as well as the history of parameters characterizing the main impact on the ionosphere from above — solar and magnetic activity.

Recently, when solving complex problems with a large number of unknowns, machine learning methods have often been adopted. These methods are a synthesis of various mathematical methods — from the approximation theory to the optimal control theory. Today, the term "machine learning" generalizes widely-known simple methods such as various regression techniques and more complex methods of classical learning (super-

vised and unsupervised) and deep learning, using complex end-to-end neural networks, reinforcement learning, etc. [Goodfellow et al., 2016]. Recently, machine learning in one form or another has often been employed to address geophysical problems [Yu, Ma, 2021] and problems of forecasting the state of the ionosphere. Sivavaraprasad et al. [2022] present a nonlinear autoregressive neural network with external input that predicts the total electron content (TEC) in the ionosphere. As input data for this model, TEC, geomagnetic index A_p values, solar activity data, time of day, geographic coordinates, etc. are utilized.

The critical frequency f_oF2 is one of the main ionospheric characteristics used for solving applied problems. In vertical sounding problems, this is the maximum frequency of an ordinary polarization radio wave reflected from the ionosphere [Hargreaves, 1982]. For oblique radio wave propagation, f_oF2 can help to estimate the maximum frequency applicable to radio communication. For global positioning with global navigation satellite systems, f_oF2 has an effect on the ionospheric correction required to improve positioning accuracy. Diurnal variation in f_oF2 as recorded by the ISTP SB RAS mid-latitude ionosonde (Irkutsk, 52°16' N, 104°17' E) is shown in Figure 1. There is a significant

were standardized and transformed from a vector-valued 3D time function (f_oF2 , $F10.7$, Dst with 1 hr time resolution) to a sequence of 90×24 matrices for each of the parameters (90 days before the current moment, 24 hours relative to the current hour) characterizing values of the corresponding parameters for the previous three months. The sequence of matrices was constructed so that each hour of each day of measurements corresponded to its own three matrices. Twenty four f_oF2 values following the current moment were target to be predicted by the three matrices (f_oF2 , $F10.7$, Dst) corresponding to the current moment.

The method of forming this dataset is as follows.

Stage 1. The entire training dataset is standardized (scaled) by types of f_oF2 , $F10.7$, Dst data. The scaled data $z(t)$ of each type is obtained from the initial one $x(t)$ as follows :

$$z(t) = (x(t) - \mu) / s, \quad (1)$$

where μ is the average value of the training dataset $x(t)$; s is the standard deviation of the training dataset $x(t)$.

Stage 2. For the three time series of f_oF2 , $F10.7$, Dst for each hour, we take a dataset of 2160 hour values preceding this hour, including this hour. As target values for training the model, we choose 24 f_oF2 values following the given hour. Each dataset of 2160 values is transformed into a 2D 90×24 matrix, where each row represents a day from 0 to 89; each column, an hour from 0 to 23.

Stage 3. For training, we generate a 4D $65736 \times 90 \times 24 \times 3$ data feature matrix and a target 65736×24 matrix, which contains daily f_oF2 sets, in other words 65736 (the number of hours in the complete dataset) 3D $90 \times 24 \times 3$ blocks and 65736 vectors 24 long (the number of forecast hours after the current hour). The generated data is divided into two datasets containing 52608 (training) and 13128 (validation) values. The fourth dimension of the data feature matrix has a dimension of 3 and is responsible for f_oF2 , $F10.7$, Dst .

Figure 2 exemplifies the standardized data on f_oF2 , $F10.7$, Dst for the time interval January 2, 2009, 00:00 LST — April 1, 2009, 23:00 LST, corresponding to April 1, 2009, 23:00 LST in the multidimensional dataset (after Stage 3). Along the vertical axis are days; along the horizontal axis are hours. The diurnal variation in f_oF2 , associated with the level of illumination, is clearly defined — to the local day correspond high values of f_oF2 ; to the local night, low values.

ARCHITECTURE OF THE NEURAL NETWORK, ITS TRAINING AND PROPERTIES

The basic assumptions underlying the neural network are as follows:

- f_oF2 is determined from a linear combination of the preceding f_oF2 , Dst , and $F10.7$ values with unknown weight coefficients;
- the weight coefficients may differ for different current moments of observations, but depend only on the shift from the local time of the current moment.

These assumptions are fulfilled in the convolutional neural network shown in Figure 3.

The initial network consists of three independent parallel convolution layers (L.I) (for each of the input parameters of f_oF2 , Dst , $F10.7$), the results of passing through which are combined (L.II) further into a $90 \times 24 \times 3$ matrix. The resulting matrices are flattened (converted into a 6480 vector) (L.III) and fed to the input of the decision layer (L.IV) — the single-layer neural network with a linear activation function, most often used in forecasting problems and performing a linear combination function of outputs of convolution layers (see Figure 3). The convolution layers L.I have been added to improve the quality of filtering of random outliers and for additional adaptive smoothing of input data obtained with different time resolutions. They have a convolution kernel with a size of 30 days \times 1 hr.

Such a network architecture is equivalent to a linear regression problem with a very large number of coefficients found by machine learning methods.

When training the network, we have applied an analogue of the gradient descent method — the adaptive momentum method (Adam), which increases the learning rate [Kingma, Va, 2014]. As a loss function, we utilized MSE — the average squared difference between the predicted values and the real one. This loss function is traditional for solving numerical value prediction problems. The neural network code is implemented in Python, using the TensorFlow library [Abadi et al., 2016].

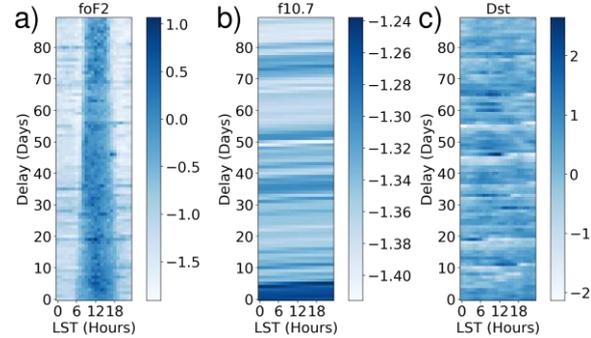


Figure 2. Initial standardized data on f_oF2 , $F10.7$, Dst from January 2, 2009, 00:00 LST to April 1, 2009, 23:00 LST (used for training the neural network), corresponding to the measurement time of April 1, 2009, 23:00 LST. For this measurement time, the time axis corresponds to local solar time

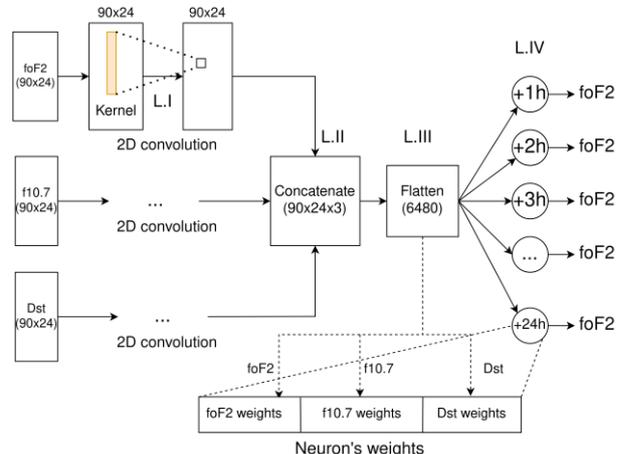


Figure 3. Neural network architecture

After training the neural network, we analyzed its coefficients to simplify the architecture of this rather general neural network. In order to simplify the architecture, we have addressed two main issues: 1) whether the network coefficients depend on the local time of the current moment or they can be considered stationary in the first approximation and thereby allow us to simplify the neural network; 2) what is the structure of the neural network coefficients and the contribution of the observed data for the previous days/hours to the predicted f_oF2 value?

