 © Authors of the article / Authors of the article

МЕДИЦИНСКАЯ ИНФОРМАТИКА/MEDICAL INFORMATICS

DOI: https://doi.org/10.60797/BMED.2026.9.7 EDN: EYCODA

A COVID-19 EPIDEMIOLOGICAL DATA STYDY USING WAVELET ANALYSIS

Research article

Elistratov S.A.[1, *]
[1] ORCID : 0000-0002-7006-6879;
[1] Ivannikov Institute for System Programming of RAS, Moscow, Russian Federation
[1] Shirshov Institute of Oceanology of Russian Academy of Sciences, Moscow, Russian Federation
[1] "Sirius" University of Science and Technology, Sochi, Russian Federation

* Corresponding author (sa.elist-ratov[at]yandex.ru)

Suggested: 15.04.2026; Accepted: 03.06.2026; Published: 26.06.2026

Abstract

This work presents a comprehensive investigation into the application of wavelet analysis to real-world pandemic data, specifically focusing on the daily new cases of COVID-19 in Russia. The inherent complexity and multi-scale nature of infectious disease dynamics, exacerbated by factors such as viral evolution and population responses, often challenge the efficacy of traditional epidemiological approaches. Wavelet analysis, with its unique ability to decompose signals into constituent frequencies localized in time, offers a powerful alternative for uncovering hidden patterns and understanding phenomena occurring across diverse temporal scales.

Our analysis delves into the dynamic behavior of COVID-19 incidence across different Russian regions. By comparing these regional dynamics at various time scales, we aim to identify periods where the behavior is consistently similar, thus revealing underlying common drivers or responses. This comparative analysis leads to a crucial conclusion regarding the specific time scales on which these similar behaviors manifest, and anomaly visualization.

The findings derived from this wavelet analysis are directly applicable to the development, generalization, and scaling of data-driven prognostic models. By understanding the time scales that exhibit robust, cross-regional similarities, we can construct more effective and adaptable predictive models. Models informed by these scale-specific insights are better positioned for broader application, reducing the need for extensive refitting across diverse geographical areas and ultimately enhancing our capacity for epidemic forecasting and management.

Keywords: COVID-19, epidemiology, wavelet analysis, dynamic systems.

ИССЛЕДОВАНИЕ ЭПИДЕМИОЛОГИЧЕСКИХ ДАННЫХ COVID-19 С ИСПОЛЬЗОВАНИЕМ ВЕЙВЛЕТ-АНАЛИЗА

Научная статья

Елистратов С.А.[1, *]
[1] ORCID : 0000-0002-7006-6879;
[1] Институт системного программирования им. В.П. Иванникова, Москва, Российская Федерация
[1] Институт океанологии им. П.П. Ширшова РАН, Москва, Российская Федерация
[1] Научно-технологический университет «Сириус», Сочи, Российская Федерация

* Корреспондирующий автор (sa.elist-ratov[at]yandex.ru)

Предложена: 15.04.2026; Принята: 03.06.2026; Опубликована: 26.06.2026

Аннотация

В данной работе представлено всестороннее исследование применения вейвлет-анализа к реальным пандемическим данным — ежедневным новым случаям COVID-19 в России. Неотъемлемая сложность и многомасштабная природа динамики инфекционных заболеваний, усугубляемая такими факторами, как эволюция вируса и реакция населения, часто представляют собой вызов для эффективности традиционных эпидемиологических подходов. Вейвлет-анализ, благодаря своей уникальной способности разлагать сигналы на составляющие частоты, локализованные во времени, предлагает мощную альтернативу для выявления скрытых закономерностей и понимания явлений, происходящих на различных временных масштабах.

Наше исследование углубляется в изучение динамики заболеваемости COVID-19 в различных регионах России. Сравнивая региональную динамику на различных временных масштабах, мы стремимся выявить периоды, когда поведение является устойчиво схожим, тем самым обнажая лежащие в основе общие движущие силы или ответные реакции. Этот сравнительный анализ приводит к ключевому выводу относительно конкретных временных масштабов, на которых проявляются эти сходные паттерны поведения, и визуализации аномалий.

