

# Метод обнаружения аномалий в природных данных на основе нейронных сетей и вейвлет-фильтрации

О.В. Мандрикова  
Институт космических исследований и распространения радиоволн ДВО РАН  
Паратунка, Россия  
oksanam1@mail.ru

Ю.А. Полозов  
Институт космических исследований и распространения радиоволн ДВО РАН  
Паратунка, Россия  
up\_agent@mail.ru

**Аннотация** — Предложен автоматизированный метод анализа природных данных и обнаружения аномалий, основанный на совмещении операций вейвлет-фильтрации с нейронной сетью NARX. Построен алгоритм вейвлет-фильтрации и способ оценки порогов, основанный на стохастическом подходе. Показано, что применение вейвлет-фильтрации подавляет шум, упрощает структуру данных и, как следствие, позволяет получить более точную нейросетевую модель NARX. На примере ионосферных данных показана эффективность метода для обнаружения ионосферных аномалий в периоды магнитных бурь.

**Ключевые слова**— анализ данных, вейвлеты, нейронные сети, ионосфера

## 1. ВВЕДЕНИЕ

Задача обработки и анализа природных данных важна для изучения процессов и явлений разной природы и актуальна в различных сферах человеческой деятельности (физика, биология, медицина, экономика и др.). Особо актуальны методы, направленные на диагностику состояний объектов и обнаружение аномалий [1], [2]. В настоящее время для таких задач активно развиваются гибридные подходы и методы [1], [2], которые позволяют повысить качество процедуры анализа данных. В работе предложен метод анализа природных данных, основанный на совместном применении операций вейвлет-фильтрации и нейронных сетей (НС) NARX [3]. Вейвлет-преобразование является гибким инструментом и широко используется в задачах обработки и анализа данных. Обширная библиотека вейвлет-фильтров и широкий набор конструкций разложения обеспечивают возможность адаптации этого инструмента для данных разной структуры [4]. Сети NARX выполняют аппроксимацию временных рядов на основе «моделей нелинейной авторегрессии с экзогенными входами» [3]. К очевидным преимуществам регрессионных моделей относятся их математическая обоснованность, формализованная методика идентификации модели и её проверки на адекватность. В данной работе полученные после вейвлет-фильтрации временные ряды подаются на вход сети NARX. Процедура вейвлет-фильтрации подавляет шум, упрощает структуру данных и повышает эффективность нейронной сети NARX. В работе предложен алгоритм вейвлет-фильтрации и способ оценки порогов. Приведена схема решения задачи обнаружения аномалий.

В качестве экспериментальных данных используются временные ряды критической частоты ионосферного слоя F2 (foF2). Ионосферные временные ряды имеют регулярный ход, а также аномалии разной формы и временной протяженности, которые наблюдаются в

периоды возмущений в околоземном пространстве. Применяемые традиционные методы анализа ионосферных данных не достаточно эффективны для обнаружения ионосферных аномалий [5]. В работе демонстрируется эффективность предлагаемого метода для обнаружения ионосферных аномалий разной частотно-временной структуры. Работа является продолжением исследования [6].

## 2. ОПИСАНИЕ МЕТОДА

В работе использовались данные критической частоты ионосферы foF2, имеющие дискретизацию один час. Регистрация данных выполняется на станции «Паратунка» (Камчаткий край, Россия, координаты станции: 53.0 N, 158.7 E) с 1969 года.

Операции подавления шума включают применение конструкции кратномасштабного анализа [4] и пороговой функции. Используя работу [7], применены стохастические пороги. Алгоритм подавления шума :

1. Вейвлет-разложение сигнала  $f_0(t)$  на компоненты:

$$f_0(t) = \sum_{j=-1}^{-m} g_j(t) + f_{-m}(t),$$

где  $f_{-m}(t) = \sum_k c_{-m,k} \phi_{-m,k}(t)$  – сглаженная компонента,  $c_{-m,k} = \langle f_0, \phi_{-m,k} \rangle$ ,  $\phi_{-m,k}(t) = 2^{-m/2} \phi(2^{-m}t - k)$  – скейлинг-функция,  $g_j(t) = \sum_k d_{j,k} \Psi_{j,k}(t)$  – детализирующие компоненты,  $d_{j,k} = \langle f_0, \Psi_{j,k} \rangle$ ,  $\Psi_{j,k}(t) = 2^{j/2} \Psi(2^j t - k)$  – вейвлет,  $j$  – уровень разложения, для исходного сигнала предполагается уровень разложения  $j = 0$ .

