Article

Approximation and Analysis of Natural Data Based on NARX Neural Networks Involving Wavelet Filtering

Oksana Mandrikova [1,*], Yuryi Polozov [1], Nataly Zhukova [2] and Yulia Shichkina [3]

1 Institute of Cosmophysical Research and Radio Wave Propagation, Far Eastern Branch of the Russian Academy of Sciences, Mirnaya St., 7, Kamchatskiy Kray, 684034 Paratunka, Russia

2 Laboratory of Big Data Technologies in Socio-Cyberphysical Systems, Saint-Petersburg Federal Research Centre of the Russian Academy of Sciences, 199178 St. Petersburg, Russia

3 Department of Technologies of Artificial Intelligence in Physiology and Medicine, Alexander Popov International Innovation Institute for Artificial Intelligence, Cybersecurity and Communications, St. Petersburg State Electrotechnical University "LETI", 197022 St. Petersburg, Russia

* Correspondence: oksanam1@mail.ru

Abstract: Recurrent neural network (RNN) models continue the theory of the autoregression integrated moving average (ARIMA) model class. In this paper, we consider the architecture of the RNN with embedded memory—«Process of Nonlinear Autoregressive Exogenous Model» (NARX). Though it is known that NN is a universal approximator, certain difficulties and restrictions in different NN applications are still topical and call for new approaches and methods. In particular, it is difficult for an NN to model noisy and significantly nonstationary time series. The paper suggests optimizing the modeling process for a complicated-structure time series by NARX networks involving wavelet filtering. The developed procedure of wavelet filtering includes the application of the construction of wavelet packets and stochastic thresholds. A method to estimate the thresholds to obtain a solution with a defined confidence level is also developed. We introduce the algorithm of wavelet filtering. It is shown that the proposed wavelet filtering makes it possible to obtain a more accurate NARX model and improves the efficiency of the forecasting process for a natural time series of a complicated structure. Compared to ARIMA, the suggested method allows us to obtain a more adequate model of a nonstationary time series of complex nonlinear structure. The advantage of the method, compared to RNN, is the higher quality of data approximation for smaller computation efforts at the stages of network training and functioning that provides the solution to the problem of long-term dependencies. Moreover, we develop a scheme of approach realization for the task of data modeling based on NARX and anomaly detection. The necessity of anomaly detection arises in different application areas. Anomaly detection is of particular relevance in the problems of geophysical monitoring and requires method accuracy and efficiency. The effectiveness of the suggested method is illustrated in the example of processing of ionospheric parameter time series. We also present the results for the problem of ionospheric anomaly detection. The approach can be applied in space weather forecasting to predict ionospheric parameters and to detect ionospheric anomalies.

Keywords: time series model; wavelet transform; neural network NARX; ionospheric parameters

MSC: 62C12; 62C20; 62L20; 68T05; 68T07

1. Introduction

Time series modeling and analysis form an important fundamental basis for the investigation of processes and phenomena of different nature. This theory can be applied in different spheres of human activities (physics, biology, medicine, economy, etc.). A separate class of problems of time series analysis is directed on the diagnostics of object states and anomaly detection. Such problems have special relevance in the area of geophysical monitoring, they are: anomaly detection in geological medium [1]; in the near-earth

space [2–4]; the prediction of tsunamis [5,6], earthquakes [7], and other catastrophic natural phenomena. Anomaly detection and identification are also very topical in medicine [8]. The most important requirements for such methods are accuracy, promptness of answer reception, as well as the adaptability to have the possibility to record fast nonstationary changes of the system or object state.

Natural data time series have a complex structure that complicates the process of construction of analysis models and methods. Classical methods of time series analysis (AR, ARMA models [9,10], stochastic approximation [11,12], etc.) do not allow us to describe data complex structures adequately, and do not satisfy the adaptation requirement [3,13]. In applications, hybrid approaches are increasingly frequently used. They are based on the combination of deterministic and stochastic methods involving elements of machine learning [3,13–19]. They allow us to improve the procedure of complex data analysis. For example, in the paper [15], simplified selection ensembles based on trees were used to model and predict the data on milk yield. Preliminary data processing included their reduction into rotating main components. The authors [15] developed a simplified selective algorithm on the index of concordance that allowed them to optimize the method's performance. In the paper [15], two linear hybrid models were constructed and investigated. Another hybrid approach is considered in the paper [13] to analyze hydrological data. The authors of the paper [13] showed the efficiency of the joint application of wavelet transform [20–22] with neural networks (NN): wavelet neural network models. A flexible apparatus of the wavelet transform made it possible to apply it successfully in data analysis application [2]. A set of wavelet decomposition schemes and a wide library of basic wavelets allow us to adapt this method for the data of different-structure and according to the investigation's aim. The authors of the papers [23,24] illustrated the efficiency of application of wavelet transform with ARIMA models to model ionospheric parameter time variations and to detect anomalies. This paper continues that work. Here, we use the combination of wavelet packets with recurrent neural networks (RNN) [25], which continue the theory of the autoregression integrated moving average (ARIMA) model class [9,10].

Neural network methods are widely applied now in different areas of experience [3,4,7,13,26,27]. However, we should note that NN efficiency depends on the properties of training data and their representativity. Though the NN apparatus includes a wide set of paradigms and allows us to approximate complicated dependencies, some difficulties and restrictions in different applications are still topical and require the development of new approaches and methods [28–31]. In particular, NN efficiency decreases significantly for very noisy and nonstationary data. For example, it is difficult for an NN to model the nonstationary not associated with seasonal regularities, especially when it has long time delays [28]. Thus, the application of an NN requires data pre-processing (suppression of noise, elimination of trends, seasonality, etc.) in most cases to obtain an optimal result [28,32–35]. In this case, the combination with different methods makes it possible to overcome the problems in NN applications. For example, it was suggested in the paper [17], to apply the LSTM neural network together with discrete wavelet decomposition and ARIMA models. A combination of discrete wavelet decomposition with neural network and ARIMA was also suggested in the paper [36] to forecast the hydrological time series.

This paper considers RNN network architecture with embedded memory—«Process of Nonlinear Autoregressive Exogenous Model» (NARX) [25,27,28,30,33]. The evident advantages of regression models are their mathematical validity, formalized method of model identification, and the test for adequacy. Moreover, the advantage of the NARX network with the training gradient algorithm is their rapid convergence and good capacity for generalization [29,30].

However, we should note that one of RNN's problems is the problem of long-term dependencies [37]. Many researchers tried to solve it. The authors of [38] showed that in certain cases, RNN are capable of reflecting time delays of not less than 100 time steps. A complex approach to the solution of the problem of long-term dependencies was proposed

by [39]. The authors [39] applied the architecture of the segmented-memory recurrent neural network (SMRNN) together with extended real-time recurrent learning (eRTRL). However, the eRTRL had a high computational complexity; thus, the authors [39] introduced an auxiliary condition in the form of extended back propagation through time (eBPTT) for SMRNN together with a layer-local unsupervised pre-training procedure. The theoretical solution of the question of vanishing gradient is represented in the paper [40]. The authors [40] used a regularization term, which prevents the error signal from vanishing during its motion back in time. It was also shown in the paper [40] that the proposed solutions improved RNN performance on the considered synthetic data sets. The authors [41] suggested a Fourier Recurrent Unit (FRU), which stabilizes gradients arising during network training. The FRU summarizes hidden states in the temporal dimension by Fourier basic functions. The experimental part of the paper [41] showed for the smaller number of parameters, the suggested architecture exceeded other RNN for many problems. One more approach to the solution of the problem of vanishing gradients and the problem of long-term dependences associated with it, was suggested in [42]. The authors [42] introduced a new recurrent unit with a residual error on the stage of training (Res-RNN network). It was shown in the paper [42] that the proposed Res-RNN was effective in standard RNN modifications.