Результаты, полученные с помощью вейвлет-анализа, напрямую применимы к разработке, обобщению и масштабированию прогностических моделей, основанных на данных. Понимая временные масштабы, демонстрирующие надежное, межрегиональное сходство, мы можем создавать более эффективные и адаптируемые предиктивные модели. Модели, основанные на этих масштабно-специфических инсайтах, лучше подходят для широкого применения, уменьшая необходимость в обширной донастройке для различных географических областей и, в конечном итоге, повышая нашу способность к прогнозированию и управлению эпидемиями.

 © Authors of the article / Authors of the article

Ключевые слова: COVID-19, эпидемиология, вейвлет-анализ, динамические системы.

Introduction

The COVID-19 pandemic has presented an unprecedented global challenge, characterized by a continuous surge of new infections and the emergence of distinct viral variants. The sheer volume and granularity of data collected during this period — encompassing daily case counts, geographical spread, demographic impact, and mortality rates — provide a rich resource for understanding the intricate dynamics of infectious disease transmission [1]. However, the complexity inherent in these datasets, further amplified by the evolution of SARS-CoV-2 into various strains (including Alpha, Delta, Omicron ones), demands analytical tools capable of dissecting processes occurring across multiple temporal scales [2].

Different viral strains exhibit varied transmissibility, virulence, and immune evasion properties, leading to the manifestation of epidemic processes at diverse scales. Short-term fluctuations might reflect localized outbreaks driven by a highly transmissible variant, while longer-term trends could be influenced by the introduction of new variants, the impact of public health interventions, or seasonal patterns [1], [2]. Traditional statistical methods, while valuable, often struggle to adequately capture these multi-scale dynamics, making it challenging to disentangle the contributions of various factors and identify critical turning points in the epidemic trajectory.

This is where wavelet analysis [3] emerges as a particularly powerful and advantageous tool. Wavelet transforms offer a unique ability to decompose time-series data into its constituent frequencies while simultaneously preserving temporal localization. This allows for the identification of patterns and anomalies that may be masked by conventional approaches [5]. For the analysis of COVID-19 dynamics, wavelet analysis provides a means to multi-scale resolution, capable to distinguish between short-term fluctuations associated with specific outbreaks and slower, persistent trends indicative of broader epidemiological shifts, as well as identification and visualization of the abrupt changes and anomalies detectable due to the sensitivity of wavelet analysis to localized features in the data [4].

This paper explores the application of wavelet analysis to detailed COVID-19 epidemiological data, demonstrating its efficiency in uncovering intricate patterns and providing a deeper understanding of the multi-scale nature of the pandemic, particularly as influenced by the evolution of its constituent viral strains.

Research methods and principles

The wavelet transform [6] is a mathematical tool that allows for the analysis of a signal in both the time and frequency domains simultaneously. Unlike the traditional Fourier analysis, which represents a signal as a sum of sinusoids of infinite duration, the wavelet transform uses short, wave-like functions called wavelets. These wavelets are localized in both time and frequency, meaning they have a finite duration and a limited frequency band.

The core idea behind the wavelet transform is to decompose a signal into a set of basis functions (wavelets) that are translated and scaled versions of a single “mother wavelet.” The output of the wavelet transform is a set of wavelet coefficients, or a field of continuous wavelet transform field. These coefficients represent the degree of similarity between the signal and the chosen wavelet at a particular scale (inversed frequency) and time location. A large coefficient indicates that the wavelet at that specific scale and time is a good representation of the signal at that point.

Continuous wavelet transform for a given signal $x(t)$ is defined by a formula:

$$CWT[x](\tau, t) = \frac{1}{|a|^2} \int_{-\infty}^{\infty} x(t')\overline{\psi}\left(\frac{t'-t}{\tau}\right) \mathrm{d}t$$

Here ψ is a mother wavelet which is required to be a continuous finite function with a finite spectrum, zero mean and finite L^2 -norm. In the form given the wavelet transform represents the signal's localized temporal behaviour, where t denotes global time and τ is time scale of the local process.

The key advantages of wavelet transform are time-frequency localization, multi-resolution analysis, adaptability for different signal types and successful detection of transient features even for purely-resolved digital sampling.

