

Article

Method of Constructing a Nonlinear Approximating Scheme of a Complex Signal: Application Pattern Recognition

Oksana Mandrikova, Bogdana Mandrikova * and Anastasia Rodomanskay

Institute of Cosmophysical Research and Radio Wave Propagation, Far Eastern Branch of the Russian Academy of Sciences, Mirnaya st, 7, Paratunka, 684034 Kamchatskiy Kray, Russia; oksanam1@ikir.ru or oksanam1@mail.ru (O.M.); pantina_anastasia@mail.ru (A.R.)

* Correspondence: bogdana@ikir.ru or 555bs5@mail.ru; Tel.: +7-919-436-12-08

Abstract: A method for identification of structures of a complex signal and noise suppression based on nonlinear approximating schemes is proposed. When we do not know the probability distribution of a signal, the problem of identifying its structures can be solved by constructing adaptive approximating schemes in an orthonormal basis. The mapping is constructed by applying threshold functions, the parameters of which for noisy data are estimated to minimize the risk. In the absence of a priori information about the useful signal and the presence of a high noise level, the use of the optimal threshold is ineffective. The paper introduces an adaptive threshold, which is assessed on the basis of the posterior risk. Application of the method to natural data has confirmed its effectiveness.

Keywords: data analysis; nonlinear approximation; wavelet packets; cosmic ray variations; geomagnetic data



Citation: Mandrikova, O.; Mandrikova, B.; Rodomanskay, A. Method of Constructing a Nonlinear Approximating Scheme of a Complex Signal: Application Pattern Recognition. *Mathematics* **2021**, *9*, 737. <https://doi.org/10.3390/math9070737>

Academic Editor: Christine Böckmann

Received: 31 January 2021

Accepted: 24 March 2021

Published: 29 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Currently, scientists are actively conducting research related to the development of methods for modeling and analyzing complex nonstationary signals [1–3]. The need to create such methods arises when carrying out a number of fundamental and applied investigations in such areas as biomedicine, geophysics, ecology, seismology, etc. When there is no possibility to make direct measurements or observations of the characteristics of a research object, the task is to determine the cause from the consequences obtained during observations or experiments (a class of inverse problems). In this case, the determination of the model parameters is based on the observation results. For example, in cosmophysics, the problem arises when determining the state of the galactic cosmic ray flux based on the data of the world network of neutron monitors [4]. In addition, an example of such a problem is to determine the state of the Earth's magnetic field based on the measurements of ground magnetic stations [5].

The recorded natural data have a complex nonstationary structure, but the main problem of such investigations is the lack of a priori information on the useful signal and the presence of a high noise level [4]. The absence of an accurate mathematical apparatus for constructing the estimates of signals with such properties results in the application of heuristic approaches and methods [6], the spectrum of which is very wide at the present time. For example, there have been successful attempts to apply machine learning methods [7], allowing one to obtain approximations of acceptable accuracy even without complete a priori data. However, a significant disadvantage of such methods is the need for periodic correction of system parameters (for example, retraining of a neural network), due to their significant dependence on external conditions. This factor significantly affects the quality and efficiency of the results obtained.

Taking into account these disadvantages, an approach based on the construction of nonlinear adaptive approximating schemes on the basis of orthonormal functions is

proposed. Since the signal distribution has a complex shape, linear approximation is not effective and nonlinear threshold estimates give better results [6]. In the article, wavelets are used as approximating functions. It is known that wavelet filtering makes it possible to effectively detect the structures of a complex signal and suppress noise [6,8–13]. Different wavelet bases approximate the signal in different ways, therefore, the choice of the best basis, in the sense of identifying certain structures, provides an effective solution to the problem. However, we should take into account the fact that if a signal contains different types of structures localized at different times, then it is impossible to build a basis adapted to all structures [6]. In this case, it is necessary to use large basis dictionaries [6,13], for example, as was proposed in [14]. Thus one can extend the class of orthogonal functions by dictionaries of linearly independent functions. An effective result, in this case, is given by the pursuit approximation. For example, dictionaries of wavelet packets and local cosine trees allow for constructing the best approximations of signals of finite length by minimizing the concave cost function [6].

However, the great computational complexity of this method makes it ineffective. Consistent pursuit algorithms [15] using the “greedy” strategy make it possible to optimize the process of constructing a basis and to obtain sufficiently accurate approximations. Even when there is no knowledge about noise, a signal can be assessed by isolating coherent structures [6]. However, when the signal energy is small relative to the noise energy, such estimates give a very small threshold [6] and their application, as was shown by the estimates made in the work, does not allow for obtaining good results. To solve the problem, we propose a method based on a heuristic approach that, by minimizing the posterior risk, makes it possible to obtain the best estimate in the absence of a priori knowledge about a useful signal and the presence of a high noise level. Applying the example of a wavelet packet dictionary, the paper shows that larger threshold increases the risk, but allow one to obtain more accurate estimates. An increase in the detection efficiency of different types of structures is achieved by applying an adaptive threshold. The effectiveness of the proposed method is confirmed by the results of experiments with real data. In addition, the comparison with the machine learning method (using the autoencoder neural network) showed the effectiveness of the proposed approach.

2. Materials and Methods

2.1. Nonlinear Signal Approximation Based on Its Expansion in Basis

In the case of nonlinear approximation, the signal $f \in H$ (H is the Hilbert space) is represented by M vectors adaptively selected from the orthonormal basis $\mathfrak{B} = \{g_m\}_{m \in \mathbb{N}}$ (\mathbb{N} are natural numbers, including 0) of the space H [6]:

$$f_M = \sum_{m \in I_M} \langle f, g_m \rangle g_m, \quad (1)$$

where I_M is a set of indices.

The approximation error is $\epsilon[M] = \|f - f_M\|^2 = \sum_{m \notin I_M} |\langle f, g_m \rangle|^2$.

Obviously, minimization of the error $\epsilon[M]$ is achieved by choosing I_M such that M vectors g_m with indices from I_M have the largest moduli of the scalar product $|\langle f, g_m \rangle|$, that is, correlate with the signal in the best way.

We arrange $\{|\langle f, g_m \rangle|\}_{m \in \mathbb{N}}$ in descending order, denoting $f_{\mathfrak{B}}[k] = \langle f, g_{m_k} \rangle$ the coefficient of rank k as:

$$|f_{\mathfrak{B}}[k]| \geq |f_{\mathfrak{B}}[k+1]|, \quad k > 0.$$

In this case, approximation (1) with the smallest error $\epsilon[M]$ is

$$f_M = \sum_{k=1}^M f_{\mathfrak{B}}[k] g_{m_k}, \quad (2)$$

It can be obtained by applying the threshold function

$$T(x) = \begin{cases} x, & \text{if } |x| \geq T, \\ 0, & \text{if } |x| < T, \end{cases}$$

for the threshold $T : f_{\mathfrak{B}}[M + 1] < T \leq f_{\mathfrak{B}}[M]$.