The first step in training was to increase the accuracy of the prediction by the network not only by selecting the neural network coefficients, but also by applying several networks of identical architecture trained independently, using different datasets (ensemble methods). The simplest ensemble method that in some cases can improve the prediction accuracy is bagging [Breiman, 1994; Opitz, Maclin, 1999], which consists in independent training of an ensemble of networks and superposition of the results predicted by each network using a fixed algorithm, usually by averaging.

Models for subsequent bagging have been trained as follows.

As a validation dataset we utilized 20 % of the dataset at its end. Each of the 200 models was trained using 80 % of data randomly selected from the remaining part. This yielded 200 models, trained using different datasets, with various weight coefficients.

Since each network is linear in our case, the bagging (averaging the results of 200 neural networks) is equivalent to averaging of the neural network coefficients corrected for possible inversion of the result by an output layer, and within the framework of bagging allows us to expect an increase in the prediction accuracy. On the other hand, the physical meaning of the averaged neural network coefficients is the contribution of certain previous f_oF2 , Dst , $F10.7$ data to the predicted f_oF2 value. From a qualitative point of view, averaging these coefficients over an ensemble of networks will reduce their variations associated with learning inaccuracies or noise in the initial data, and thus will allow us to more accurately determine the contribution of previous data. To study the effect of the number of averages on the prediction accuracy, we have analyzed ensembles containing 1 (without bagging), 10, 30, 100, and 200 models.

To implement bagging, the output neuron coefficients are averaged taking into account possible inversion by the output layer according to the following algorithm:

- in a cycle for each network of the ensemble on the basis of the day closest to the moment of observation, we find the sign of the decision coefficient for each hour: -1 if the weight value is negative; 1 if it is positive;
- at -1 , the network coefficients corresponding to this hour are inverted (Figure 4, *a*);
- the final coefficient of the final matrix of coefficients is obtained by averaging this coefficient over the ensemble of trained networks.

The algorithm described above implements the following transformations (using the processing coeffi-

cients responsible for f_oF2 as an example) of the decision layer:

$$W_{i,j}^l = \frac{1}{N} \sum_{k=1}^N \left(INV \left(w_{i,j,k}^l, w_{i,0,k}^l \right) \right), \quad (2)$$

where k is the number of the network in the ensemble; i, j are the hours and days of delay from -23 to 0 and from 0 to 89 respectively; l is the number of the output neuron from $+1$ to $+24$, each responsible for the prediction of f_oF2 for the l th hour of 24 forecast hours; N is the number of trained models in the processed ensemble; $INV \left(w_{i,j,k}^l, w_{i,0,k}^l \right)$ is the function of conditional inversion depending on the sign of the multiplicative coefficient; $W_{i,j}^l$ is the matrix of output neuron coefficients, averaged within the ensemble of models, responsible for measuring f_oF2 ; $w_{i,j,k}^l$ is the initial matrix of output neuron coefficients responsible for f_oF2 for one (k -th) model. The weight inverting function is as follows:

$$INV \left(w, w_0 \right) = \begin{cases} w, & \text{if } w_0 \geq 0 \\ -w, & \text{if } w_0 < 0 \end{cases}, \quad (3)$$

where w is a weight coefficient; w_0 is a weight coefficient on the day closest to the moment of observation with the number of neuron, the number of the network of the ensemble, and the hour number of this coefficient are the same as in w . The "+" sign in front of the output neuron number is used to emphasize that the neuron outputs a predicted f_oF2 value.

In the initial neural network, each output neuron is responsible for the corresponding forecast hour after the moment of observation. Number output neurons from $+1$ to $+24$. The moment of observation stands for the moment of time starting from which we want to get a forecast for the next 24 hours. The relationship between the numbering of output neurons, the predicted hour and the moment of observation is shown in Figure 4, *b*.

Before obtaining the final ensemble averaged forecast described by Formulas (2), (3), we have trained the necessary number of models (see the architecture in Figure 3) and have analyzed the dependence of the forecast results on their number (1, 10, 30, 100, and 200 models).

Consider further calculations using the ensemble of 200 models as an example. We have independently trained 200 models. For each of the three input parameters of the initial neural network (f_oF2 , $F10.7$, Dst), we have obtained $24 \times 90 \times 24$ coefficient matrices corresponding to the forecast hours after the moment of observation. These matrices resulted from averaging of the corresponding matrices in each of the 200 models according to Formulas (2), (3). Call these matrices averaged neurons. Figure 5 shows the coefficients of the neuron responsible for the $+12$ th forecast hour relative to the moment of observation according to the data on f_oF2 measured at the -12 th (*a*) and -18 th (*b*) hours relative to the moment of observation. Along the horizontal axis is the delay before the moment of observation (in days); along the vertical axis is the weight coefficient. The scheme for obtaining these coefficients is given in Figure 6, *a*. From Figure 5 we can infer that f_oF2 for $+12$ th

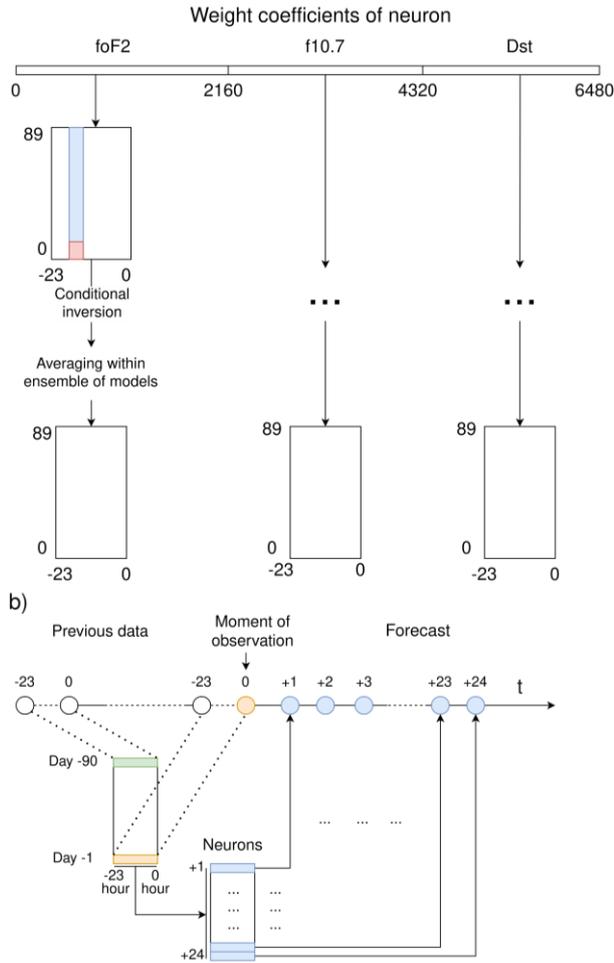


Figure 4. Scheme for averaging weight coefficients within an ensemble of trained models for one of the 24 output neurons (a) and a neural network model prediction scheme (b)