One of the effective solutions of the problem of long-term dependencies for RNN is the long short-term memory (LSTM) architecture [43,44]. Multiple investigations caused the development of the architecture and its application in different fields. For example, in the paper [45], the authors showed the LSTM application for the graph of pump operation in drinking water production. The system, obtained by the authors [45], made it possible to take into account such information as day of a week, and holidays when solving the problem of long-term dependencies. The development of LSTM architecture was presented by the authors [46], who used the model of improved seasonal-trend decomposition LSTM (ISTL-LSTM) to forecast bus passenger traffic during COVID-19. The model [46], based on STL, several functions, and three neural LSTM networks made it possible to forecast the daily bus passenger traffic in Beijing during the pandemic. As the authors [47] showed, the application of LSTM is possible in hybrid format. In the paper [47], the combination of two types of NN was applied. That was determined by a different dimension of input data. The convolution neural network (CNN) was used to process two-dimensional precipitation maps and LSTM to process one-dimensional output CNN data and to calculate the downstream flux. The results [47] showed that the CNN-LSTM model was useful to estimate water supply and to make flood warning. The paper [48] demonstrated the results of the comparison of the architectures RNN, LSTM, and GRU. It was shown that LSTM and GRU are often better than RNN in the accuracy of approximation and data forecast, but their convergence takes more time.

At the same time, despite the illustrated examples of successful application of LSTM networks, this architecture has significant complexity compared to standard RNN. The disadvantage of LSTM is the long time for training and, as a consequence, it requires a long machine time [49]. At the same time, LSTM does not guarantee the complete solution of the problems of gradient explode or their vanishing. They occur rarely and quite slowly (during a large number of time steps) [40,49,50]. In spite of the great diversity of the current modification of LSTM, they do not give a significant benefit compared to the initial LSTM [49].

In the paper, we suggest optimizing the process of data modeling by NARX network involving wavelet filtering. The proposed procedure of wavelet filtering allows us to decrease the noise level and to improve NARX network efficiency. Compared to LSTM, the suggested method does not require long-term time series recorded into memory for a retrospective analysis. That makes it possible to use standard RNN without serious risks to obtain the problem of long-term dependencies [37,40], which are reduced by the simplification of input vectors by wavelet filtering. Wavelet filtering is based on wavelet packet construction with the use of stochastic thresholds. In the paper, we introduce

the algorithm of wavelet filtering and propose the technique for estimating stochastic thresholds to obtain solutions with defined confidence coefficient. Moreover, we consider a scheme of implementation of the approach for the problem of anomaly detection in natural data.

The paper considers the ionospheric parameter time series (ionospheric layer F2 critical frequency, foF2). The ionospheric time series have regular variation and anomalies of different forms and time durations. The anomalies are observed during increased solar and geomagnetic activities [3,51]. In seismically active regions, ionospheric anomalies may also occur during earthquakes [51]. The detection of ionospheric anomalies is important in different aspects of life such as space craft operation, radio communication, navigation system operation, etc. The applied traditional methods of time series analysis (median method, moving average, ARIMA models) are not efficient enough to detect ionospheric anomalies [3,52,53]. In the paper, we show that the application of the wavelet filtering procedure makes it possible to obtain a more accurate NARX model of ionospheric parameter time variation. We compare the method with a direct application of NARX networks that also confirms its efficiency. On the example of the analysis of the data during a magnetic storm, the possibility of application of the method to detect anomalies in the space weather problem is illustrated.

2. Method Description

2.1. Wavelet Filtering with Stochastic Thresholds

There is a discrete noisy signal $y(t_n)$ ($n \in \mathbb{N}$, $\mathbb{N}$ are natural numbers including zero)

$$y(t_n) = f(t_n) + e(t_n), \tag{1}$$

where $y(t_n)$ is the recorded data, $f(t_n)$ is a useful signal, and $e(t_n)$ is noise.

To detect the signal structure, according to the paper [20], we apply packet wavelet decompositions [21,22]:

$\mathcal{B}_j^p = \left\{ \phi_j^p \left(t - 2^j k \right) \right\}_{k \in \mathbb{N}}$, where $\mathcal{B}_j^p$ is the basis of the space V_j^p of the wavelet packet tree (Figure 1) generated by the scaling function $\phi_j^p(t) = 2^{-j/2} \phi^p \left(2^{-j} t \right)$;
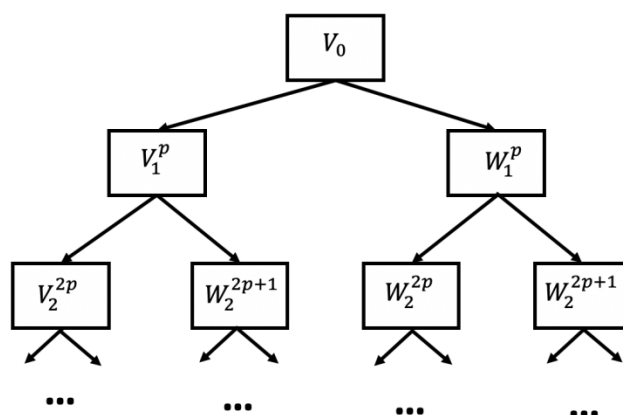

Figure 1. Wavelet packet tree.

$\mathfrak{B}_j^p = \left\{ \Psi_j^p \left(t - 2^j k \right) \right\}_{k \in \mathbb{N}}$, where $\mathfrak{B}_j^p$ is the basis of the space W_j^p of the wavelet packet tree (Figure 1) generated by the wavelet $\Psi_j^p(t) = 2^{-j/2} \Psi^p \left(2^{-j} t \right)$. When moving downwards along the tree of the space V_j^p, W_j^p are divided into orthogonal subspaces

$$V_j^p = V_{j+1}^{2p} \oplus W_{j+1}^{2p+1}; \ W_j^p = W_{j+1}^{2p} \oplus W_{j+1}^{2p+1}. \tag{2}$$

Signal $y(t_n)$ in the wavelet packet space at the decomposition level m has the form

$$y(t_n) = y^p_m(t_n) + \sum_j g^p_j(t_n), \tag{3}$$

where $y^p_m(t_n) = \sum_k c^p_{m,k} \phi^p_{m,k}(t_n)$ is the smoothed component; the coefficients $c^p_{m,k} = \left\langle y, \phi^p_{m,k} \right\rangle$; $\phi^p_{m,k}(t_n) = 2^{-\frac{m}{2}} \phi^p_m(t_n - 2^m k)$, $g^p_j(t_n) = \sum_k d^p_{j,k} \Psi^p_{j,k}(t_n)$ are detailing components; the coefficients $d^p_{j,k} = \left\langle y, \Psi^p_{j,k} \right\rangle$, $\Psi^p_{j,k}(t_n) = 2^{-j/2} \Psi^p_j(t_n - 2^j k)$ are the wavelet.

To determine the decomposition level m, we apply the NAS algorithm suggested in the paper [54]. The NAS algorithm allows us to construct a wavelet packet tree by suppressing noises and detecting signal coherent structures (the algorithm is in Appendix A).

To estimate the signal $\widetilde{f}$, according to the paper [54], we apply the following hard threshold to the absolute values of the coefficients $d^p_{j,k}$ of the components $g^p_j(t_n)$ in each tree node

$$P_{T^p_j}\left(d^p_{j,k}\right) = \begin{cases} d^p_{j,k}, & if \ \left|d^p_{j,k}\right| > T^p_j, \\ 0, & if \ \left|d^p_{j,k}\right| \le T^p_j. \end{cases} \tag{4}$$

The estimate $\widetilde{f}$, based on (4), is

$$\widetilde{f}(t_n) = y^p_m(t_n) + \sum_{j,k} T^p_j\left(d^p_{j,k}\right) \Psi^p_{j,k}(t_n). \tag{5}$$

The risk of such an estimate for $\widetilde{f} \in L^2(\mathbb{R})$ ($L^2(\mathbb{R})$, Lebesgue space [55]), is [56]

$$r\left(\widetilde{f}, f\right) = E\left\{ \|\widetilde{f} - f\|^2 \right\}, \tag{6}$$

where E is the mathematical expectation; $\|\cdot\|$ is the norm.