Then, (2) takes the form

$$f_M = \sum_{m=1}^{+\infty} T(\langle f, g_m \rangle) g_m, \tag{3}$$

Approximation error (3) is

$$\epsilon[M] = \|f - f_M\|^2 = \sum_{k=M+1}^{+\infty} |f_{\mathfrak{B}}[k]|^2.$$

Considering the base dictionary D , which is the union of orthonormal bases in the space of signals of finite length N :

$$D = \cup_{\lambda \in \Lambda} \mathfrak{B}^\lambda,$$

the cost of f approximation in the basis $\mathfrak{B}^\lambda = \{g_m^\lambda\}_{1 \leq m \leq N}$ can be estimated by the concave Schur sum [16]

$$C(f, \mathfrak{B}^\lambda) = \sum_{m=1}^N \Phi\left(\frac{|(f, g_m^\lambda)|^2}{\|f\|^2}\right), \Phi(x) = -x \ln x, \tag{4}$$

and the basis \mathfrak{B}^α , minimizing the error, is defined as

$$(f, \mathfrak{B}^\alpha) = \min_{\lambda \in \Lambda} C(f, \mathfrak{B}^\lambda).$$

In the case of wavelet packet bases [6,8], each node of the wavelet packet tree corresponds to the space W_j^p defining an orthonormal basis $\mathfrak{B}_j^p = \{\Psi_j^p(2^j t - m)\}_{m \in \mathbb{N}}$ [6,8,17]. The space W_j^p is divided into orthogonal subspaces

$$W_j^p = W_{j+1}^{2p} \oplus W_{j+1}^{2p+1}. \tag{5}$$

The basis that minimizes the error is the basis O_j^p of the space W_j^p [16]:

$$O_j^p = \begin{cases} O_{j+1}^{2p} \cup O_{j+1}^{2p+1}, & \text{if } C(f, O_{j+1}^{2p}) + C(f, O_{j+1}^{2p+1}) < C(f, \mathfrak{B}_j^p), \\ \mathfrak{B}_j^p, & \text{if } C(f, O_{j+1}^{2p}) + C(f, O_{j+1}^{2p+1}) \geq C(f, \mathfrak{B}_j^p). \end{cases} \tag{6}$$

Recursive calculation of the bases (6) when moving from the bottom up the tree allows one to determine the wavelet packet basis that minimizes the cost (4).

2.2. Nonlinear Approximation of a Noisy Signal

Let us have a discrete signal $f[n]$, which is defined for $0 \leq n < N$ ($n \in \mathbb{N}$) and is polluted with noise:

$$X[n] = f[n] + V[n], \tag{7}$$

where $X[n]$ are recorded discrete data, $f[n]$ is the signal, $V[n]$ is noise.

Using mapping (3) gives the estimate

$$\tilde{F} = DX = \sum_{m=0}^{N-1} T(\langle X, g_m^\lambda \rangle) g_m^\lambda. \tag{8}$$

The risk of estimating \tilde{F} is

$$r(D, f) = E\left\{\|\tilde{F} - f\|^2\right\},$$

where E is the mathematical expectation.

Following Wald’s theory [18], signals f will be considered as elements of a special set θ , disregarding the probability distribution on this set. Since we do not know the probability distribution of signals, to minimize the risk $r(D, f)$, we can use the minimax approach [6]. The problem is to determine such an operator D that minimizes the maximum risk (minimax risk) [6]:

$$r_O(\theta) = \inf_{D \in O} \sup_{f \in \theta} E\left\{\|\tilde{F} - f\|^2\right\},$$

O is a set of operators performing mapping (8).

It is clear that to minimize the risk $r_O(\theta)$, the threshold T in (8) should be chosen so that there is a high probability that it is greater than the maximum level of noise coefficients $|V_{\mathfrak{B}^\lambda}[m]|$, $V_{\mathfrak{B}^\lambda}[m] = \langle V, g_m^\lambda \rangle$ (see (7)). It was proved in [19] that in the case of white noise with variance σ^2 , the risk close to $r_O(\theta)$ gives the threshold $T_O = \sigma\sqrt{2 \ln N}$:

$$\tilde{F}_O = \sum_{m=0}^{N-1} T_O \left(\langle X, g_m^\lambda \rangle \right) g_m^\lambda \tag{9}$$

$$T_O \left(\langle X, g_m^\lambda \rangle \right) = \begin{cases} \langle X, g_m^\lambda \rangle, & \text{if } |\langle X, g_m^\lambda \rangle| \geq T_O, \\ 0, & \text{if } |\langle X, g_m^\lambda \rangle| < T_O. \end{cases}$$

The risk of estimate (9) is associated with the approximation error f in the basis \mathfrak{B}^λ and can be estimated as [6]

$$r(f) = r(D, f) = \sum_{m=0}^{N-1} \min \left(|f_{\mathfrak{B}^\lambda}[m]|^2, \sigma^2 \right),$$

where $f_{\mathfrak{B}^\lambda}[m] = \langle f, g_m^\lambda \rangle$.

Therefore, the risk of the resulting estimate $r(f)$ depends on the basis.

In addition, when the noise V is not white and $\sigma_m^2 = E\left\{|V_{\mathfrak{B}^\lambda}[m]|^2\right\}$ depends on each vector g_m^λ of the basis, Donohoe and Johnston [19] showed that the threshold estimate with $T_m = \sigma_m\sqrt{2 \ln N}$ also gives a risk close to $r(f)$.

Note that for wavelet packet bases (see (5)), the noise variance σ^2 can be estimated [19] as: $\bar{\sigma}^2 = \frac{M_X}{0.6745}$, where M_X is the median of the set $\{|\langle X, \Psi_{j,m}^p \rangle|\}_{0 \leq m < N/2^j}$, $\Psi_{j,m}^p = \Psi_j^p(2^j t - m)$ is the basis of the space W_j^p for $p = 1$.

In turn, it follows from Jaffard’s theorem [20] (Jaffard’s theorem is given below) and the equivalence of continuous and discrete wavelet expansions [8,17] that when choosing a threshold higher than the maximum amplitude of the noise coefficients, we, using (8), suppress noise and with a high probability keep the coefficients in the vicinity of the signal structural features. It is also clear that the threshold needs to be scaled for better approximation.

Theorem (Jaffard) [20]: If $f \in L^2(\mathbb{R})$ satisfies the Lipschitz condition $\alpha \leq n$ at the point v , then $\exists A$:

$$\forall (a, b) \in \mathbb{R} \times \mathbb{R}^+ \quad |Wf(a, b)| \leq Ab^{\alpha+\frac{1}{2}} \left(1 + \left| \frac{a-v}{b} \right|^\alpha \right),$$

$$W_\Psi f(a, b) = |b|^{-\frac{1}{2}} \int_{-\infty}^{+\infty} f(t) \Psi \left(\frac{t-a}{b} \right) dt.$$

Conversely, if $\alpha < n$ is non-integer and $\exists A, \alpha' < \alpha$:

$$\forall(a, b) \in R \times R^+ |Wf(a, b)| \leq Ab^{\alpha+1/2} \left(1 + \left| \frac{a-v}{b} \right|^{\alpha'} \right),$$

then f satisfies the Lipschitz condition α at the point v .

In this case, the wavelet Ψ must have n zero moments and n derivatives with fast decay. The evidence of Jaffard’s theorem is given in [6].