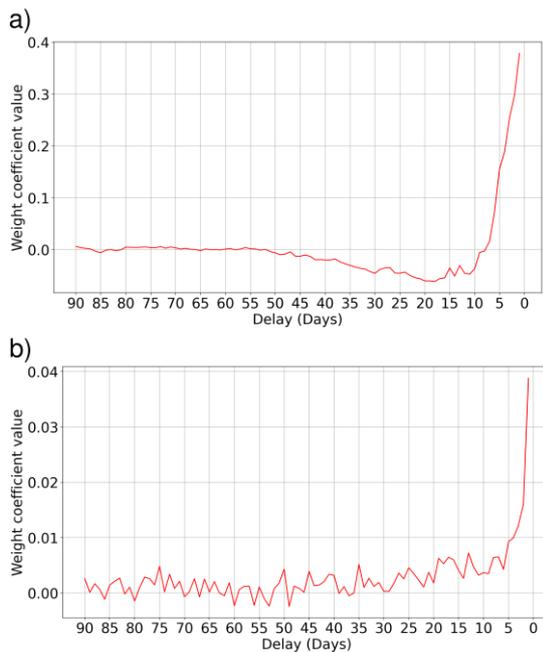


Figure 5. Coefficients of a neuron responsible for the +12th forecast hour after the moment of observation. The coefficients correspond to the -12th (a) and -18th (b) observation hours before the moment of observations during the preceding days

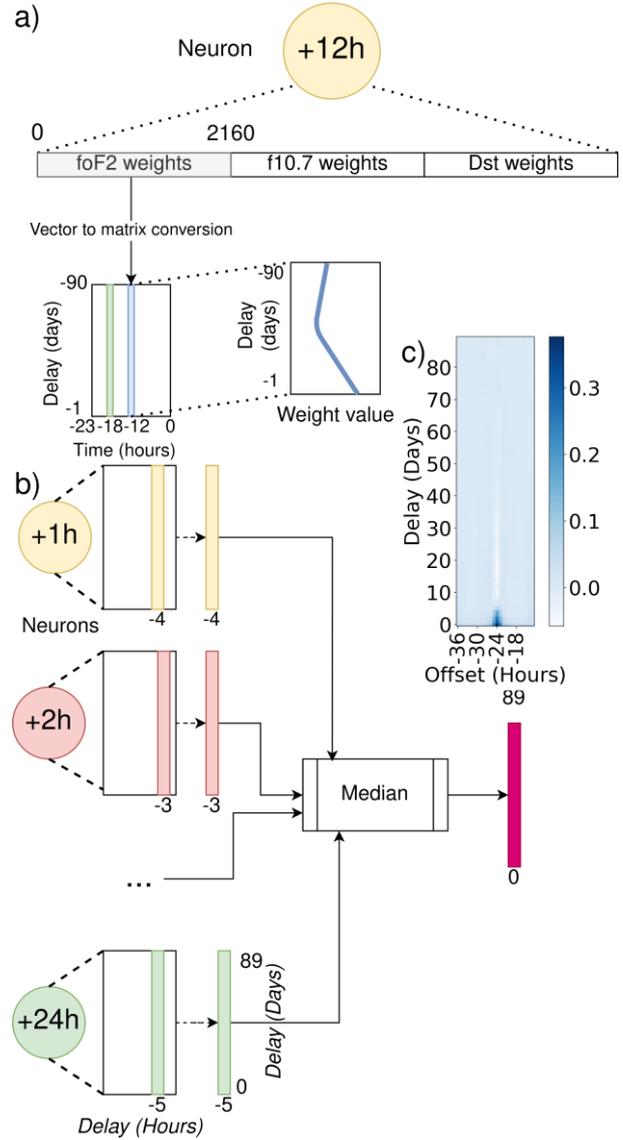


Figure 6. Identification of neuron coefficients corresponding to processing of f_oF2 observations at the -18th and -12th hours before the moment of observations, during the preceding days for the +12th forecast hour (a); scheme for obtaining response coefficients responsible for the influence of the previous hour, -29 hours from the forecast hour, -36 to -13 hours from the forecast hour (b); median matrix for the preceding hours, from -36 to -13 hours from the forecast hour (response matrix θ for f_oF2) (c)

forecast hour relative to the moment of observation is more strongly affected by the data, obtained on the preceding days, responsible for the -12th hour relative to the moment of observation, than the data, collected on the preceding days, responsible for the -18th hour relative to the moment of observation because the corresponding weight coefficients are on average higher. This means that the main contribution to the predicted f_oF2 value for the forecast hour is made by observations shifted from it by 24 hrs back on the days of previous observations.

When analyzing the coefficients of individual neurons responsible for the forecast hours, we observed that the main contribution to the predicted f_oF2 value is made by observations during the previous hour, -24 hours from

the forecast hour, as well as during the adjacent hours on the days of previous observations.

Analysis of all the 24 forecast hours showed a similar picture.

Thus, the values in the matrix of coefficients shift depending on the neuron number (forecast hour) and have a maximum at the corresponding hour. Thus, when switching to the time reference system associated with the forecast hour, these functions will not depend on it. It is therefore possible for the analysis to transform all the matrix coefficients to a time reference system related not to the local time of the observation moment, but to the forecast hour (shifted matrices). For the reference level of the shift, we take the -24 hour shift from the forecast hour. Accordingly, assume that the interval of adjacent hours is from -36 to -13 from the forecast hour.

The independence of the shape of these shifted matrices from the forecast hour we have demonstrated allows us to average the matrices over an ensemble composed of 24 shifted matrices, each corresponding to its own forecast hour.

Figure 6, *b* illustrates the formation of median coefficients in columns for the -29 hour shift from the forecast hour, each cell of a new matrix is a median value of corresponding cells of 24 initial shifted matrices.

To confirm the weak dependence of the shifted matrix coefficients on the forecast hour, we demonstrate what the coefficients used to calculate the median look like. Figure 7 displays weight coefficients of the shifted matrices (for different forecast hours) for f_oF2 , corresponding to -29 hours (*a*) and -24 hours (*b*) from the forecast hour during previous days. Colors of the curves correspond to different forecast hours.