It is obvious that to minimize the risk r, the threshold T^p_j should be likely higher than the noise coefficient maximum level. As it was shown in the papers [21,22], outside the neighborhoods containing signal local features, absolute values of the coefficients $\left|d^p_{j,k}\right|$ with respect to the argument k are close to zero. Local features in a signal are observed during anomalies and are rare events; thus, based on the thee-sigma rule [57], we can state with high confidence ($\alpha \approx 0.99$) that the values $d^p_{j,k}$ with respect to the argument k are within the interval $(\mu_j - 3\sigma_j; \ \mu_j + 3\sigma_j)$, where $\mu_j \approx 0$ is the mathematical expectance of the value $d^p_{j,k}$, and σ_j is the standard deviation $d^p_{j,k}$.

Then, in the case of normal distribution of the value $d^p_{j,k}$, according to the paper [34], we can estimate the thresholds T^p_j for each level j with the defined confidence coefficient as

$$T^p_j = t_{1-\frac{\alpha}{2}, N-1} \hat{\sigma}_j, \tag{7}$$

where $t_{\alpha,N}$ are α-quantiles Student's distribution [57]; $\hat{\sigma}_j$ is the sample standard diviation of the value $d^p_{j,k}$ which is estimated during the periods without anomalies in the data.

We should note that the risk of estimate (5) is also associated with the error of the approximation f in the basis $\mathfrak{B}$ that should also be taken into account. We can determine the basis for the approximation f as, for example, it was suggested in the paper [20] using the Schur's function [58] and the concordant fitting algorithm [59].

Based on the described operations, we obtain the following algorithm of wavelet filtering:

1. Decomposition of signal y into wavelet packets

$$y(t_n) = y^p_m(t_n) + \sum_j g^p_j(t_n), \tag{8}$$

where $y_m^p(t_n) = \sum_k c^p{}_{m,k} \phi^p{}_{m,k}(t_n)$, $c^p{}_{m,k} = \left\langle y, \phi^p{}_{m,k} \right\rangle$, $\phi^p{}_{m,k}(t_n) = 2^{-\frac{m}{2}} \phi_m^p(t_n - 2^m k)$;
$g_j^p(t_n) = \sum_k d^p{}_{j,k} \Psi^p{}_{j,k}(t_n)$, $d^p{}_{j,k} = \left\langle y, \Psi^p{}_{j,k} \right\rangle$, $\Psi^p{}_{j,k}(t_n) = 2^{-j/2} \Psi_j^p(t_n - 2^j k)$.

2. Application of the threshold function to the coefficients $d^p{}_{j,k}$ of the components g_j^p:

$$P_{T_j^p}\left(d^p{}_{j,k}\right) = \begin{cases} d^p{}_{j,k}, \ if \ \left|d^p{}_{j,k}\right| \geq T_j^p \\ 0, \ if \ \left|d^p{}_{j,k}\right| < T_j^p \end{cases}, \tag{9}$$

where $T_j^p = t_{1-\frac{\alpha}{2}, N-1} \hat{\sigma}_j$, $t_{\alpha,N}$ are the α-quantiles Student's distribution; $\hat{\sigma}_j$ is the sample standard deviation of the value $d^p{}_{j,k}$.

3. Wavelet reconstruction of the signal

$$\widetilde{f}(t_n) = y_m^p(t_n) + \sum_{j,k} P_{T_j^p}\left(d^p{}_{j,k}\right) \Psi^p{}_{j,k}(t_n). \tag{10}$$

2.2. Application of NARX Network

After the wavelet filtering, the signal $\widetilde{f}$ is approximated by the neural NARX network [4,25,27,28]. The architecture of the recurrent NARX PA network [28] is illustrated in Figure 2.
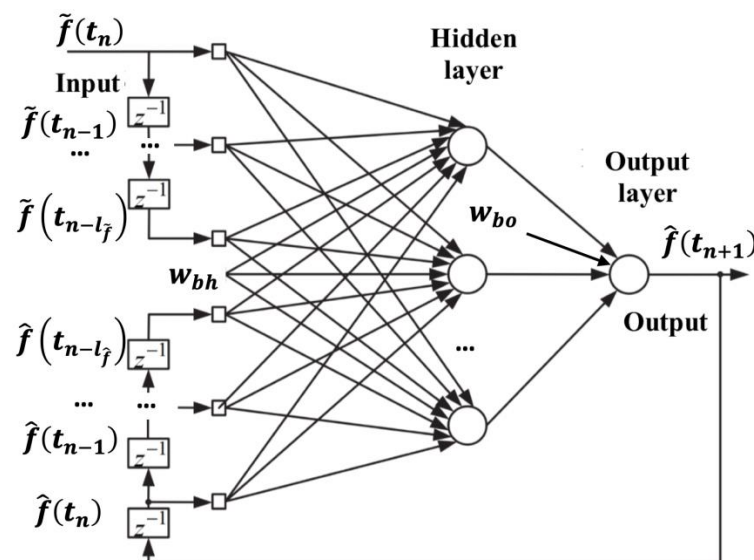

Figure 2. Architecture of the recurrent NARX PA network.

According to the NARX PA network architecture, the network input is $\widetilde{f}(t_n)$, and the network output is $\hat{f}(t_n + 1)$, i.e., the network predicts the data by one step ahead.

The vector, sent to the network hidden layer neurons, consists of the following components:

- current $\widetilde{f}(t_n)$ and preliminary values $\widetilde{f}(t_{n-1}), \ldots, \widetilde{f}\left(t_{n-l_{\widetilde{f}}}\right)$;
- output values $\hat{f}(t_n), \hat{f}(t_{n-1}), \ldots, \hat{f}\left(t_{n-l_{\hat{f}}}\right)$.

The analytical representation of the signal model based on NARX PA has the form

$$\hat{f}(t_{n+1}) = F_o\left(w_{bo} + \sum_{h=1}^{D} w_{ho} \cdot F_h\left(w_{bh} + \sum_{i=0}^{l_{\widetilde{f}}} w_{ih}\widetilde{f}(t_{n-i}) + \sum_{z=0}^{l_{\hat{f}}} w_{zh}\hat{f}(t_{n-z})\right)\right), \tag{11}$$

where $\widetilde{f}(t_n)$ и $\hat{f}(t_n)$ is the NN input signal and its approximation by the network, respectively; $l_{\widetilde{f}}, l_{\hat{f}}$ is the number of delay lines; w_{zh} are weight coefficients of the values arriving from NN output to the hidden layer neurons; w_{ih} are weight coefficients of input values arriving to the hidden layer neurons; w_{bh} and w_{bo} are constant terms for the hidden and output layers, respectively; F_h and F_o are activation functions for the neurons of the hidden and output layers, respectively; w_{ho} are weight coefficients for the values arriving on output layer neurons; D is the number of hidden layer neurons.

It is known that network architecture affects forecast efficiency. NARX model architecture depends on the size of the embedded memory of input $(l_{\widetilde{f}})$, of output $(l_{\hat{f}})$ and the neuron number in the hidden layer. It is not a trivial task to determine these parameters. The number of the delay lines of input $l_{\widetilde{f}}$ and output $l_{\hat{f}}$ can be determined by minimizing the network error [28]. The technique to determine these parameters and the obtained results are represented below in this paper.