2. Применение пороговой функции к коэффициентам компонент  $g_j(t)$ :

$$T(d_{j,k}) = \begin{cases} 0, & \text{если } |d_{j,k}| \leq T_j \\ d_{j,k}, & \text{если } |d_{j,k}| > T_j \end{cases}, \quad (1)$$

где  $T_j = t_{1-\frac{\alpha}{2}, N-1} \hat{\sigma}_j$ ,  $t_{\alpha, N} - \alpha$ -квантили распределения Стьюдента,  $\hat{\sigma}_j$  – выборочное стандартное отклонение, уровни разложения  $j = -1, -m$ .

3. Вейвлет-восстановление сигнала:

$$\tilde{f}_0(t) = \sum_{j,k} T(d_{j,k}) \Psi_{j,k}(t) + f_{-m}(t).$$

Пороги  $T_j$  в (1) определяются как  $T_j = t_{1-\frac{\alpha}{2}, N-1} \hat{\sigma}_j$ , где  $t_{\alpha, N} - \alpha$ -квантили распределения Стьюдента.

Для построения нейросетевой модели использовались сети архитектуры NARX с обратными связями. В сети использовались блоки линий задержки

входа и выхода  $l_f = l_f$ , значения которых подаются на нейроны скрытого слоя, что позволяет регулировать глубину ретроспективного анализа. Значение выхода нейронной сети имеет вид:

$$\hat{f}_0(t+1) = F[\tilde{f}_0(t), \tilde{f}_0(t-1), \dots, \tilde{f}_0(t-l_x), \hat{f}_0(t), \hat{f}_0(t-1), \dots, \hat{f}_0(t-l_y)].$$

где  $F(\cdot)$  - функция отображения нейронной сети.

### 3. РЕЗУЛЬТАТЫ ПРИМЕНЕНИЯ МЕТОДА ДЛЯ ДАННЫХ ИОНОСФЕРЫ

На рис. 1 показаны результаты обработки данных в период магнитной бури 5-6 августа 2019 г. Начало бури указано красной вертикальной линией. Для анализа уровня геомагнитной активности на рис. 1е показаны