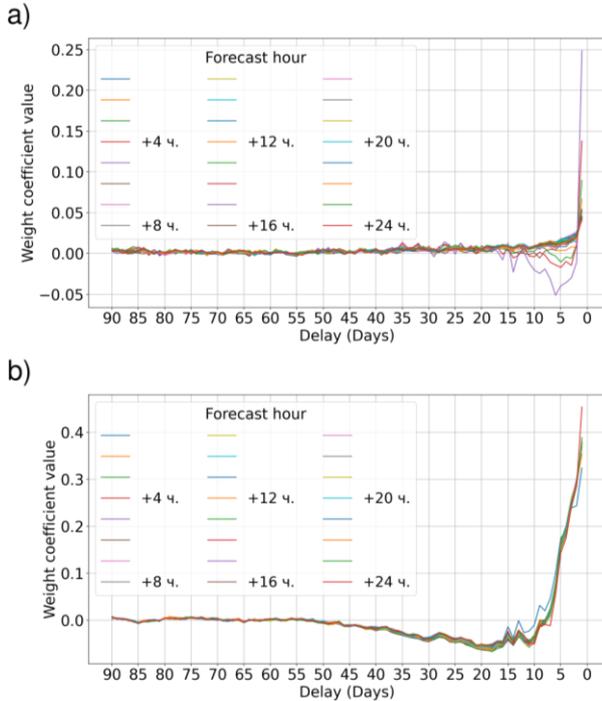


Figure 7. Weight coefficients of shifted matrices of coefficients corresponding to -29 hours from the forecast hour, depending on the number of the previous day (X axis) and the forecast hour (color) (*a*); -24 hours from the forecast hour, depending on the number of the previous day (X axis) and the forecast hour (color) (*b*)

The coefficients of the shifted matrices in the first approximation are seen to be independent of the forecast hour; they depend only on the relative delay between the forecast hour and the previous data involved in calculating the predicted value. We can therefore statistically process the shifted matrices, using an ensemble made up of shifted matrices for each of the 24 hours of the forecast.

Stationarity of the shifted matrix allows us to redefine its coefficients through additional statistical processing, thereby simplifying the network architecture and reducing the number of free coefficients. We therefore analyze only the median shifted matrix of coefficients we have obtained.

Thus, the predicted f_oF2 value in the first approximation is a superposition of the f_oF2 values at previous moments of time, and their contribution to the predicted value does not depend on local time, but only on the relative delay between the data used for the forecast and the moment at which we make this forecast.

These coefficients are described by the 3D $90 \times 24 \times 3$ shifted matrix in coordinates (delay in days, time shift, parameter), where the third coordinate is responsible for one of the measurements of f_oF2 , $F10.7$, Dst . This shifted matrix will be further referred to as the response matrix.

Let us take a closer look at the calculation of this matrix, using the f_oF2 measurement data processing as an example:

$$\theta_{k,j} = \text{median}_l \left(W_{SHIFT(l,k),j}^l \right), \quad (4)$$

where θ is the 90×24 response matrix for f_oF2 ; W is a matrix of neural network ensemble averaged coefficients obtained by calculation from Formula (2); k is relative shifts of the previous hour from the forecast hour from -36 to -13 ; l is a forecast hour from $+1$ to $+24$; j is the delay day number from 0 to 89; $SHIFT(l,k)$ is the function for calculating the shifted hour number depending on the shift and the forecast hour, which converts relative shifts into hour delays relative to zero, which must be obtained from the matrix W ; median is the function that returns a median column vector of size 90 within the processed neurons. For example, Figure 7, *a* presents 24 curves corresponding to different forecast hours, from which the median curve is calculated for -29 hours from the forecast hour.

Since the shift coefficients can go beyond the matrix W , during calculations it is transformed by the $SHIFT(l,k)$ function that corrects this problem:

$$SHIFT(l,k) = \begin{cases} l+k, & \text{if } l+k \in [-23,0] \\ (l+k)+24, & \text{if } l+k < -23. \\ (l+k)-24, & \text{if } l+k > 0 \end{cases} \quad (5)$$

CALCULATION OF MODEL COEFFICIENTS FOR INPUT $F107$ AND Dst INDICES

We have considered above only the influence of measured f_oF2 values on its predicted values. Using this model, we calculated the contribution of $F10.7$ and Dst to the predicted f_oF2 value. The response matrices for $F10.7$

and Dst are shown in Figure 8.

Since the input measured data is standardized according to Formula (1), we can estimate the relative contribution of f_oF2 , $F10.7$, and Dst to the predicted f_oF2 value by comparing the absolute values of the coefficients of response matrices with each other [Bring, 1994]. The comparison shows that the influence of $F10.7$ and Dst on the predicted f_oF2 value is on average inconsiderable compared to the f_oF2 data, which can be explained by the architecture of our fairly simple network.

Panels $a-f$ indicate that the result of the f_oF2 forecast should be most strongly affected by $F10.7$ and Dst observed on the day preceding the forecast day. Panels e, f

present the weight coefficients obtained as in Figure 7, which demonstrate that the response matrices are stationary.

Panels g, h show response matrices for $F10.7$ and Dst when forecasting f_oF2 , which were calculated using the same algorithm as for the response to f_oF2 . From the analysis of Figure 8 we can conclude that the forecast of f_oF2 is more strongly influenced by $F10.7$ and Dst measured on the day closest to the moment of observation. The influence of $F10.7$ and Dst on the f_oF2 forecast on the day closest to the forecast day should be taken into account more carefully; it is not examined here.