The Bayesian regularization algorithm is used for network training. This algorithm updates the weights and the shifts leading to the network neurons according to the Levenberg–Marquardt optimization [25]. The Bayesian regularization technique allows one to form a combination of neuron and weight number of the hidden layer in such a way that the network has the highest degree of generalization [28]. The regularization function minimizes the linear combination of squared errors and weight coefficients of the network at the stage of training. That makes it possible to optimize the neuron number of the hidden layer and to avoid the overtraining effect.

2.3. Scheme of Method Realization

The scheme of method realization is illustrated in Figure 3. Disorder in the system evidently indicates an anomaly in the data. A disorder can be detected based on the analysis of the vector of neural network summary errors estimated in the time window of the length $\mathcal{L} = 2l + 1$

$$\varepsilon_i = \sum_{i=i-l}^{i+l} \left| \hat{f}(i) - \widetilde{f}(i) \right|. \tag{12}$$
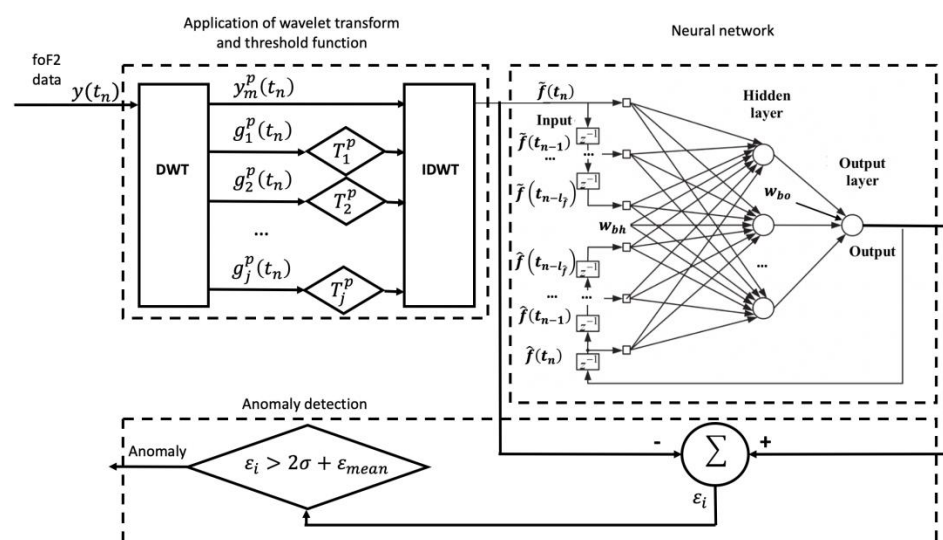

Figure 3. Scheme of method realization.

In this case, we can consider that there is an anomaly in the data if

$$\varepsilon_i > 2\sigma + \varepsilon_{mean}, \tag{13}$$

where σ is the standard deviation of network summary errors, it is estimated during the periods without anomalies; ε_{mean} is the average of network summary errors during the periods without anomalies.

3. Results of Method Application for Ionospheric Data

In the paper, we used the ionospheric layer F2 critical frequency (foF2) data for the period 1969–2015. The data were recorded at IKIR FEB RAS (Paratunka site, Kamchatskiy kray, Russia) from 1968 up to the present time. According to the suggested method, the input and reference values of NN were obtained after the wavelet filtering procedure. In the operations of wavelet filtering, we used orthonormal third-order Daubechies [21], which were determined by minimizing foF2 data approximation errors [60]. To evaluate the method, NN were also trained using the foF2 initial data without wavelet filtering.

When forming training and control NN samples, we took into account the dependence of foF2 data time variation on the season and solar activity level. Thus, NN were constructed for different seasons and different levels of solar activity separately. Two periods of solar activity were considered, the period of high solar activity and the period of low solar activity (the level of solar activity was estimated by mean monthly values of radio emission at the wavelength of f10.7 [61]). To obtain NARX models describing foF2 data regular time variation, data for the periods of calm ionosphere were used in the training process. For each NN, training samples contained one vector of the length from 2000 to 4000 counts. We constructed 24 networks.

When constructing NN, input and output delay lines $l_{\widetilde{f}} = l_{\hat{f}} = 2$, $l_{\widetilde{f}} = l_{\hat{f}} = 3$, $l_{\widetilde{f}} = l_{\hat{f}} = 5$ were used. The parameters $l_{\widetilde{f}}$ and $l_{\hat{f}}$ were determined according to the investigation results of [24]. In the paper [24], an autocorrelation function (ACF) and partial autocorrelation function (PACF) were studied to determine the order of ARIMA models of the foF2 series. It was shown [24] that after wavelet filtering and obtaining the first differences, AR models of foF2 series had the orders 2 and 3 depending on the season and solar activity level. However, we should note that in the general case, the question of the determination of $l_{\widetilde{f}}$ and $l_{\hat{f}}$ is important and requires additional study. As was shown in the paper [28], the NARX model quality significantly depends on the size of input and output embedded memory.

The selection of NN inner architecture was based on NN error estimates. The standard deviations of errors (STD) for the networks were determined as

$$\text{STD} = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} \left(e_i - \bar{e} \right)^2}, \tag{14}$$

where $e = \frac{1}{N} \sum_{i=1}^{N} e_i$, $e_i = \hat{f}_0(i) - \widetilde{f}_0(i)$.

As an example, Table 1 shows NN error estimates for wintertime depending on the neuron number of the hidden layer. An analysis of Table 1 shows that the STD error decreased when the number of neurons was 20 in the hidden layer. Then, the network error remained unchanged. Based on these results, we determined the wintertime network architecture with the number of neurons of the hidden layer equal to 20. Making similar investigations, we determined the number of neurons of the hidden layer were also equal to 20 for the summertime.

Table 1. Construction of network architecture for the data during winter season and high solar activity.

Neuron Number	4	8	12	16	20	24	28
STD	0.54	0.48	0.42	0.42	0.40	0.40	0.41

Table 2 shows the STD errors of NN for different delay lines. An analysis of the results shows that the number of NN approximations increased insignificantly as the number of delay lines grew for the NN trained without wavelet filtering. The application of wavelet

filtering made it possible to increase significantly the NN approximation quality especially during low solar activity. That confirmed the efficiency of the proposed method.

Table 2. Standard deviations of neural network errors.

Season	Delay 2	Delay 2 Wavelet Filtering	Delay 3	Delay 3 Wavelet Filtering	Delay 5	Delay 5 Wavelet Filtering
Winter (low solar activity)	0.71	0.49	0.69	0.5	0.69	0.49
Summer (low solar activity)	0.6	0.33	0.54	0.32	0.46	0.27
Winter (high solar activity)	1.3	0.87	1.31	0.85	1.3	0.87
Summer (high solar activity)	0.54	0.5	0.52	0.51	0.49	0.49

The results of NN quality estimates, based on the control set data during the periods without anomalies, are illustrated in Table 3. The analysis shows significant improvement of NN approximation quality when using the wavelet filtering procedure. The application of wavelet filtering made it possible to decrease significantly the average error (ε_{mean}) and STD (σ) for different delay lines. We should also note that the application of wavelet filtering allowed us to use the low number of input and output delay lines when the network performance quality was good.

Table 3. Results of NN performance (summer and winter, low solar activity).

Delay Lines $l_{\tilde{y}}=l_{\tilde{f}}$	Season	ε_{mean}^{y} without Wavelet Filtering	σ^{y} without Wavelet Filtering	$\tilde{\varepsilon}_{mean}^{f}$ with Wavelet Filtering	$\tilde{\sigma}^{f}$ with Wavelet Filtering
2	Sum	1.3	0.6	0.48	0.33
3	Sum	1.42	0.54	0.47	0.32
5	Sum	1.27	0.46	0.46	0.27
2	Win	1.39	0.71	0.65	0.49
3	Win	1.24	0.69	0.67	0.50
5	Win	1.21	0.69	0.64	0.49

The test for the adequacy of the obtained NARX models was based on the Ljung-Box test [62]:

$$Q = M(M+2)\sum_{s=1}^{L}\frac{\rho_s^2}{M-s},$$ (15)

where M is the observations number, ρ_s is the autocorrelation of the s-th order, and L is the number of lags under the check. If $Q > \chi_{1-\alpha, L}^2$, where $\chi_{1-\alpha, L}^2$ is the quantile of chi-square distribution with L degrees of freedom, then the presence of autocorrelation of the L-th order in the time series is admitted.