We get the following algorithm for constructing an approximating scheme (ACAS):

1. We decompose the signal X into wavelet packets (see (5)):

$$W_j^0 : W_j^0 = \oplus_{i=0}^I W_{j_i}^p, \left\{ \Psi_{j_i}^p(2^i t - m) \right\}_{m \in \mathbb{N}}$$

is a basis of the space $W_{j_i}^p$.

2. Based on the estimation of normalized energies, we determine the tree branches corresponding to the structural components of the signal: the basis $B_{j_i}^p$ of the space $W_{j_i}^p$ is the basis:

$$B_{j_i}^p = \begin{cases} \left\{ \Psi_{j_i}^p(2^i t - m) \right\}_{m \in \mathbb{N}}, & \text{if } \sum_{m \in I^p} \frac{|\langle X, \Psi_{j_i}^p \rangle|^2}{\|X\|^2} \geq \sum_{m \in I^{2p}} \frac{|\langle X, \Psi_{j_{i+1}, m}^{2p} \rangle|^2}{\|X\|^2} + \sum_{m \in I^{2p+1}} \frac{|\langle X, \Psi_{j_{i+1}, m}^{2p+1} \rangle|^2}{\|X\|^2}, \\ \left\{ \Psi_{j_{i+1}}^{2p} \right\}_{m \in \mathbb{N}} \cup \left\{ \Psi_{j_{i+1}}^{2p+1} \right\}_{m \in \mathbb{N}}, & \text{if } \sum_{m \in I^p} \frac{|\langle X, \Psi_{j_i}^p \rangle|^2}{\|X\|^2} < \sum_{m \in I^{2p}} \frac{|\langle X, \Psi_{j_{i+1}, m}^{2p} \rangle|^2}{\|X\|^2} + \sum_{m \in I^{2p+1}} \frac{|\langle X, \Psi_{j_{i+1}, m}^{2p+1} \rangle|^2}{\|X\|^2}, \end{cases} \quad (10)$$

where the set of indices $I^l, l = P, 2P, 2P + 1$ is defined as follows: the index $m \in I^l$, if $|\langle X, \Psi_{j_i, m}^l \rangle| \geq T_{j_i}$, threshold $T_{j_i} = K * \sigma_{j_i}^l$, where $X, \Psi_{j_i, m}^l$ is the mean of the set $\{|\langle X, \Psi_{j_i, m}^l \rangle|\}_{0 \leq m < L}$, L is the number of elements.

The nodes of the wavelet packet tree selected on the basis of (10) determine the components that have the greatest correlation with the basis (coherent structures).

The threshold $T_{j_i} = K * \sigma_{j_i}^l$ can be estimated by posterior risk [21].

The threshold splits the space of values of the analyzed function into two nonoverlapping areas Θ_0 and Θ_1 . When using a certain threshold for a given state h_s , the loss average can be determined as

$$R_s(x) = \sum_{z=0}^1 \prod_{sz} P \left\{ x \in \frac{\Theta_z}{h_s} \right\},$$

where \prod_{sz} is the loss function, $P\{x \in \Theta_z/h_s\}$ is the conditional probability of falling within the area Θ_z , if in reality there is a state $h_s, s \neq z, s, z$ are the state indices (“/” sign means the conditional probability). Averaging the conditional risk function over all states h_s , we obtain the average risk:

$$R = \sum_{s=0}^1 p_s R_s,$$

where p_s is the prior probability of the state h_s .

When we do not know the probability distribution of the signal, the posterior probabilities $P\{h_s/x\}, s = 0, 1$ are the most complete characteristics of the states h_s with available a priori data. For a simple loss function $\prod_{sz} = \begin{cases} 1, & s \neq z, \\ 0, & s = z, \end{cases}$ the posterior risk is

$$R = \sum_{s \neq z} P\{h_s/x \in \Theta_z\}.$$

3. Results

3.1. Detection of Geomagnetic Pulsations in Geomagnetic Data

Geomagnetic data are the Earth’s magnetic field variations, which are obtained by magnetometer direct measurements at a magnetic station network [5]. Analysis of geomagnetic data is important in solving practical problems of space weather forecasting [5,22]. The negative impact of geomagnetic anomalies on technical objects determines the importance of developing methods for their detection. The main source of impact is geomagnetic

pulsations (short-period variations of the geomagnetic field) and their detection is of key importance [23]. The structure of geomagnetic data is complex. They contain local features of different structure and duration. Therefore, detection of geomagnetic anomalies is a complex and urgent task. The results of applying the method to geomagnetic data are shown in Figures 1–5.

Figure 1 shows the histograms of two nodes of the wavelet packet tree, the dashed line shows the estimated σ . The results support the need to adapt the threshold to the scale. The tree constructed on the basis of the ACAS up to the 6th level using Coiflet 3 is shown in Figure 2. Geomagnetic pulsations are determined by the detailing nodes of the tree, therefore the nodes $(j_i, 0)$ are not analyzed below. Table 1 and Figure 3 present the error estimation results, which show that Coiflet 5 provides the smallest losses for different periods of solar activity. The results of detection of geomagnetic pulsations using different wavelets are shown in Figure 4.

The results of risk estimates (errors of the 1st and 2nd kind) are presented in Table 2. They show that the threshold coefficient $K = 2.5$ gives the best results. Risks were estimated based on statistical modeling. As a comparison, Figure 5 shows the results of the application of the threshold $T_o = \sigma\sqrt{2 \ln N}$ and the threshold $T_{j_i} = K * \sigma_{j_i}^l$, with $K = 2.5$. It can be seen that threshold T_o , which is close to optimal one, does not allow us to suppress the noise. The results confirm the effectiveness of the proposed method.

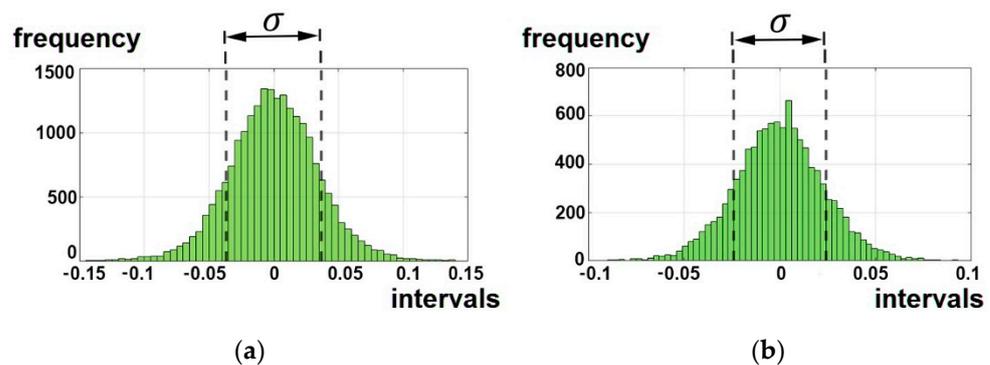


Figure 1. Histograms of tree nodes, Daubechies 3: (a) node (2, 1); (b) node (5, 1).