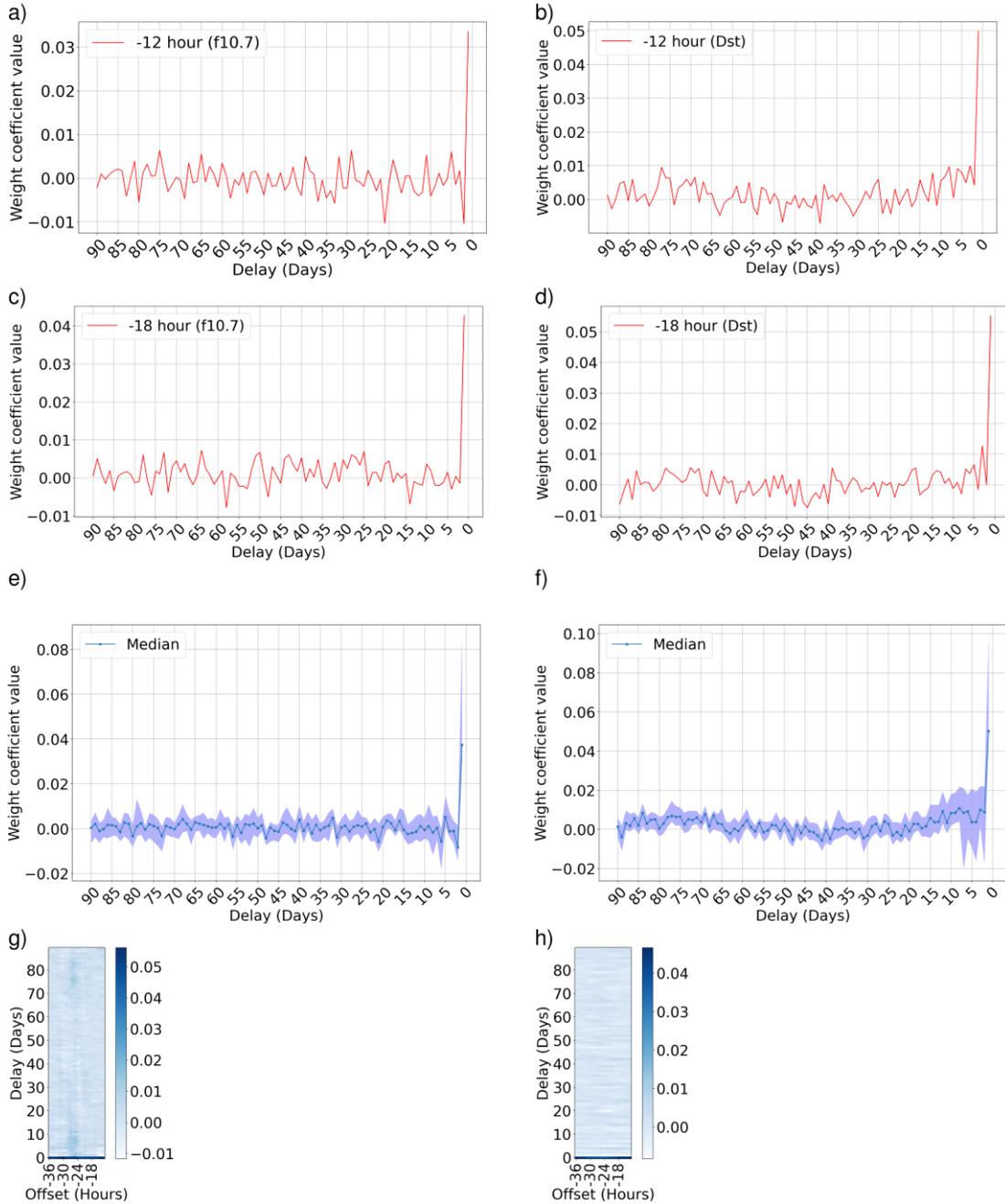


Figure 8. Weight coefficients of a neuron responsible for the +12th forecast hour, which correspond to -12th (a, b) and -18th (c, d) hour before the moment of observation for $F10.7$ (a, c) and Dst (b, d). The median weight coefficients of shifted matrices of coefficients corresponding to -24 hours from the forecast hour on the previous days for $F10.7$ (e) and Dst (f). Response matrices for $F10.7$ (g) and Dst (h) for the f_oF2 forecast

FINAL f_oF2 FORECAST MODEL

To adopt the neural network model, we retrain it, taking into account characteristics of the response matrices — their relative independence of the moment of observation demonstrated above. We can therefore fix the output layer coefficients and the response matrices obtained by the corresponding ensembles (1, 10, 30, 100, 200 independently trained models), as well as retrain the network to refine the shape of the convolution kernel.

For the final architecture of the model (see Figure 3), the output layer coefficients have been obtained at the previous stage, they do not change during the training process. Note that in the output layer of the final model, the response coefficient matrices are shifted for each neuron so that a maximum response corresponds to a

forecast hour. The shift was made according to the formula

$$W_{i,j}^{l,k} = \theta_{SHIFTOUT(l,i),j}^k, \quad (6)$$

where θ^k is the 90×24 response matrix of the k th criterion; k is the criterion number, one criterion of f_oF2 , $F10.7$, Dst ; W is the matrix of output layer coefficients; i is the index of the weight matrix column from 0 to 23; l is a forecast hour from +1 to +24; j is a delay day index from 0 to 89; $SHIFTOUT(l, i)$ is the function for calculating the shifted hour number depending on the response matrix index responsible for the relative shifts of the preceding hour from the forecast hour, and forecast hour (7). The range of output $SHIFTOUT(l, i)$ values is from -36 to -13 :

Table 1

Quality metrics of the IRI model and neural network models with different methods for obtaining coefficients of the model

Number of independent trainings for averaging coefficients	Pearson	RMSE, MHz	MAPE, %	R^2
1	0.93	0.595	10.502	0.863
10	0.929	0.596	10.486	0.862
30	0.929	0.596	10.468	0.862
100	0.93	0.596	10.532	0.862
200	0.928	0.598	10.445	0.861
IRI Model	0.876	0.823	13.803	0.738

$$SHIFTOUT(l, i) = \begin{cases} i-l-23, & \text{if } i-l-23 \in [-36, -13] \\ (i-l-47), & \text{if } i-l-23 > -13 \\ i-l+1, & \text{if } i-l-23 < -36 \end{cases}. \quad (7)$$

Thus, in the last layer of the final neural network, output neurons differ in coefficients since the response matrix for each neuron is shifted in a special way according to Formula (6).

For each forecast hour, the initial data is transformed following the method described in Section “Initial Data for Network Training”. These are the f_oF2 , $F10.7$, and Dst matrices.

When training the final model, response matrices were selected from the corresponding ensemble of coefficient matrices of the initial models (1, 10, 30, 100, 200 models) and shifted according to Formula (6) to generate matrices of output layer coefficients of the final neural network.

When training the network (finding convolution coefficients), the initial dataset was divided as follows: 60 % — training dataset, 20 % — validation dataset, 20 % — test dataset.

Four metrics have been used to assess performance of the model: Pearson correlation coefficient, root-mean-square error (RMSE), mean absolute percentage error (MAPE), and coefficient of determination (R^2). We compared the real data from the test dataset, the data predicted by the final model, and the data predicted by the IRI model. Figure 9, *a* presents the results of the forecast of the diurnal f_oF2 variation. Table 1 lists quality metrics of

the final model when generating response matrices, using ensembles of 1, 10, 30, 100, and 200 independent models, as well as compares them with the IRI model’s quality metrics. We have used the IRI-2016 model with URSI.

The data allows us to conclude that there is practically no improvement in the quality of the model with an increase in the number of independent trainings of response matrix coefficients. Thus, stationarity of the response matrix (its independence of the moment of observation) is the main factor affecting the quality metrics of this model.

For comparison, Salimov, Khmel'nov [2020] give quality metrics of the f_oF2 forecast +24 hours for the Recurrent Neural LSTM Network, using Irkutsk data: Pearson correlation coefficient — 0.923, RMSE — 0.605 MHz, MAPE — 10.07 %. These quality metrics are similar to those of our model, which allows for the conclusion that it is important to take into account previous ionospheric observations when developing predictive models.