The results of the Ljung-Box test are presented in Table 4. The analysis of the results shows that for the networks constructed without wavelet filtering, Ljung-Box test values exceeded significantly the corresponding critical value $\chi_{1-\alpha, L}^2$. That indicates the correlation of network errors and, as a sequence, the not-quite-good-enough quality of foF2 data approximation. We should note that the error correlation grew significantly for large lags L that were evidently associated with the presence of long time dependencies (foF2 diurnal variation) in foF2 data. As was mentioned in the paper [28], in the systems with large time dependencies for the training algorithms based on gradient, information on step gradient m in the past vanished at large m (the problem of vanishing gradients). Thus, the application of RNN faces problems when modeling data with long time dependencies, especially when forecasting nonlinear nonstationary signals [28]. It was also shown in the paper [25] that the problem of vanishing gradients made the investigation of long-term dependencies in training algorithms, based on gradient, difficult and in some cases almost

impossible. When the delay lines grew, the data approximation quality increased (Table 4). However, the adequacy was confirmed only for NARX models obtained with wavelet filtering. For the network delay lines $l_{\tilde{f}} = l_{\hat{f}} = 3$, $l_{\tilde{f}} = l_{\hat{f}} = 5$, according to the Ljung-Box test, network errors were uncorrelated. That confirms their adequacy to foF2 data and shows the efficiency of the suggested method.

Table 4. Ljung-Box test values (summer, low solar activity).

Delay Lines $l_x = l_y$	Lag Number L	Q_y without Wavelet Filtering	$Q_{\tilde{f}}$ with Wavelet Filtering	$\chi^2_{1-\alpha,\ L}$
2	1	0.04	1.47	3.84
2	4	30.4	16.3	9.48
2	8	69.1	28.1	15.5
2	12	128	41.7	21
3	1	0.02	0.04	3.84
3	4	0.9	1.9	9.48
3	8	12.6	9.6	15.5
3	12	64.9	18.2	21
5	1	0.5	0.05	3.84
5	4	1.4	3.8	9.48
5	8	15.6	15.2	15.5
5	12	44.6	19.9	21

Figure 4 shows the results of NN performance during a moderate magnetic storm on 16–17 July 2017. A red dashed line in Figure 4 indicates the time of the magnetic storm beginning. To analyze the geomagnetic activity index, DST index values are illustrated in Figure 4e [63]. During the strongest geomagnetic disturbances on 16 July 2017, the DST index reached the minimum of −72 nT (Figure 4e). An analysis of foF2 data (Figure 4a) shows insignificant changes in fluctuation amplitude during the magnetic storm that was determined by ionospheric disturbance occurrences. Median value data (dashed lines in Figure 4a) also confirmed the presence of an anomaly in the ionosphere during the magnetic storm. The processing results (Figure 4b–d,f–h) illustrated a significant increase in NN errors during the strongest geomagnetic disturbances that indicated anomalous changes in the data. The comparison of the results of the NN trained on the initial data (Figure 4b–d) with the results of the NN obtained after wavelet filtering (Figure 4f–h) confirmed the significant improvement of NN quality when using wavelet filtering. Errors of the NN trained after the wavelet filtering procedure were close to zero. In the error vector of the NN trained on the initial data, an oscillation process was observed which was likely to be associated with foF2 diurnal variation. The result confirms the assumption mentioned above that the application of RNN faced problems when modeling the data with long time dependencies (problem of vanishing gradients). The result also confirms the efficiency of the suggested wavelet filtering procedure to improve the NARX performance quality when modeling nonstationary and noisy data.

Comparing the results of NN with a different number of delay lines, we have learned that for the delay lines $l_{\tilde{f}} = l_{\hat{f}} = 5$ (Figure 4h), anomalous changes were detected on a longer interval that agreed well with the obtained median values of foF2 (dashed line in Figure 4a). A comparison of the NARX results with the median method shows a higher efficiency of the NN. Due to the significant nonstationarity of foF2 data time variation during the magnetic storm, there were errors in median values during the period after the storm on 18 July 2017. There were no errors in the NN model.
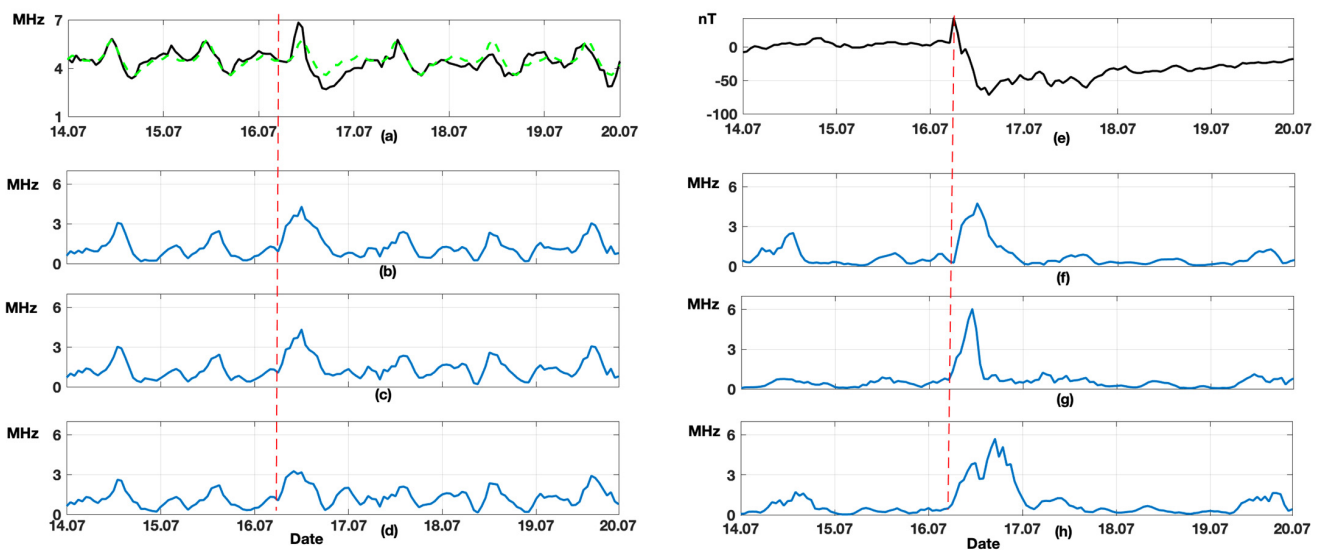
Figure 4. Results of data processing for the period 14–19 July 2017. (**a**) foF2 initial data (black) and foF2 median (green); (**b**–**d**) NN errors with delays 2, 3, and 5, respectively, obtained without wavelet filtering; (**e**) DST; (**f**–**h**) errors with delays 2, 3, and 5, respectively, obtained with wavelet filtering. Red dashed line is the magnetic storm beginning.

Figure 5 shows the results of data processing during a weak magnetic storm on 5–6 August 2019. The red vertical line indicates the magnetic storm beginning (Figure 5). Results in Figure 5 are similar to those illustrated above for the event on 16–17 July 2017. The estimated median values (Figure 5a) show long changes in foF2 data time variation during the magnetic storm. An analysis of NN errors also shows their increase during the event that indicates anomaly occurrence in the ionosphere. The comparison of the results of NN performance without wavelet filtering (Figure 5b–d) and with wavelet filtering (Figure 5f–h) shows a significant decrease in NN errors based on the suggested approach. That is similar to the results of the event on 16–17 July 2017. The NN with the delay lines $l_{\tilde{f}} = l_{\hat{f}} = 5$ (Figure 5h) shows the best results. It has the least errors and the anomalous period in ionospheric data is clearly detectable.