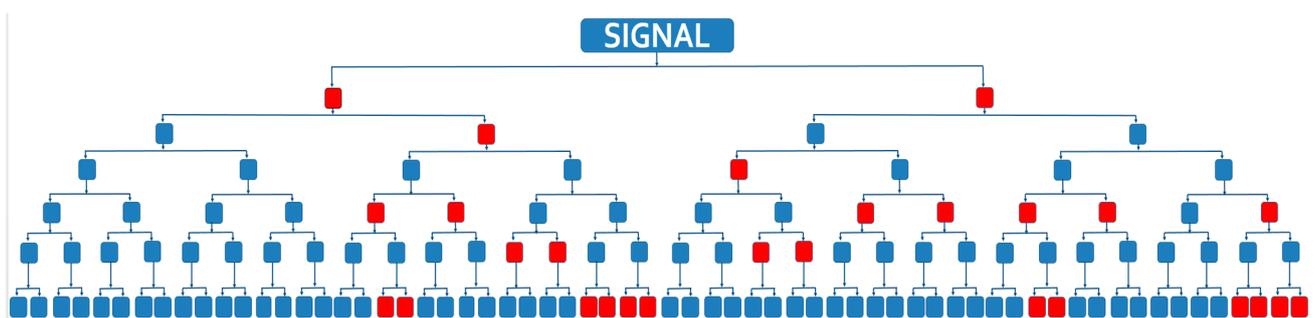


Figure 2. Constructed tree.

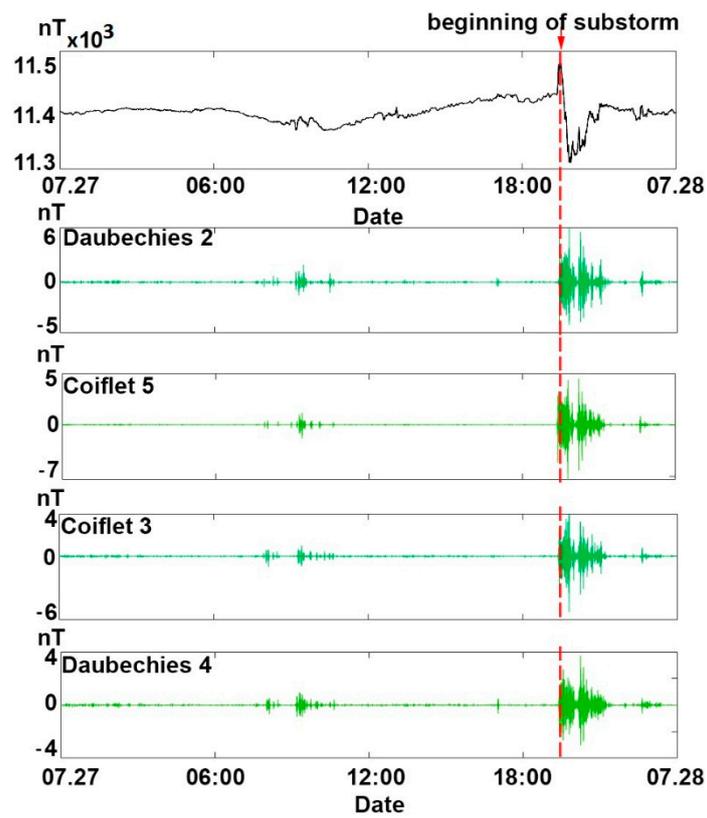


Figure 3. Estimation of the approximation error for different bases.

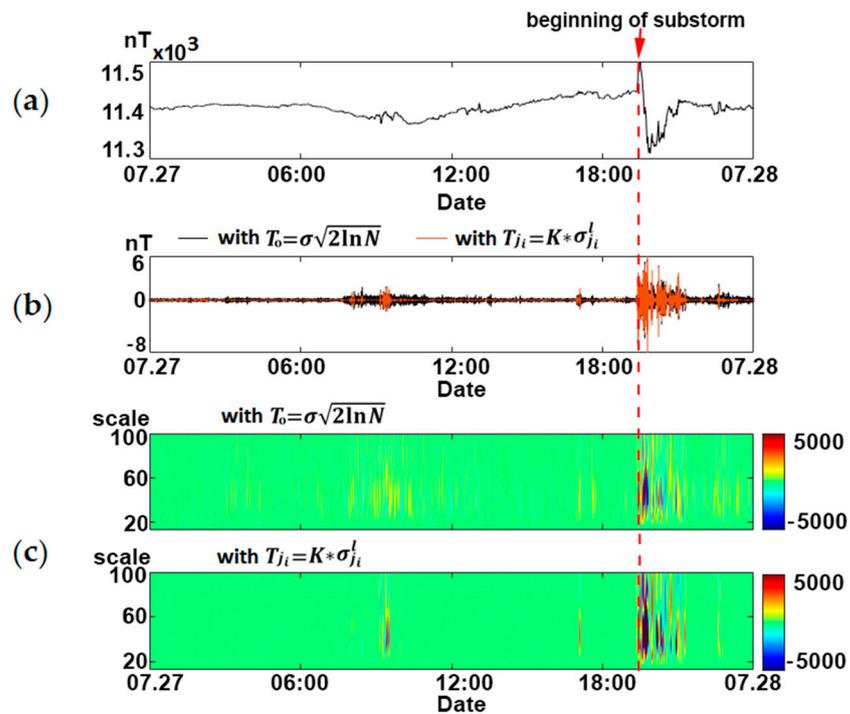


Figure 4. (a) Signal; (b,c) method application (without nodes $(j_i,0)$).

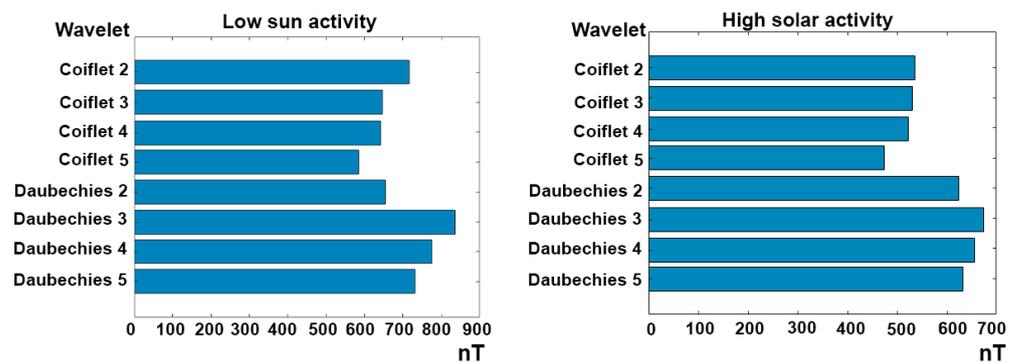


Figure 5. Results with different thresholds.

Table 1. Estimation of the approximation error.

Wavelet, Ψ	Number of Tree Nodes	Approximation Error, $\epsilon[M]$
Daubechies 2	44	57.95443552
Daubechies 2	44	57.95443552
Daubechies 3	44	95.03347346
Daubechies 4	37	73.25697093
Daubechies 5	61	63.16288707
Coiflet 2	42	77.63664436
Coiflet 3	40	70.45379894
Coiflet 4	44	61.81353236
Coiflet 5	48	56.32646977

Table 2. Risks.