To analyze the behavior of the model from the data with gaps, we have conducted an experiment: specially added gaps to the test dataset so that their distribution was indistinguishable from the distribution of gaps in the training dataset (the resulting value of p -value is 0.11 according to the Mann-Whitney test when comparing the distributions of gaps of modified test and training datasets). The quality metrics of the model based on averaged coefficients from 200 training sessions for such a test dataset are as follows: Pearson correlation coefficient — 0.923; RMSE — 0.598 MHz; MAPE — 10.45 %; R^2 — 0.86. These quality metrics practically coincide with those of the model in the unchanged test dataset (see the second last row of Table 1), which al-

allows us to conclude that the model is resistant to omissions in data, which are most characteristic of current measurements with the Irkutsk digisonde.

To analyze the performance of the model, depending on the forecast hour, we have calculated the corresponding metrics separately for each forecast hour. They are shown in Figure 10. It can be inferred that the most qualitative forecast is given for the 1st hour, then, by the 24th hour, the quality metrics become slightly worse.

Figure 9 *c, d* plots the f_oF2 forecast during geomagnetic storms (four days after and a day before a decrease in the geomagnetic index K_p below 6). It can be seen

that the model reacts worse to a peak of the geomagnetic storm and adapts to quiet conditions with some delay. Thus, the proposed model is more likely to describe an undisturbed ionosphere and is not recommended for use in highly disturbed conditions due to error growth.

Figure 11 presents results of the model in different seasons. It can be seen that in the first approximation the model traces the diurnal variation satisfactorily, regardless of the season.

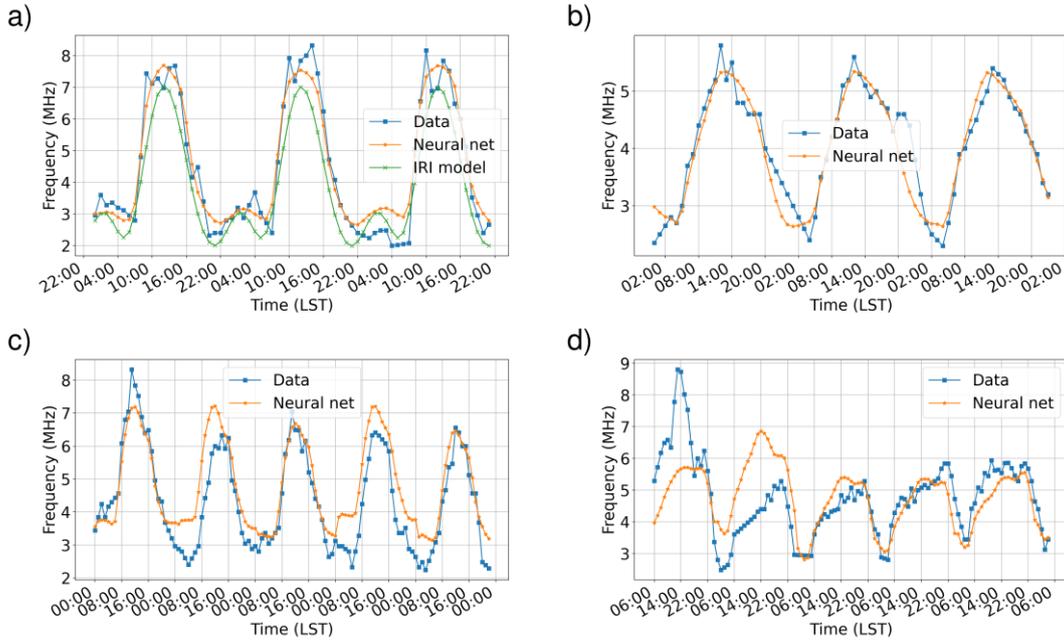


Figure 9. Forecast of f_oF2 with the neural network and the IRI model and the diurnal f_oF2 variation from December 14, 2015 00:00 LST to December 16, 2015 22:00 LST (a), from October 25, 2016 00:00 LST to October 29, 2016 23:00 LST (c), from May 8, 2016 06:00 LST to May 13, 2016 05:00 LST (d) for Irkutsk for the final model with coefficients calculated from the ensemble of 200 independent training sessions; f_oF2 forecast, its diurnal variation for Arti ($56^{\circ}25' N$, $58^{\circ}32' E$) from April 13, 2017 00:00 LST to April 16, 2017 00:00 LST (b)

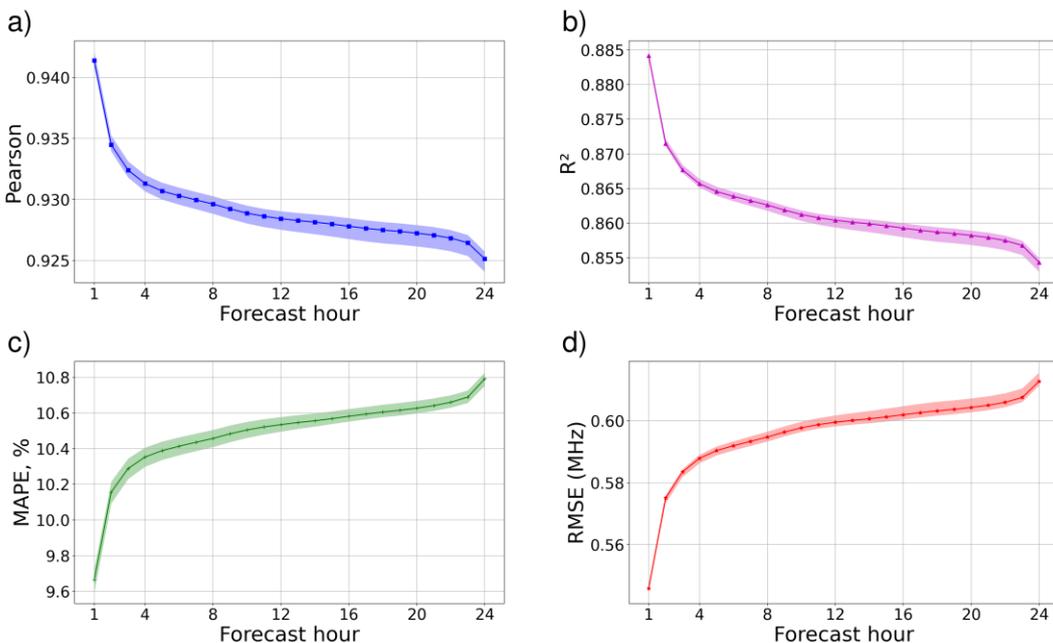


Figure 10. Quality metrics of models as function of forecast hour. Lines are arithmetic means of the models, color is the spread of values for different final models based on responses calculated using 1, 10, 30, 100, 200 initial models

MODEL TEST WITH DATA FROM OTHER MID-LATITUDE IONOSONDES

When developing forecast models, their generalizing ability is important, which can be tested by quality matrices, using different test datasets that do not overlap with the training dataset. In this work, as such test datasets in addition to the Irkutsk test dataset, we have selected data from three mid-latitude ionosondes close in latitude to Irkutsk, but significantly spaced in longitude: Arti (Russia, 56°25' N, 58°32' E), Warsaw (Poland, 52°13' N, 21°02' E), Mohe (China, 52°58' N, 122°31' E) for 2017.