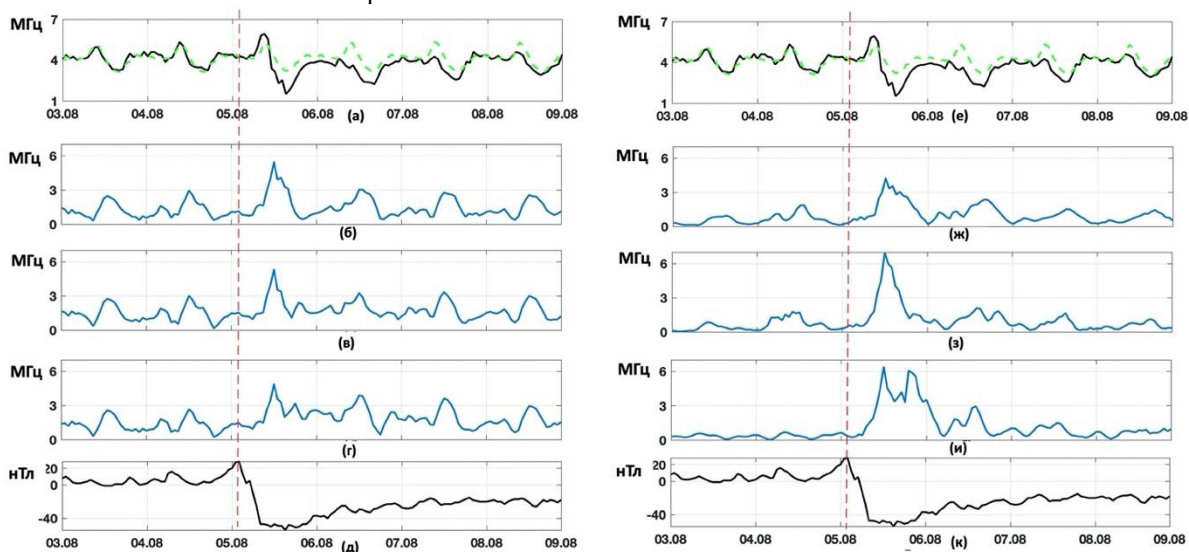


Рисунок 1. Результаты обработки данных за период 03 – 08 августа 2019 года. (а),(е) – исходные данные foF2 (черн.) и медиана foF2 (зел.); (б) - (г) – ошибки НС с задержками 2, 3 и 5, соответственно, полученные без применения вейвлет-фильтрации; (ж) - (з) – ошибки с задержками 2, 3 и 5, соответственно, полученные с применением вейвлет-фильтрации; (д), (к) – DST. Красный пунктир – начало магнитной бури.

### 4. ЗАКЛЮЧЕНИЕ

Применение метода показало его эффективность в задаче моделирования и анализа ионосферных данных. Предлагаемая процедура вейвлет-фильтрации позволяет повысить качество работы нейронных сетей NARX и дает возможность получить адекватную модель для зашумленных и нестационарных данных.

На примере магнитной бури 5-6 августа 2019 г., подтверждена возможность метода для обнаружения ионосферных аномалий по данным foF2 во время магнитосферных возмущений. Метод может быть применен для мониторинга состояния ионосферы при выполнении прогноза космической погоды.

### БЛАГОДАРНОСТИ

Работа выполнена в рамках ГЗ «Физические процессы в системе ближнего космоса и геосфер при солнечных и литосферных воздействиях», регистрационный номер: АААА-А21-121011290003-0. В работе использовалось оборудование Центра коллективного пользования «Северо-восточный гелиогеофизический центр» СКР\_558279.

### ЛИТЕРАТУРА

[1] Wu, X. The development of a hybrid wavelet-ARIMA-LSTM model for precipitation amounts and drought analysis / X. Wu, J. Zhou, H. Yu, D. Liu, K. Xie, Y. Chen, J. Hu, H. Sun, F. Xing // Atmosphere. – 2021. – Vol. 12(1). – P. 74. DOI: 10.3390/atmos12010074.

значения DST индекса геомагнитной активности. Оцененные значения скользящей медианы (рис. 1а) показывают длительные изменения во временном ходе данных foF2 во время бури. Анализ ошибок НС подтверждает рост ошибок в период события, что свидетельствует о возникновении аномалий в данных. Сравнение результатов НС без вейвлет-фильтрации (рис. 5б-г) и с вейвлет-фильтрацией (рис. 5ж-и) показывает значительное уменьшение ошибок НС на основе предложенного в работе подхода. Наилучшие результаты показывает НС с линиями задержки  $l_f = l_f = 5$  (рис. 1з), которая имеет наименьшие ошибки и четко детектирует аномальный период в ионосферных данных.

[2] Sebastian, D.E. Multi-scale association between vegetation growth and climate in India: A wavelet analysis approach / D.E. Sebastian, S. Ganguly, J. Krishnaswamy, K. Duffy, R. Nemani, S. Ghosh // Remote Sensing. – 2019. – Vol. 11(22). – P. 2703. DOI: 10.3390/rs11222703.

[3] Haykin, S. S. Neural networks: a comprehensive foundation / S. S. Haykin, 2nd ed-e. – Upper Saddle River, N.J: Prentice Hall, 1999. – 842 p.

[4] Mallat, S. G. A wavelet tour of signal processing / S. G. Mallat. – San Diego: Academic Press, 1999. – 620 p.

[5] Danilov, A.D. Ionospheric F-region response to geomagnetic disturbances / A.D. Danilov // Advances in Space Research. – 2013. – Vol. 52(3). – P. 343–366. DOI: /10.1016/j.asr.2013.04.019.

[6] Mandrikova, O. Hybrid Model for time series of complex structure with ARIMA components / O. Mandrikova, N. Fetisova, Y. Polozov // Mathematics. – 2021. – Vol. 10(9). – P. 1122. DOI: 10.3390/math9101122.

[7] Mandrikova, O. Hybrid method for detecting anomalies in cosmic ray variations using neural networks autoencoder / O. Mandrikova, B. Mandrikova // Symmetry. – 2022. – Vol. 14(4). – P. 744. DOI: 10.3390/math9070737.