Signal/Noise	Threshold Coefficient K = 2.5		Threshold Coefficient K = 1.5	
	Part of Detected (%)	Part of False (%)	Part of Detected (%)	Part of False (%)
1	89	4	87	13
0.8	81	7	49	15
0.7	72	10	66	17

3.2. Detection of Sporadic Features in Neutron Monitor Data

Cosmic ray dynamics are studied using neutron monitor data. Neutron monitor data represent particle counts per unit time and reflect the secondary cosmic ray intensity. In addition to the useful information, the data contain a high level of noise, including natural and human-made interferences [4]. Periodic variations correspond to the regular course and anomalous (sporadic) features characterize Forbush effect occurrences and strong ground level enhancements (GLE events). Sporadic features have different shapes and durations, and their detection is a difficult problem [4]. Of particular interest is the problem of detecting low-amplitude sporadic features that serve as predictors of magnetic storms [4].

Figures 6–11 show the results of applying the method to the neutron monitor (NM) data of the Inuvik station (USA, [24]). Figure 6 shows the histograms of two nodes of the wavelet packet tree. The dashed line shows the estimated σ . The results are similar to geomagnetic data.

To assess the effectiveness of the proposed method, its results were compared with the results of the autoencoder method [25]. Autoencoder is a deep feed forward neural network using back propagation and unsupervised learning. The autoencoder network is described in detail in [25]. The work used undercomplete autoencoders. Minimizing the approximation error, the undercomplete autoencoders allow one to isolate dependencies in the data and to suppress noise. The logistic sigmoidal function was used as the encoder transfer function:

$$f(z) = \frac{1}{1 + e^{-z}}$$

As the transfer function of the decoder, a linear transfer function was used:

$$f(z) = z.$$

To train the network, we used NM data for 2013–2015. Figure 7 shows the results of NM data processing by different methods. The threshold $T_o = \sigma\sqrt{2 \ln N}$, the threshold $T_{j_i} = K * \sigma_{j_i}^l$ and the autoencoder network were used. It is clear that the threshold $T_o = \sigma\sqrt{2 \ln N}$ does not allow us to suppress all the noise. The error of network approximation increases sharply in the test data. That is associated with the effect of retraining and the need to adapt the network. The best results are obtained by the proposed method with the threshold $T_{j_i} = 2.5 * \sigma_{j_i}^l$. Table 3 shows the errors of different methods. Analysis of Table 3 also confirms the effectiveness of the proposed method. Figures 8 and 9 show the results of detecting Forbush effects in NM data based on the proposed method and autoencoder. Data were processed sequentially. First, the data were processed based on the ACAS (algorithm for constructing an approximating scheme). Then, the anomaly detection algorithm described in Appendix A [26] was applied. Alternatively, the data were processed by the autoencoder network, then the anomaly detection algorithm was applied. Forbush effects were detected on 4, 6, 8, 10, 11, 12 September 2014 [27]. In Figure 9, the periods of Forbush effects are marked with red vertical lines. The results show that application of both the autoencoder and the proposed method makes it possible to detect anomalies. It can be seen that the Forbush effect on 22 December was successfully detected based on the proposed method (Figure 9a), and the Forbush effect on 31 December was detected by a neural network (Figure 9b). For detailed comparison of the results, Figures 10 and 11 show periods of Forbush effects of different structures. Figure 10 shows the Forbush effect of a multiscale structure, which was detected by the proposed method. As the result in Figure 11 shows, the Forbush effect of a narrower spectrum was detected by the autoencoder.

Thus, the joint application of these methods increases the efficiency of the problem solution. Estimation of the efficiency of the proposed method is illustrated in Table 4. It shows that its efficiency (over 86%) exceeds that of the autoencoder neural network.

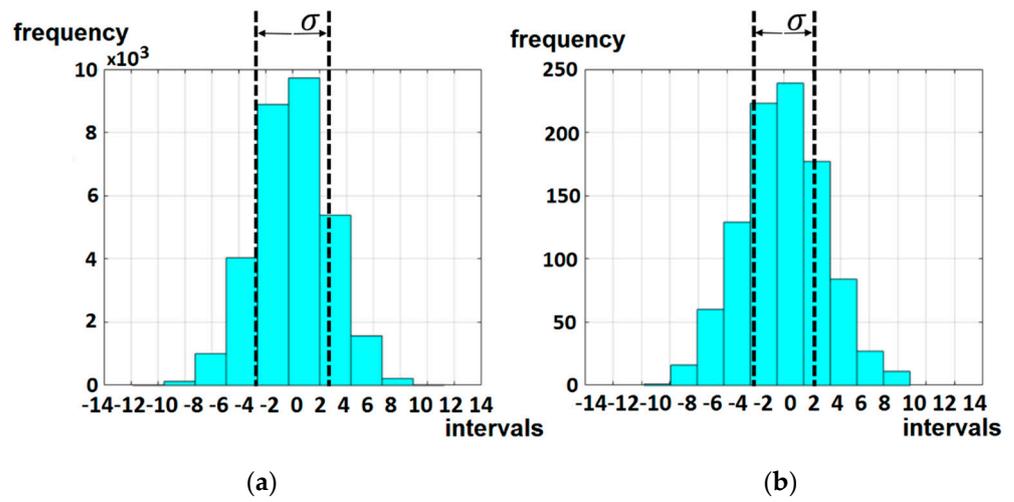


Figure 6. Histograms of tree nodes, Coiflets 3: (a) node (2, 1); (b) node (5, 1).

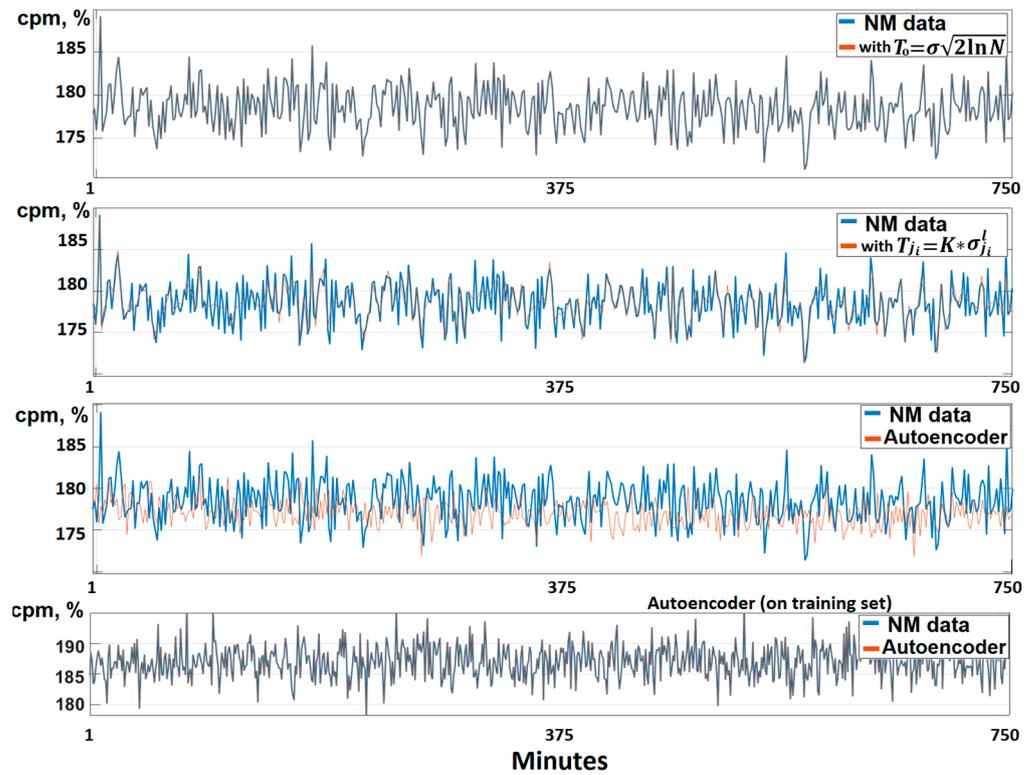


Figure 7. Neutron monitor (NM) data processing results.