Table 5 shows quantitative estimates of the NN performance during the events described above. The estimates were carried out separately during calm and disturbed periods. An analysis of the results from Table 5 shows that wavelet filtering allows one to decrease significantly the average values of network errors and their STD during calm periods. A decrease in the NN error level by more than 3 times is observed after the wavelet filtering procedure. During the anomalous period, the STD of the network trained after the wavelet filtering may increase by 17 times. That makes it possible to detect anomalous changes in the ionosphere accurately. We should also note that as the number of delay lines increase, the quality of anomaly detection by this network improves. The results confirm the efficiency of the suggested method.
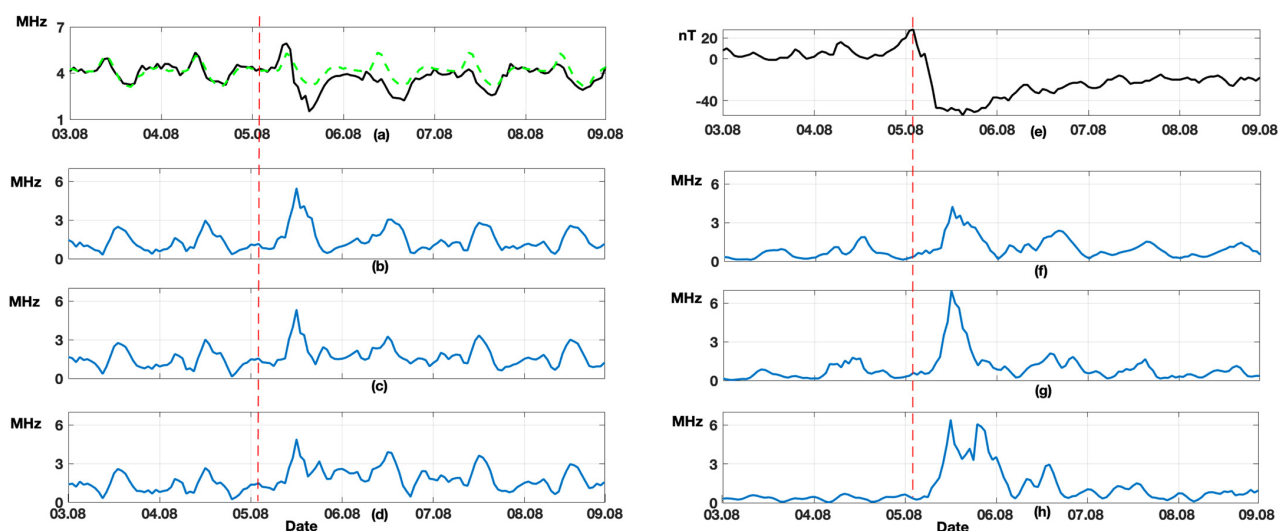
Figure 5. Results of data processing for the period 3–8 August 2019. (**a**) foF2 initial data (black) and foF2 median (green); (**b**–**d**) NN errors with delays 2, 3, and 5, respectively, obtained without wavelet filtering; (**e**) DST; (**f**–**h**) errors with delays 2, 3, and 5, respectively, obtained with wavelet filtering. Red dashed line is the magnetic storm beginning.

Table 5. Results of NN performance.

Delay Lines $l_{\tilde{f}} = l_{\tilde{f}}$	Year	ε_{mean}^{y} without Wavelet Filtering	σ^{y} without Wavelet Filtering	$\tilde{\varepsilon}_{mean}^{\tilde{f}}$ with Wavelet Filtering	$\sigma^{\tilde{f}}$ with Wavelet Filtering
		calm period			
2	2017	1.02	0.63	0.39	0.27
3	2017	1.05	0.57	0.37	0.24
5	2017	0.98	0.53	0.24	0.15
2	2019	1.17	0.61	0.42	0.29
3	2019	1.42	0.64	0.31	0.24
5	2019	1.31	0.58	0.33	0.12
		anomaly period			
2	2017	1.86	1.19	1.81	1.4
3	2017	1.83	1.11	1.6	1.64
5	2017	1.65	0.94	2.58	1.7
2	2019	1.83	1.41	1.68	1.2
3	2019	2.02	1.06	2.18	1.96
5	2019	2.16	0.97	2.88	2.05

4. Conclusions

The application of the method showed its efficiency in the problem of ionospheric data modeling and analysis. The suggested procedure of wavelet filtering allows us to improve the NARX neural network performance quality and gives the possibility to obtain an adequate model for noisy and nonstationary data.

As was mentioned in the papers [25,28], data modeling based on NARX has some difficulties associated with the presence of long time dependences caused by the "vanishing gradient". It was shown on the example of ionospheric data that the wavelet filtering procedure makes it possible to solve this problem if there is a long period. The Ljung-Box test confirmed the adequacy of the obtained neural network models.

On the example of the magnetic storms that occurred on 16–17 July 2017 and on 5–6 August 2019, we confirmed the possibility to apply the method for the detection of ionospheric anomalies based on foF2 data during magnetospheric disturbances. A comparison of the NARX network with the median method, traditionally used for ionospheric data

analysis, showed the NN efficiency. The ionospheric data time variation change during the magnetic storms under analysis entailed error occurrences in the estimates of median values, which were absent in the NN model. The method can be used to monitor the ionosphere state during space weather forecasting.

We plan to continue the investigation in this direction involving ionospheric data from other regions. We also plan to apply the developed method for a more detailed study of ionospheric parameters during increased solar activity and magnetic storms.

Author Contributions: Conceptualization, O.M.; methodology, O.M. and Y.P.; software, Y.P.; validation, Y.P. and N.Z.; formal analysis, O.M. and Y.S.; investigation, O.M. and Y.P.; resources, N.Z. and Y.S.; data curation, Y.P.; writing—original draft preparation, Y.P.; writing—review and editing, O.M.; project administration, O.M.; funding acquisition, Y.S. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

Appendix A

NAS algorithm [54]:

(1) signal X is decomposed into wavelet packets

$W_j^0 : W_j^0 = \oplus_{i=0}^I W_{j_i}^{p_i}$, $\left\{ \Psi_{j_i}^{p_i}(2^{j_i}t - m) \right\}_{m \in \mathbb{N}}$ is the basis of the space $W_{j_i}^{p_i}$;

(2) based on the estimate of normalized energies, we determine the tree branches corresponding to signal structural components: basis $B_{j_i}^{p_i}$ of the space $W_{j_i}^{p_i}$ is the basis

$$B_{j_i}^{p_i} = \begin{cases} \left\{ \Psi_{j_i}^{p_i}(2^{j_i}t - m) \right\}_{m \in \mathbb{N}}, & \text{if } \sum_{m \in I^{p_i}} \left| \left\langle X, \Psi_{j_i,m}^{p_i} \right\rangle \right|^2 \geq \sum_{m \in I^{2p_i}} \left| \left\langle X, \Psi_{j_i+1,m}^{2p_i} \right\rangle \right|^2 + \sum_{m \in I^{2p_i+1}} \left| \left\langle X, \Psi_{j_i+1,m}^{2p_i+1} \right\rangle \right|^2 \\ \left\{ \Psi_{j_i+1}^{2p_i} \right\}_{m \in \mathbb{N}} \cup \left\{ \Psi_{j_i+1}^{2p_i+1} \right\}_{m \in \mathbb{N}}, & \text{if } \sum_{m \in I^{p_i}} \left| \left\langle X, \Psi_{j_i,m}^{p_i} \right\rangle \right|^2 < \sum_{m \in I^{2p_i}} \left| \left\langle X, \Psi_{j_i+1,m}^{2p_i} \right\rangle \right|^2 + \sum_{m \in I^{2p_i+1}} \left| \left\langle X, \Psi_{j_i+1,m}^{2p_i+1} \right\rangle \right|^2 \end{cases} \tag{A1}$$

where the index set I^l, $l = p_i$, $2p_i$, $2p_i + 1$ is determined as follows: index $m \in I^l$, if $\left| \left\langle X, \Psi_{j_i,m}^l \right\rangle \right| \geq T_{j_i}$, threshold $T_{j_i} = K * \sigma_{j_i}^l$, $\sigma_{j_i}^l = \sqrt{\frac{1}{L} \sum_{m=1}^L \left(\left\langle X, \Psi_{j_i,m}^l \right\rangle - \overline{\left\langle X, \Psi_{j_i,m}^l \right\rangle} \right)^2}$, where the coefficient of the threshold K is determined by estimating a posterior risk, $\overline{\left\langle X, \Psi_{j_i,m}^l \right\rangle}$ is the average of the set $\left\{ \left| \left\langle X, \Psi_{j_i,m}^l \right\rangle \right| \right\}_{0 \leq m < L}$, L is the element number.