The test was carried out as follows.

- Ionosonde data was reduced to the desired format (a set of 90×24 matrices and vectors of size 24).
- The data acquired using the standardization coefficients μ and s , obtained through standardization of Irkutsk training datasets, was scaled according to Formula (1). These coefficients are applied equally to data from all radars and are listed in Table 2. For the initial and final models, the data will differ due to different methods of separating the datasets.

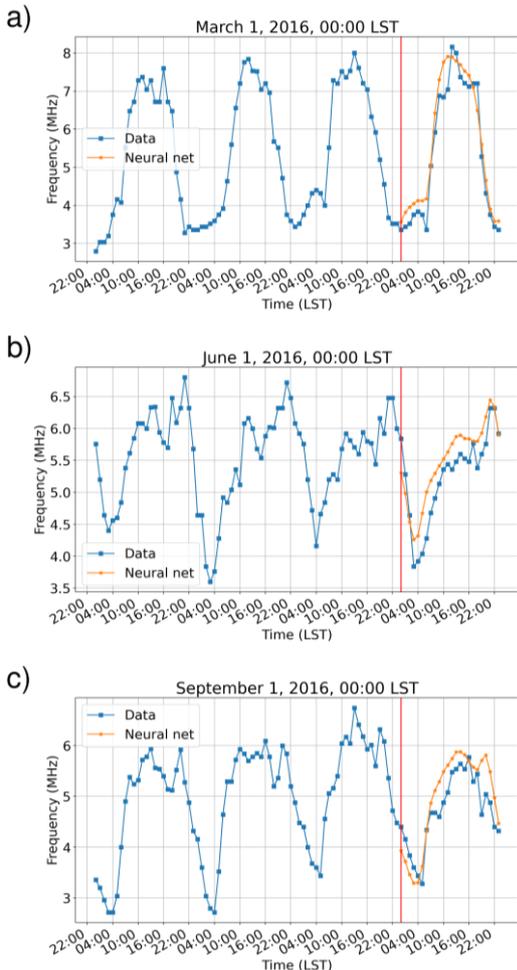


Figure 11. Forecast of the critical frequency f_oF2 over Irkutsk after measurements on March 1, 2016, 00:00 LST (a); June 1, 2016, 00:00 LST (b); September 1, 2016, 00:00 LST (c). The measurement points coincide with the beginning of the vertical line

Table 2

Coefficients of standardization of Irkutsk training datasets

Criterion		f_oF2 , MHz	$F10.7$	Dst
Initial model (Model A)	μ	5.403	110.795	-9.814
	s	1.989	31.779	15.051
Final	μ	5.068	100.864	-9.458
	s	1.672	26.995	14.628

- This data was used for forecasting with the neural network model we propose.

- For comparison, we forecasted f_oF2 , using a model constructed and trained without assuming that the response matrix is stationary (Model A).

- The predicted f_oF2 value was compared with the value measured at this point.

The forecast quality matrices obtained from testing the mid-latitude ionosonde data are given in Table 3. The forecast of f_oF2 and its diurnal variation for Arti are plotted in Figure 9, b.

It follows from Table 3 that the quality metrics of the final model weakly depend on the number of models in an ensemble, but these metrics are better than those of the initial Model A. Thus, we can conclude that taking into account the stationarity of response matrices improves the quality of the forecast. The result also suggests a good generalizing ability of the constructed models, whose absolute error is ~ 0.6 MHz, and weakly depends on the longitude of station. This allows us to conclude that the model we have trained is suitable for the prediction of f_oF2 according to the data from mid-latitude ionosondes.

CONCLUSIONS

We have studied the influence of 90 days preceding the moment of observation on the forecast of the critical frequency f_oF2 . The f_oF2 forecast model one day in advance was obtained based on convolution networks from f_oF2 , $F10.7$, and Dst data. Coefficients of the model have been derived by averaging the coefficients from 200 independent trainings of the initial model, followed by averaging the network's coefficients, assuming that response matrices are stationary, as well as through additional training of final models to redefine the coefficients of the input convolution layer.

We have presented a method for averaging coefficients of the model over the ensemble of independently trained models and the ensemble of 24 shifted matrices, taking into account their stationarity, which is a kind of analogue of the ensemble method (bagging) for this network architecture.

We have demonstrated that the contribution of the parameters to the predicted f_oF2 value in the first approximation does not depend on local time (with a forecast for 24 hours or less), but only on the time interval between the corresponding measured value and the moment for which the forecast is made. This indicates the stationarity of the response matrices. We have shown that the main contribution to the predicted f_oF2

Table 3

Quality metrics of models according to data from different mid-latitude ionosondes. Model A is constructed and trained without assuming that the response matrix is stationary (initial model). Quality metrics of Model A are averaged over a set of 200 independent model trainings

	Model	Pearson	RMSE, MHz	MAPE, %	R^2
Arti	Model A	0.848	0.669	14.242	0.683
	1 (final model)	0.874	0.613	13.506	0.735
	10 (final model)	0.872	0.596	12.818	0.750
	30 (final model)	0.870	0.609	13.194	0.739
	100 (final model)	0.867	0.629	13.966	0.721
	200 (final model)	0.875	0.591	12.584	0.754
Mohe	Model A	0.821	0.632	12.622	0.624
	1 (final model)	0.847	0.556	11.080	0.710
	10 (final model)	0.844	0.554	10.763	0.712
	30 (final model)	0.842	0.560	11.042	0.705
	100 (final model)	0.842	0.567	11.353	0.699
	200 (final model)	0.845	0.554	10.687	0.713
Warsaw	Model A	0.809	0.708	14.279	0.613
	1 (final model)	0.841	0.626	12.398	0.698
	10 (final model)	0.838	0.622	12.072	0.702
	30 (final model)	0.841	0.621	12.192	0.703
	100 (final model)	0.837	0.634	12.599	0.691
	200 (final model)	0.839	0.622	12.015	0.703

value is made by the data obtained during the next few days before the forecast, the contribution of the remaining days decreases significantly, which is understandable from a physical point of view.

We have demonstrated that taking into account the response matrix stationarity makes it possible to improve the quality metrics of the model forecast compared to the model that omits averaging of weight coefficients (see Table 3).

It has been shown that with an increase in the number of independent models used to average coefficients from 1 to 200, the quality of the forecast practically does not improve.