The wavelet transform is connected with a spectrogram [7] done with the windowed Fourier transform, a Fourier transform extension developed to analyze how the frequency content of a signal changes over time. It addresses a principle limitation of the standard Fourier transform, which provides frequency information for the entire signal duration but loses all temporal information. The method works by dividing the continuous signal into many small, overlapping segments A windowing function (e.g., Hanning, Hamming) is applied to each segment to smoothly taper its edges and reduce spectral leakage. Then, a standard Fourier transform is performed on each of these windowed segments.

$$TFD[x](f, t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x(t')e^{i2\pi f t'} w(t' - t)\mathrm{d}t'$$

where w is a window function.

Based on the wavelet transform, wavelet scattering transform (WST) [8] represents a sophisticated evolution of wavelet analysis, designed to extract robust, invariant, and hierarchical features from complex signals. It addresses limitations of traditional methods by systematically decomposing a signal across multiple scales and translations, while simultaneously ensuring invariance to deformations and translations, which are common in real-world data.

The WST operates through a series of cascaded layers, each employing a wavelet transform followed by a non-linear operation. This layered structure allows for the progressive capture of increasingly complex signal properties.

At the foundation of the WST lies the zero-order scattering coefficient S_0 . This is essentially a low-pass filtered version of the input signal. It is obtained by applying a smoothing filter (preliminary a Gaussian filter) or by averaging the wavelet coefficients over all scales and translations from the first wavelet transform represents the coarse, low-frequency component of the signal, essentially capturing its overall structure without fine details.

 © Authors of the article / Authors of the article

The first layer of the WST generates the first-order scattering coefficients (S_1). For each scale and translation of the initial wavelet transform, a set of coefficients is produced. These coefficients, which represent localized spectral information, are then subjected to a non-linear function (typically a logarithm or a similar compression) and then low-pass filtered. This process effectively captures information about the signal's spectral content at various scales and orientations, but with a degree of translation invariance built in. coefficients thus represent features that describe the presence and intensity of specific frequency components at different locations, robust to minor shifts.

A key principle underpinning the WST's ability to capture multi-scale information is the exponential expansion of scales. The wavelet filters are applied at scales that increase exponentially (or according to a geometric progression). This ensures that the decomposition covers a wide range of temporal or spatial resolutions, from very fine details to very coarse structures. This exponential expansion is crucial for capturing the full spectrum of underlying phenomena, from rapid fluctuations to long-term trends, which is particularly relevant when analyzing signals with diverse temporal characteristics, such as epidemiological data influenced by different viral strains operating at different timescales.

Subsequent layers of the WST build upon the information captured in the previous layers. For instance, the second-order scattering coefficients (S_2) are derived by applying a wavelet transform to the coefficients. This allows the WST to capture relationships and interactions between features detected in the first layer, providing information about the co-occurrence and combined patterns of spectral components. This hierarchical composition continues, creating a rich feature representation that effectively summarizes the signal's structure across multiple levels of complexity and scales.

Main results

As a data, the new cases distributed by the regions of Russia will be used. An example is represented on Figure 1. The data has typical oscillations with a sharp peak in the beginning of 2022 corresponding the Omicron strain appearance.

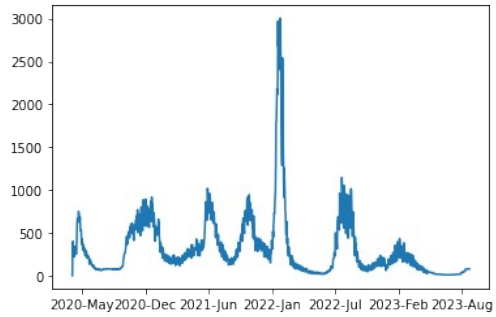

Figure 1 - Overall behaviour of new cases in Moscow
DOI: https://doi.org/10.60797/BMED.2026.9.7.1

Note: example region

The data considered contains about 1200 counts. Despite the frequent records respectively other epidemiological data, they are not enough for the spectrogram of a good quality (Figure 2). The point is that the spectrogram is represened in "time-frequency" axis; the typical frequencies are respectively low, and the existing amount of data does not allow to resolve the spectrogram efficiently both in frequential and temporal localization. As a result, the spectral lines are heavily enwiden. Spectrogram are proper for more data (which is rare in epidemiology) or in orher fields of science (e.g. physics [9]).