References

1. Alperovich, L.; Eppelbaum, L.; Zheludev, V.; Dumoulin, J.; Soldovieri, F.; Proto, M.; Bavusi, M.; Loperte, A. A New Combined Wavelet Methodology: Implementation to GPR and ERT Data Obtained in the Montagnole Experiment. *J. Geophys. Eng.* **2013**, *10*, 025017. [CrossRef]
2. Bailey, R.L.; Leonhardt, R. Automated Detection of Geomagnetic Storms with Heightened Risk of GIC. *Earth Planets Space* **2016**, *68*, 99. [CrossRef]
3. 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]
4. Boussaada, Z.; Curea, O.; Remaci, A.; Camblong, H.; Mrabet Bellaaj, N. A Nonlinear Autoregressive Exogenous (NARX) Neural Network Model for the Prediction of the Daily Direct Solar Radiation. *Energies* **2018**, *11*, 620. [CrossRef]
5. Chierici, F.; Embriaco, D.; Pignagnoli, L. A New Real-Time Tsunami Detection Algorithm: Tsunami detection algorithm. *J. Geophys. Res. Oceans* **2017**, *122*, 636–652. [CrossRef]
6. Kim, S.-K.; Lee, E.; Park, J.; Shin, S. Feasibility Analysis of GNSS-Reflectometry for Monitoring Coastal Hazards. *Remote Sens.* **2021**, *13*, 976. [CrossRef]
7. Perol, T.; Gharbi, M.; Denolle, M. Convolutional Neural Network for Earthquake Detection and Location. *Sci. Adv.* **2018**, *4*, e1700578. [CrossRef] [PubMed]
8. Amigó, J.M.; Small, M. Mathematical Methods in Medicine: Neuroscience, Cardiology and Pathology. *Philos. Trans. R. Soc. A* **2017**, *375*, 20170016. [CrossRef] [PubMed]
9. Marple, S.L. *Digital Spectral Analysis: With Applications*; Prentice-Hall signal processing series; Prentice-Hall: Englewood Cliffs, NJ, USA, 1987; ISBN 978-0-13-214149-9.
10. Box, G.E.P.; Jenkins, G.M. *Time Series Analysis: Forecasting and Control, Revised Edition*; Holden-Day series in time series analysis and digital processing; Holden-Day: San Francisco, CA, USA, 1976; ISBN 978-0-8162-1104-3.
11. Liu, J.; Kumar, S.; Palomar, D.P. Parameter Estimation of Heavy-Tailed AR Model With Missing Data Via Stochastic EM. *IEEE Trans. Signal Process.* **2019**, *67*, 2159–2172. [CrossRef]
12. Robbins, H.; Monro, S. A Stochastic Approximation Method. *Ann. Math. Stat.* **1951**, *22*, 400–407. [CrossRef]
13. Estévez, J.; Bellido-Jiménez, J.A.; Liu, X.; García-Marín, A.P. Monthly Precipitation Forecasts Using Wavelet Neural Networks Models in a Semiarid Environment. *Water* **2020**, *12*, 1909. [CrossRef]
14. Pavlicko, M.; Vojteková, M.; Blažeková, O. Forecasting of Electrical Energy Consumption in Slovakia. *Mathematics* **2022**, *10*, 577. [CrossRef]
15. Gocheva-Ilieva, S.; Yordanova, A.; Kulina, H. Predicting the 305-Day Milk Yield of Holstein-Friesian Cows Depending on the Conformation Traits and Farm Using Simplified Selective Ensembles. *Mathematics* **2022**, *10*, 1254. [CrossRef]
16. Li, S.; Wang, Q. India's Dependence on Foreign Oil Will Exceed 90% around 2025—The Forecasting Results Based on Two Hybridized NMGM-ARIMA and NMGM-BP Models. *J. Clean. Prod.* **2019**, *232*, 137–153. [CrossRef]
17. Wu, X.; Zhou, J.; Yu, H.; Liu, D.; Xie, K.; Chen, Y.; Hu, J.; Sun, H.; Xing, F. The Development of a Hybrid Wavelet-ARIMA-LSTM Model for Precipitation Amounts and Drought Analysis. *Atmosphere* **2021**, *12*, 74. [CrossRef]
18. Mbatha, N.; Bencherif, H. Time Series Analysis and Forecasting Using a Novel Hybrid LSTM Data-Driven Model Based on Empirical Wavelet Transform Applied to Total Column of Ozone at Buenos Aires, Argentina (1966–2017). *Atmosphere* **2020**, *11*, 457. [CrossRef]
19. Mehdizadeh, S.; Fathian, F.; Adamowski, J.F. Hybrid Artificial Intelligence-Time Series Models for Monthly Streamflow Modeling. *Appl. Soft Comput.* **2019**, *80*, 873–887. [CrossRef]
20. Mallat, S.G. *A Wavelet Tour of Signal Processing*; Academic Press: San Diego, CA, USA, 1999; ISBN 978-0-12-466606-1.
21. Daubechies, I. *Ten Lectures on Wavelets*; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 1992; ISBN 978-0-89871-274-2.
22. Chui, C.K. *An Introduction to Wavelets*; Wavelet analysis and its applications; Academic Press: Boston, MA, USA, 1992; ISBN 978-0-12-174584-4.
23. Mandrikova, O.; Polozov, Y.; Fetisova, N.; Zalyaev, T. Analysis of the Dynamics of Ionospheric Parameters during Periods of Increased Solar Activity and Magnetic Storms. *J. Atmos. Sol.-Terr. Phys.* **2018**, *181*, 116–126. [CrossRef]
24. Mandrikova, O.; Fetisova, N.; Polozov, Y. Hybrid Model for Time Series of Complex Structure with ARIMA Components. *Mathematics* **2021**, *9*, 1122. [CrossRef]
25. Haykin, S.S. *Neural Networks: A Comprehensive Foundation*, 2nd ed.; Prentice Hall: Upper Saddle River, NJ, USA, 1999; ISBN 978-0-13-273350-2.
26. Singh, J.; Barabanov, N. Stability of Discrete Time Recurrent Neural Networks and Nonlinear Optimization Problems. *Neural Netw.* **2016**, *74*, 58–72. [CrossRef]
27. Ma, Q.; Liu, S.; Fan, X.; Chai, C.; Wang, Y.; Yang, K. A Time Series Prediction Model of Foundation Pit Deformation Based on Empirical Wavelet Transform and NARX Network. *Mathematics* **2020**, *8*, 1535. [CrossRef]
28. Diaconescu, E. The use of NARX neural networks to predict chaotic time series. *WSEAS Trans. Comp. Res.* **2008**, *3*, 182–191.
29. Lin, T.; Horne, B.G.; Tino, P.; Giles, C.L. Learning Long-Term Dependencies in NARX Recurrent Neural Networks. *IEEE Trans. Neural Netw.* **1996**, *7*, 1329–1338. [CrossRef] [PubMed]

30. Gao, Y.; Er, M.J. NARMAX Time Series Model Prediction: Feedforward and Recurrent Fuzzy Neural Network Approaches. *Fuzzy Sets Syst.* **2005**, *150*, 331–350. [CrossRef]

31. Tsungnan, L.; Giles, C.L.; Horne, B.G.; Kung, S.Y. A Delay Damage Model Selection Algorithm for NARX Neural Networks. *IEEE Trans. Signal Process.* **1997**, *45*, 2719–2730. [CrossRef]

32. Dorffner, G. Neural Networks for Time Series Processing. *Neural Netw. World* **1996**, *6*, 447–468.

33. Mandic, D.P.; Chambers, J.A. *Recurrent Neural Networks for Prediction: Learning Algorithms, Architectures, and Stability*; Wiley series in adaptive and learning systems for signal processing, communications, and control; John Wiley: Chichester, UK; New York, NY, USA, 2001; ISBN 978-0-471-49517-8.