The final model provides the following prediction quality metrics: Pearson correlation coefficient — 0.928; root-mean-square error (RMSE) — 0.598 MHz; mean absolute percentage error (MAPE) — 10.45 %; coefficient of determination — 0.861. Compared to the IRI model, the model we propose yields the best quality metrics (MAPE is 13.803 % for IRI and 10.45 % for our model). As for the acceptable accuracy of the model, the requirement for RMSE depends on the specific problem being solved. Obviously, the smaller it is, the better. For qualitative problems of estimating radio wave propagation, RMSE ~0.7 MHz can probably be considered sufficiently accurate, which makes it possible to separate the magnetoionic components in the mid-latitude ionosphere.

Using data from ionosondes significantly spaced in longitude (Arti, Warsaw, Mohe), we have shown that the constructed model can be employed to predict the ionosphere according to data from other mid-latitude ionosondes without changing coefficients of this model, which indicates a good generalizing ability of the model.

The work was financially supported by the Ministry of Science and Higher Education of the Russian Federation. The results were obtained using the equipment of Shared Equipment Center «Angara» [<https://ckp-rf.ru/ckp/3056>].

REFERENCES

- Abadi M., Barham P., Chen J., Chen Z., Davis A., Dean J., Devin M., Ghemawat S. TensorFlow: A system for large-scale machine learning. *Proc. OSDI*. 2016, pp. 265–283. DOI: [10.5281/zenodo.4724125](https://doi.org/10.5281/zenodo.4724125).
- Barkhatov N.A., Revunov S.E., Urjadov V.P. Artificial neural network technology for forecasting critical frequency of ionospheric layer F2. *Proceedings of Higher Educational Institutions. Radiophysics*. 2005, vol. 48, iss. 1, pp. 1–15. (In Russian).
- Bilitza D., Mckinnell L.-A., Reinisch B., Fuller-Rowell T. The International Reference Ionosphere (IRI) today and in the future. *J. Geodesy*. 2011, vol. 85. DOI: [10.1007/s00190-010-0427-x](https://doi.org/10.1007/s00190-010-0427-x).
- Boulch A., Cherrier N., Castaings T. *Ionospheric activity prediction using convolutional recurrent neural networks*. 2018. DOI: [10.48550/arXiv.1810.13273](https://doi.org/10.48550/arXiv.1810.13273).
- Breiman L. Bagging Predictors. *Technical Report*. 1994, No. 421.

- Bring J. How to standardize regression coefficients. *The American Statistician*. 1994, vol. 48, no. 3, pp. 209–213. DOI: [10.2307/2684719](https://doi.org/10.2307/2684719).
- Consultative Committee on International Radio (CCIR) *Atlas of Ionospheric Characteristics Report 340*. International Telecommunication Union, Geneva, Switzerland, 1967.
- Galkin I.A., Reinisch B., Vesnin A.M., Huang X. Assimilation of sparse continuous groundbased ionosonde data into IRI using NECTAR model morphing. *The 1st URSI Atlantic Radio Science Conference (URSI AT-RASC)*. Las Palmas, 2015, pp. 1–8. DOI: [10.1109/URSI-AT-RASC.2015.7303112](https://doi.org/10.1109/URSI-AT-RASC.2015.7303112).
- Goodfellow I., Bengio Y., Courville A. *Deep Learning*. MIT Press, 2016, 800 p.
- Hargrivi J.K. Upper atmosphere and solar-terrestrial connections. *Introduction to Physics of the Near-Earth Space Environment*. Leningrad, Gidrometeoizdat, 1982, 351 p. (In Russian).
- Kingma D.P., Ba J.A. A Method for Stochastic Optimization. *International Conference on Learning Representations*. 2014. DOI: [10.48550/arXiv.1412.6980](https://doi.org/10.48550/arXiv.1412.6980).
- Lundberg S., Lee, S.-I. A Unified Approach to Interpreting Model Predictions. arXiv. 2017. DOI: [10.48550/ARXIV.1705.07874](https://doi.org/10.48550/ARXIV.1705.07874).
- Opitz D., Maclin R. Popular ensemble methods: An empirical study. *J. Artificial Intelligence Res.* 1999, vol. 11, pp. 169–198. DOI: [10.1613/jair.614](https://doi.org/10.1613/jair.614).
- Ratovsky K.G., Potekhin A.P., Medvedev A.V., Kurkin V.I. Modern Digital Ionosonde DPS-4 and its capabilities. *Solar-Terr. Phys.* 2004, iss. 5, pp. 102–104. (In Russian).
- Rush C., Fox M., Bilitza D., Davies K., McNamara L., Stewart F., PoKempner M. Ionospheric mapping — an update of f_oF2 coefficients. *Telecommun. J.* 1989, vol. 56, iss. 3, pp. 179–182.
- Salimov B.G., Khmel'nov A.E. Prediction of the critical frequency of the ionosphere f_oF2 using a neural recurrent LSTM network. *Proc. Conference "Lyapunovskie Chteniya". ISDC SB RAS*, Irkutsk, 2020, pp. 60–61. (In Russian).
- Sivavaraprasad G., Lakshmi Mallika I., Sivakrishna K., Venkata Ratnam D. A novel hybrid machine learning model to forecast ionospheric TEC over low-latitude GNSS stations. *Adv. Space Res.* 2022, vol. 69, iss. 3, pp. 1366–1379. DOI: [10.1016/j.asr.2021.11.033](https://doi.org/10.1016/j.asr.2021.11.033).
- Smirnov V.F., Stepanov A.E. New capabilities in studies of the high-latitude ionosphere: DPS-4 digisonde first results on measurements of localization and dynamics of large-scale ionospheric structures in Yakutsk. *Solar-Terr. Phys.* 2004, no. 5 (118), pp. 105–106. (In Russian).
- Yu S., Ma J. Deep learning for geophysics: Current and future trends. *Rev. Geophys.* 2021, vol. 59, iss. 3, e2021RG000742. DOI: [10.1029/2021RG000742](https://doi.org/10.1029/2021RG000742).
URL: <http://irimodel.org/IRI-2016> (accessed June 27, 2019).
URL: <https://ckp-rf.ru/catalog/ckp/3056> (accessed October 3, 2018).
URL: <https://omniweb.gsfc.nasa.gov/form/dx1.html> (accessed August 31, 2020).
- Original Russian version: Salimov B.G., Bergardt O.I., Khmel'nov A.E., Ratovsky K.G., Kusonsky O.A., published in *Solnechno-zemnyaya fizika*. 2023. Vol. 9. Iss. 1. P. 61–73. DOI: [10.12737/szf-91202307](https://doi.org/10.12737/szf-91202307). © 2023 INFRA-M Academic Publishing House (Nauchno-Izdatelskii Tsentr INFRA-M)
- How to cite this article*
Salimov B.G., Bergardt O.I., Khmel'nov A.E., Ratovsky K.G., Kusonsky O.A. Application of convolution neural networks for critical frequency f_oF2 prediction. *Solar-Terrestrial Physics*. 2023. Vol. 9, Iss. 1. P. 56–67. DOI: [10.12737/stp-91202307](https://doi.org/10.12737/stp-91202307).