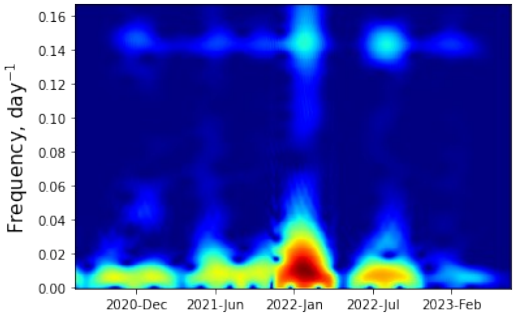

Figure 2 - Spectrogram for Moscow data
DOI: https://doi.org/10.60797/BMED.2026.9.7.2

Wavelet scalogram is capable to solve this problem. Despite the maintaining the principle limitation that the diagram cannot be resolved in the both axes, it turns out to be more representative, partially because it is noted in time scales instead of frequencies, partially because the wavelet transform is more robust for abrupt spectral component appearance. On Figure 3, one can see the typical time scales of ~200 days, with the appearance of smaller scales indicating faster processes in the

 © Authors of the article / Authors of the article

vicinity of Jan 2022 which is connected with Omicron strain. As it can be note, time scales on the wavelet scalogram are more comprehensible than in the spectrogram and visualizes anomalies.

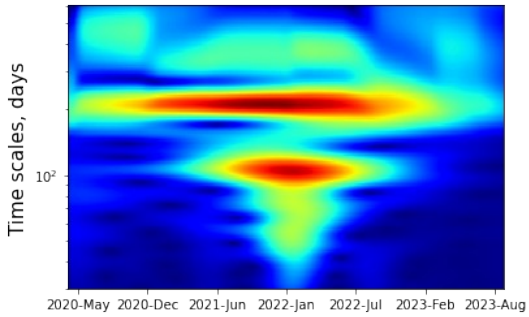

Figure 3 - Wavelet scalogram for Moscow
DOI: https://doi.org/10.60797/BMED.2026.9.7.3

To compare the difference in the behaviour in different regions, the wavelet-coherence [10] was computed. As a measure of the closeness of the wavelets, corresponding the average closeness of the behaviours on the corresponding time scale, Pearson correlation coeffecient was used. The result of the comparison every region with the Moscow data as a reference is represented on Figure 4. It is seen that the main correspondence beween the regions behaviour is localized at the time scales of 110-210 days. The deep burgundy-colored line is the Moscow itself with the correlation 1 for all the scales, leaved on the diagram as a reference.

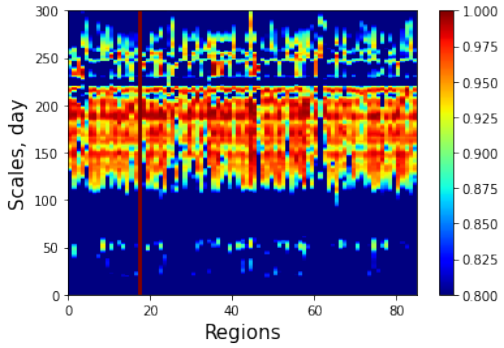

Figure 4 - Wavelet-coherence for Moscow as refercnce region
DOI: https://doi.org/10.60797/BMED.2026.9.7.4

As a deeper look, let us consider wavelet scattering transform (WST) made applied to our data. It is a powerful feature extraction technique that generates a hierarchical and invariant representation of signals. Its effectiveness hinges on a set of defining parameters and the specific nature of the coefficients S_0 and S_1 it produces. WST is governed by parameters such as J representing the total number of scales (or decomposition levels) to be analyzed, and Q which dictates the number of wavelets used to cover each frequency octave. A larger J allows for the examination of features across a broader spectrum of resolutions, from fine to coarse, while a higher Q provides finer discrimination within each frequency band. Together, these parameters define the grid of wavelet filters used, tailoring the analysis to the signal's expected characteristics and the desired level of feature detail. It should be noted that scales investigated has an exponential expansion.

At the foundation of the WST lies the zero-order scattering coefficient S_0 which essentially captures the signal's overall, low-frequency component, acting as a smooth, low-pass filtered version of the input. It represents the global, slow-varying aspects of the signal and is highly invariant to minor translations and deformations because it averages out localized variations. Figure 5 represents the action of S_0 in comparison with the other smoothing methods, including window smoothing (SMA) with 14-days Hanning window and EMD-based smoothing [11]. WST's S_0 turns out to be the heaviest filter, partially losing the information on the sharp-processes intervals.