Table 3. Estimation of the approximation error.

Method	Approximation Error, $\epsilon[M]$
Autoencoder, training	310.2346
Autoencoder testing	4.2870×10^5
Threshold $T = \sigma\sqrt{2 \ln N}$, Coiflet 1	7.2196×10^{-10}
Threshold $T_{j_i} = K * \sigma_{j_i}^l$, $K = 1$, Coiflet 1	176.0616
Threshold $T_{j_i} = K * \sigma_{j_i}^l$, $K = 1.5$, Coiflet 1	273.2969
Threshold $T_{j_i} = K * \sigma_{j_i}^l$, $K = 2.5$, Coiflet 1	376.4855
Threshold $T_{j_i} = K * \sigma_{j_i}^l$, $K = 1$, Coiflet 2	177.1038
Threshold $T_{j_i} = K * \sigma_{j_i}^l$, $K = 1.5$, Coiflet 2	273.45
Threshold $T_{j_i} = K * \sigma_{j_i}^l$, $K = 2.5$, Coiflet 2	375.741
Threshold $T_{j_i} = K * \sigma_{j_i}^l$, $K = 1$, Coiflet 3	176.9381
Threshold $T_{j_i} = K * \sigma_{j_i}^l$, $K = 1.5$, Coiflet 3	274.2327
Threshold $T_{j_i} = K * \sigma_{j_i}^l$, $K = 2.5$, Coiflet 3	376.1264

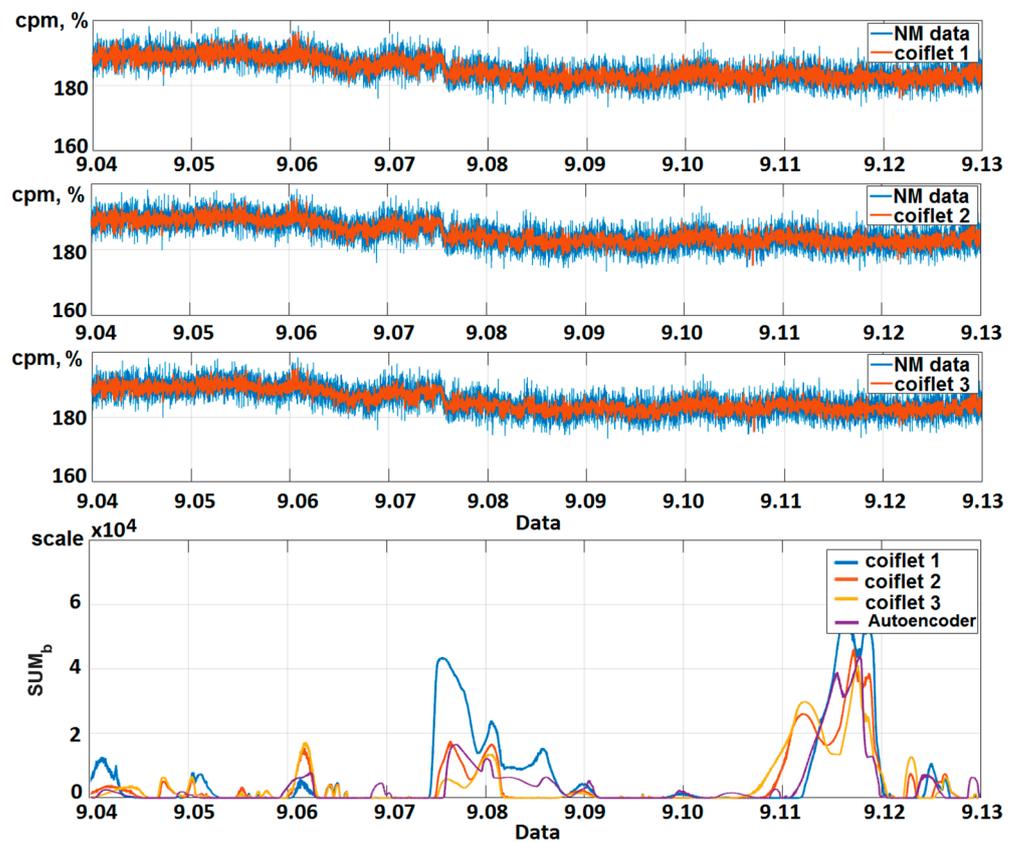


Figure 8. NM data processing results, the threshold used was $T_{j_i} = 2.5 * \sigma_{j_i}^l$.

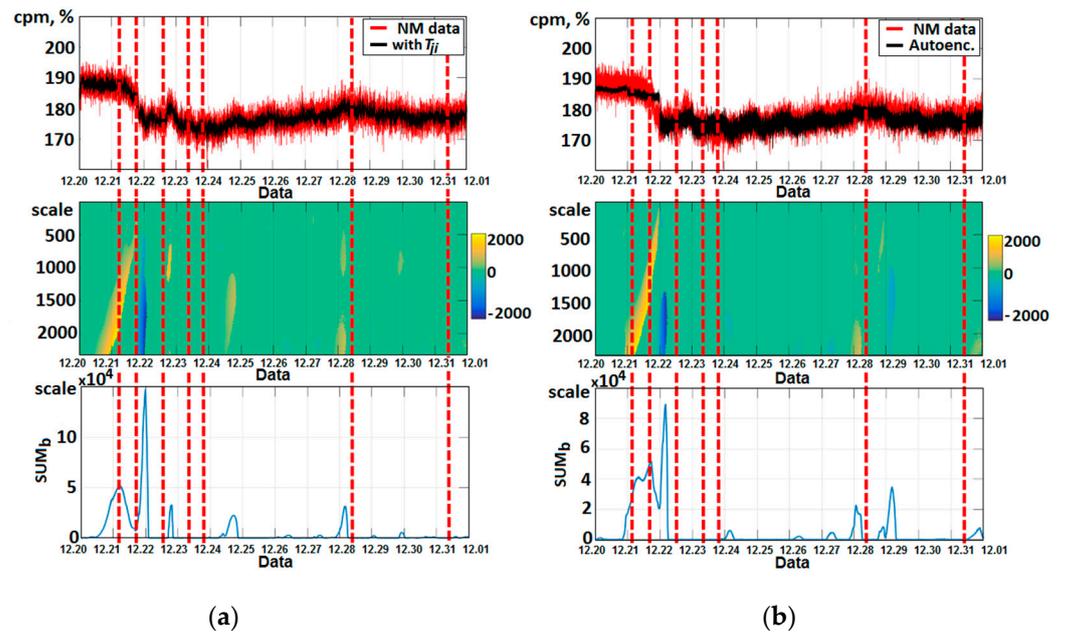


Figure 9. NM data processing results (2014.03.25): (a) the proposed method, the threshold used was $T_{j_i} = 2.5 * \sigma_{j_i}^l$; (b) autoencoder.