34. Mandrikova, O.; Mandrikova, B. Hybrid Method for Detecting Anomalies in Cosmic ray Variations Using Neural Networks Autoencoder. *Symmetry* **2022**, *14*, 744. [CrossRef]

35. Mandrikova, O.; Mandrikova, B. Method of Wavelet-Decomposition to Research Cosmic Ray Variations: Application in Space Weather. *Symmetry* **2021**, *13*, 2313. [CrossRef]

36. Phan, T.-T.-H.; Nguyen, X.H. Combining Statistical Machine Learning Models with ARIMA for Water Level Forecasting: The Case of the Red River. *Adv. Water Resour.* **2020**, *142*, 103656. [CrossRef]

37. Bengio, Y.; Simard, P.; Frasconi, P. Learning Long-Term Dependencies with Gradient Descent Is Difficult. *IEEE Trans. Neural Netw.* **1994**, *5*, 157–166. [CrossRef]

38. Schaefer, A.M.; Udluft, S.; Zimmermann, H.-G. Learning Long-Term Dependencies with Recurrent Neural Networks. *Neurocomputing* **2008**, *71*, 2481–2488. [CrossRef]

39. Glüge, S.; Böck, R.; Palm, G.; Wendemuth, A. Learning Long-Term Dependencies in Segmented-Memory Recurrent Neural Networks with Backpropagation of Error. *Neurocomputing* **2014**, *141*, 54–64. [CrossRef]

40. Pascanu, R.; Mikolov, T.; Bengio, Y. On the difficulty of training Recurrent Neural Networks. In Proceedings of the 30th International Conference on Machine Learning, Atlanta, GA, USA, 17–19 June 2013.

41. Zhang, J.; Lin, Y.; Song, Z.; Dhillon, I.S. Learning Long Term Dependencies via Fourier Recurrent Units. *Int. Conf. Mach. Learn.* **2018**, *80*, 5815–5823. [CrossRef]

42. Yue, B.; Fu, J.; Liang, J. Residual Recurrent Neural Networks for Learning Sequential Representations. *Information* **2018**, *9*, 56. [CrossRef]

43. Gers, F.A.; Schmidhuber, J.; Cummins, F. Learning to Forget: Continual Prediction with LSTM. *Neural Comput.* **2000**, *12*, 2451–2471. [CrossRef]

44. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]

45. Kühnert, C.; Gonuguntla, N.M.; Krieg, H.; Nowak, D.; Thomas, J.A. Application of LSTM Networks for Water Demand Prediction in Optimal Pump Control. *Water* **2021**, *13*, 644. [CrossRef]

46. Jiao, F.; Huang, L.; Song, R.; Huang, H. An Improved STL-LSTM Model for Daily Bus Passenger Flow Prediction during the COVID-19 Pandemic. *Sensors* **2021**, *21*, 5950. [CrossRef]

47. Li, P.; Zhang, J.; Krebs, P. Prediction of Flow Based on a CNN-LSTM Combined Deep Learning Approach. *Water* **2022**, *14*, 993. [CrossRef]

48. Noh, S.-H. Analysis of Gradient Vanishing of RNNs and Performance Comparison. *Information* **2021**, *12*, 442. [CrossRef]

49. Greff, K.; Srivastava, R.K.; Koutnik, J.; Steunebrink, B.R.; Schmidhuber, J. LSTM: A Search Space Odyssey. *IEEE Trans. Neural Netw. Learn. Syst.* **2017**, *28*, 2222–2232. [CrossRef] [PubMed]

50. Gers, F. *Long Short-Term Memory in Recurrent Neural Networks*; EPFL: Lausanne, Switzerland, 2001. [CrossRef]

51. Danilov, A.D. Ionospheric F-Region Response to Geomagnetic Disturbances. *Adv. Space Res.* **2013**, *52*, 343–366. [CrossRef]

52. Tebabal, A.; Radicella, S.M.; Nigussie, M.; Damtie, B.; Nava, B.; Yizengaw, E. Local TEC Modelling and Forecasting Using Neural Networks. *J. Atmos. Sol.-Terr. Phys.* **2018**, *172*, 143–151. [CrossRef]

53. Dmitriev, A.V.; Suvorova, A.V.; Klimenko, M.V.; Klimenko, V.V.; Ratovsky, K.G.; Rakhmatulin, R.A.; Parkhomov, V.A. Predictable and Unpredictable Ionospheric Disturbances during St. Patrick's Day Magnetic Storms of 2013 and 2015 and on 8–9 March 2008. *J. Geophys. Res. Space Phys.* **2017**, *122*, 2398–2423. [CrossRef]

54. Mandrikova, O.; Mandrikova, B.; Rodomanskay, A. Method of Constructing a Nonlinear Approximating Scheme of a Complex Signal: Application Pattern Recognition. *Mathematics* **2021**, *9*, 737. [CrossRef]

55. Rudin, W. Functional Analysis. In *International Series in Pure and Applied Mathematics*, 2nd ed.; McGraw-Hill: New York, NY, USA, 1991; ISBN 978-0-07-054236-5.

56. Korostelev, A.P.; Korosteleva, O. *Mathematical Statistics: Asymptotic Minimax Theory*; Graduate studies in mathematics; American Mathematical Society: Providence, RI, USA, 2011; ISBN 978-0-8218-5283-5.

57. Berger, J.O. *Statistical Decision Theory and Bayesian Analysis*, 2nd ed.; Springer series in statistics; Springer: New York, NY, USA, 1993; ISBN 978-0-387-96098-2.

58. Zhang, X.-M. Schur-Convex Functions and Isoperimetric Inequalities. *Proc. Am. Math. Soc.* **1998**, *126*, 461–470. [CrossRef]

59. Priestley, M.B.; Priestley, M.B. *Spectral Analysis and Time Series*; Probability and mathematical statistics; Repr.; Elsevier: Amsterdam, The Netherlands; Heidelberg, Germany, 2004; ISBN 978-0-12-564922-3.

60. Mandrikova, O.; Polozov, Y.; Geppener, V. Method of Ionospheric Data Analysis Based on a Combination of Wavelet Transform and Neural Networks. *Procedia Eng.* **2017**, *201*, 756–766. [CrossRef]

61. Tapping, K.F. The 10.7 Cm Solar Radio Flux ($F_{10.7}$): F10.7. *Space Weather* **2013**, *11*, 394–406. [CrossRef]

62. Ljung, G.M.; Box, G.E. On a measure of lack of fit in time series models. *Biometrika* **1978**, *65*, 297–303. [CrossRef]
63. Sugiura, M. Hourly values of equatorial Dst for the IGY. *Ann. Int. Geophys. Year* **1964**, *35*, 7–45.