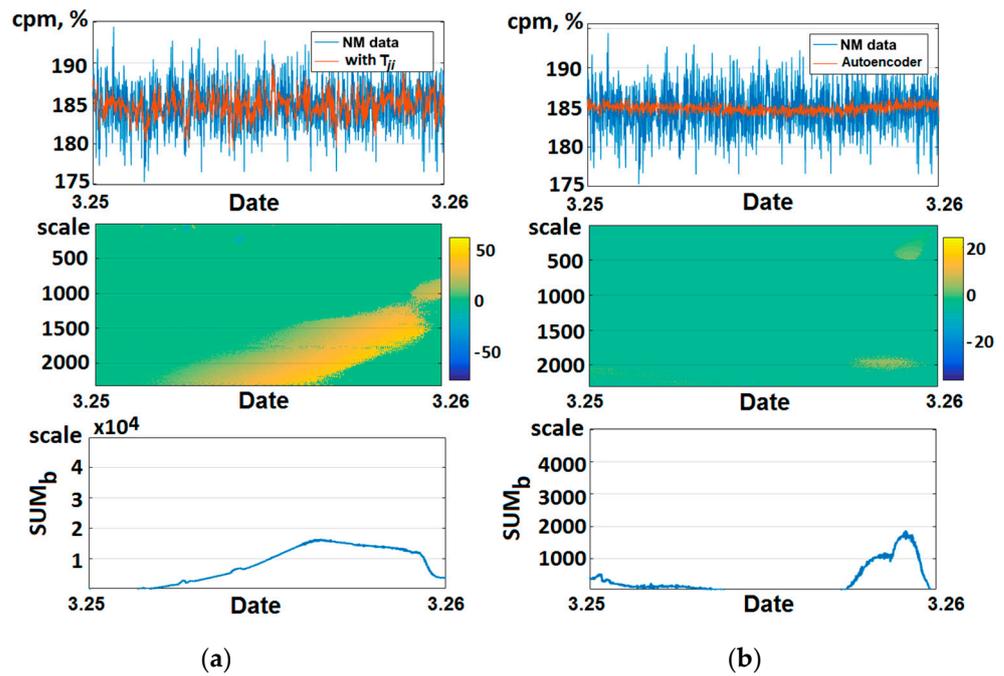


Figure 10. NM data processing results (2014.03.29 from 04:00 to 16:00): (a) the proposed method, the threshold used was $T_{j_i} = 2.5 * \sigma_{j_i}^l$; (b) autoencoder.

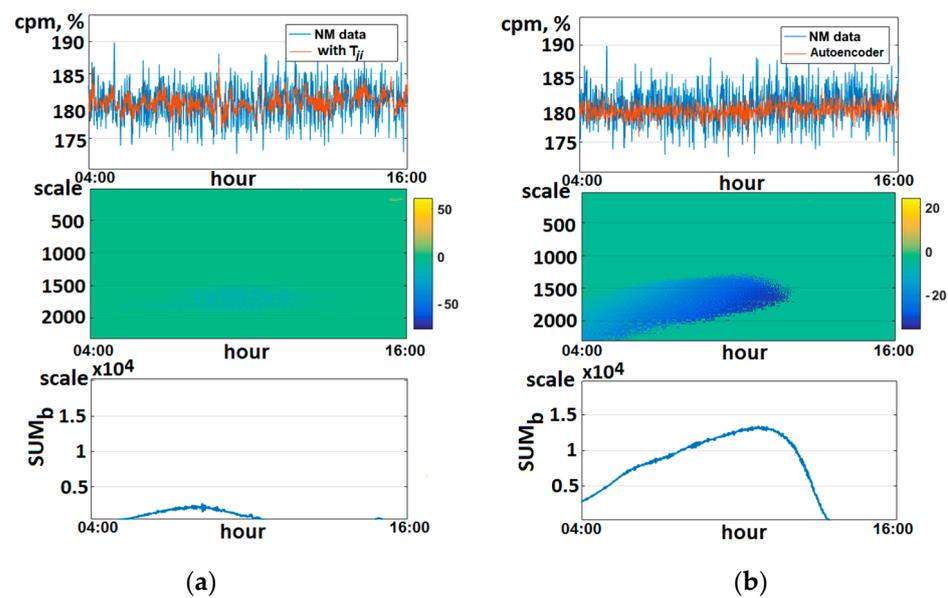


Figure 11. NM data processing results: (a) the proposed method, the threshold used was $T_{ji} = 2.5 * \sigma_{ji}^l$; (b) autoencoder.

Table 4. Estimation of the efficiency of the proposed method.

Year	The Number of Forbush Effects in the Signal	Proposed Method	Autoencoder
2013	98	Detected: 86%	Detected: 79%
		Not detected: 14%	Not detected: 21%
2014	96	False alarm: 16 events	False alarm: 11 events
		Detected: 89%	Detected: 84%
2015	91	Not detected: 11%	Not detected: 16%
		False alarm: 12 events	False alarm: 9 events
		Detected: 84%	Detected: 76%
		Not detected: 16%	Not detected: 24%
		False alarm: 10 events	False alarm: 8 events

4. Conclusions

In this paper, a new method for the identification of structures of a complex signal and noise suppression based on nonlinear approximating schemes is proposed. The experimental results confirmed the effectiveness of the method for the tasks of analyzing natural data and detecting anomalies. Estimates have shown:

1. Application of the proposed adaptive threshold increases the error of nonlinear approximation, but increases the efficiency of anomaly recognition in a complex signal.
2. For signals of an a priori defined structure, a threshold close to optimal is more effective, since it provides the construction of a nonlinear approximating scheme with minimal risk. In the case of white noise, an optimal threshold can be obtained by minimizing the cost function. In addition, the use of this method allows you to control the resulting risk.
3. Comparison of the proposed method with the autoencoder neural network confirmed its high efficiency for natural signals (over 86%) and showed its effectiveness for detecting the features of a multiscale structure. However, the frequency of false alarms (error of the first kind) in the autoencoder is less than that of the proposed method. The autoencoder also detects narrow spectrum features more efficiently.

Author Contributions: Formal analysis, B.M. and A.R.; Methodology, O.M.; Project administration, O.M.; Software, B.M. and A.R. All authors have read and agreed to the published version of the manuscript.

Funding: The work was carried out according to the Subject AAAA-A21-121011290003-0 “Physical processes in the system of near space and geospheres under solar and lithospheric influences” IKIR FEB RAS.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors are grateful to the institutes that support the neutron monitor stations (<http://cosray.unibe.ch/>, <http://spaceweather.izmiran.ru/rus/fds2015.html>) (accessed on 11 November 2020) and ground-based magnetometers (www.inrtermagnet.org) (accessed on 11 November 2020) the data of which were used in the work.

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

Appendix A

Algorithm for detection of anomalies in cosmic rays dynamics and estimation of their intensity [26].

Step 1. Continuous wavelet transform

$$(W_{\Psi}f_{b,s}) := |s|^{-\frac{1}{2}} \int_{-\infty}^{+\infty} f(t) \Psi\left(\frac{t-b}{s}\right) dt,$$

$$f \in L^2(\mathbb{R}), s, b \in \mathbb{R}, s \neq 0$$

Step 2. Application of threshold function P_{T_s} :

$$P_{T_s}(W_{\Psi}f_{b,s}) = \begin{cases} W_{\Psi}f_{b,s}, & \text{if } (W_{\Psi}f_{b,s} - W_{\Psi}f_{b,s}^{med,l}) \geq T_s^l \\ 0, & \text{if } |W_{\Psi}f_{b,s} - W_{\Psi}f_{b,s}^{med,l}| < T_s^l \\ -W_{\Psi}f_{b,s}, & \text{if } (W_{\Psi}f_{b,s} - W_{\Psi}f_{b,s}^{med,l}) < -T_s^l \end{cases}$$

where $W_{\Psi}f_{b,s}^{med,l}$ is the median value calculated in a moving time window of length l . $T_s^l = U * \sigma_s^l$ is the threshold,

$$\sigma_s^l = \sqrt{\frac{1}{l} \sum_{k=1}^l (W_{\Psi}f_{b,s} - \overline{W_{\Psi}f_{b,s}})^2}$$

is the standard deviation calculated in a moving time window of length l , $W_{\Psi}f_{b,s}$ is the average, U is the threshold coefficient.

Step 3. Estimation of anomaly intensity:

$$SUM_s(t) = \sum_b P_{T_s}(W_{\Psi}f_{b,s}),$$

which is positive in the case of CR local increase and is negative in the case of CR local decrease.

References

1. Perez-Sanchez, A.V.; Perez-Ramirez, C.A.; Valtierra-Rodriguez, M.; Dominguez-Gonzalez, A.; Amezcua-Sanchez, J.P. Wavelet Transform-Statistical Time Features-Based Methodology for Epileptic Seizure Prediction Using Electrocardiogram Signals. *Mathematics* **2020**, *8*, 2125. [CrossRef]
2. Shestakov, O. Wavelet Thresholding Risk Estimate for the Model with Random Samples and Correlated Noise. *Mathematics* **2020**, *8*, 377. [CrossRef]
3. 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]
4. Abunina, M.A.; Belov, A.V.; Eroshenko, E.A.; Abunin, A.A.; Yanke, V.G. Ring of Stations Method in Cosmic Rays Variations Research. *Sol. Phys.* **2020**, *295*, 1–20. [CrossRef]

5. Zaitsev, A.N.; Dalin, P.A.; Zastenker, G.N. Sharp variations in the flux of solar wind ions and their response to disturbances of the earth's magnetic field. *Geomagn. Aeron.* **2002**, *6*, 752–759.
6. Mallat, S. *A Wavelet Tour of Signal Processing*; Academic Press: London, UK, 1999; p. 620.
7. Singh, J.; Barabanov, N. Stability of discrete time recurrent neural networks and nonlinear optimization problems. *Neural Netw.* **2016**, *74*, 58–72. [[CrossRef](#)] [[PubMed](#)]
8. Chui, C.K. *An Introduction in Wavelets*; Academic Press: New York, NY, USA, 1992; p. 264.
9. Mandrikova, O.V.; Solovev, I.S.; Zalyaev, T.L. Methods of analysis of geomagnetic field variations and cosmic ray data. *Earth Planet Space* **2014**, *66*, 1–17. [[CrossRef](#)]
10. Alperovich, L.; Zheludev, V.; Hayakawa, M. Application of a wavelet technique for the detection of earthquake signatures in the geomagnetic field. *Nat. Hazards Earth Syst. Sci.* **2001**, *1*, 75–81. [[CrossRef](#)]
11. Alperovich, L.; Zheludev, V.; Hayakawa, M. Use of wavelet analysis for detection of seismogenic ULF emissions. *Radio Sci.* **2003**, *38*, 1093. [[CrossRef](#)]
12. Cao, K.; Zeng, X. Adaptive Wavelet Estimations in the Convolution Structure Density Model. *Mathematics* **2020**, *8*, 1391. [[CrossRef](#)]
13. Herley, C.; Kovacevic, J.; Ramchandran, K.; Vetterli, M. Tilings of the time-frequency plane: Construction of arbitrary orthogonal bases and feist tiling algorithms. *IEEE Trans. Signal Proc.* **1993**, *41*, 3341–3359. [[CrossRef](#)]
14. Chen, S.; Donoho, D. *Atomic Decomposition by Basis Pursuit*; Technical Report; Stanford University: Stanford, CA, USA, 1995.
15. Mallat, S.G.; Zhang, Z.F. Matching pursuits with time-frequency dictionaries. *IEEE Trans. Signal Process.* **1993**, *41*, 3397–3415. [[CrossRef](#)]
16. Coifman, R.R.; Wickerhauser, M.V. Entropy-based algorithms for best basis selection. *IEEE Trans. Inf. Theory* **1992**, *38*, 713–718. [[CrossRef](#)]
17. Daubechies, I. *Ten Lectures on Wavelets*; CBMS-NSF Lecture Notes; SIAM: Philadelphia, PA, USA, 1992.
18. Wald, A. *Statistical Decision Functions*; John Wiley & Sons: New York, NY, USA; Chapman & Hall: London, UK, 1950.
19. Donoho, D.L.; Johnstone, I.M. Ideal spatial adaptation via wavelet shrinkage. *Biometrika* **1994**, *81*, 425–455. [[CrossRef](#)]
20. Jaffard, S. Pointwise smoothness, two-microlocalization and wavelet coefficients. *Publ. Mat.* **1991**, *35*, 155–168. [[CrossRef](#)]
21. Bansal, A.K. *Bayesian Parametric Inference*; Narosa Publishing House Pvt. Ltd.: New Delhi, India, 2007.
22. Kuznetsov, V.D. Space weather and risks of space activities. *Space Technol. Technol.* **2014**, *3*, 3–13.
23. Despirac, I.V.; Kleimenova, N.G.; Gromova, L.I.; Gromov, S.V.; Malysheva, L.M. Supersubstorms during storms on 7–8 September 2017. *Geomagn. Aeron.* **2020**, *60*, 308–317. [[CrossRef](#)]
24. Real Time Data Base for the Measurements of High-Resolution Neutron Monitor. Available online: www.nmdb.eu (accessed on 1 November 2020).
25. Ian, G.; Yoshua, B.; Aaron, C. *Deep Learning*; Random House: New York, NY, USA; p. 808.
26. 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](#)]
27. IZMIRAN Space Weather Forecast Center. Catalog of Forbush Effects and Interplanetary Disturbances. Available online: <http://spaceweather.izmiran.ru/rus/fds2019.html> (accessed on 11 November 2020).