 © Authors of the article / Authors of the article

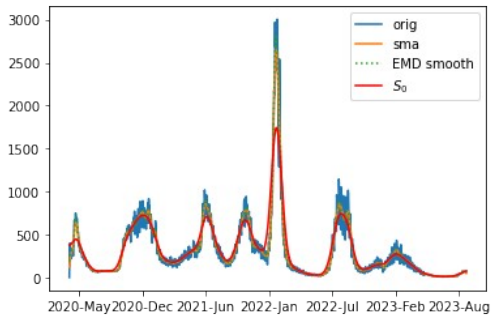

Figure 5 - Illustration of S_0 work and different smoothings for Moscow data
DOI: https://doi.org/10.60797/BMED.2026.9.7.5

However, building upon S_0 , the first-order scattering coefficients S_1 can be useful for the data's behaviour analysis on the different scales are derived from the initial layer of the WST. For each scale and translation of the wavelet transform applied to the input signal, a set of coefficients is generated. These localized spectral features are then processed through a non-linear function followed by a low-pass filter with a window characteristic for a current scale, a process repeated for all specified scales J . This cascading approach ensures that the resulting coefficients are largely invariant to translations of the original signal. Each coefficient quantifies the presence and magnitude of a particular spectral characteristic (defined by its scale and translation) across different signal locations, with minimal sensitivity to their exact position. The ensemble of coefficients thus provides a translation-invariant description of the signal's spectral landscape across various resolutions.

The idea of the application is the following: to compare the local behaviour on the different time scales using S_1 coefficients for different regions. As the wavelet scattering is translationary-invariant, it is expected to be more representative than the wavelet coherence. We continue to use Pearson correlation as a measure between the current region's data and the reference one, but compute a set of coherence diagram for different reference region, after which the data is averages over the reference region to avoid the artifacts. The result is shown on Figure 6. It is seen that the low scales do not correlate at all; the maximum correlation is observed on the scales of 110-210 days (as it was seen on the wavelet-coherence diagram, Figure 4). The large-scale processes with high correlation are too global may be non-representative in the scales of our problem.

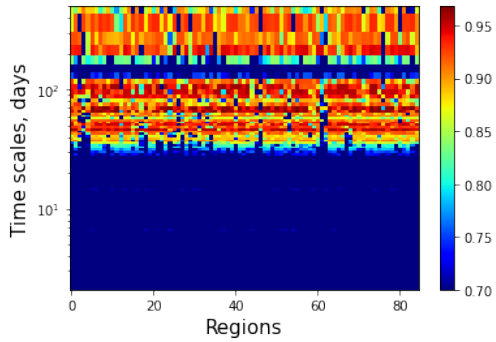

Figure 6 - S1-coherence averaged by reference regions
DOI: https://doi.org/10.60797/BMED.2026.9.7.6

For the enhanced interpretability, we also averaged the S_1-correlation across all analyzed regions. This process generates a consolidated function that illustrates the overall correlation strength across different time scales (Figure 7). The scales corresponding to the maximum value of this function represent those at which the collective behavior of the regional data is most congruent. This finding has significant implications for prognostic modeling. Specifically, models developed to capture dynamics at these scales of peak correlation are more likely to be generalizable across various geographical areas. In contrast, models focused on shorter time scales, which may capture more localized or transient phenomena, would typically necessitate refitting for each distinct region.

 © Authors of the article / Authors of the article

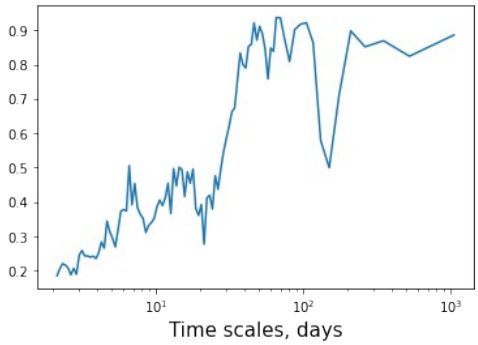

Figure 7 - S1-coherence averaged by reference and target regios
DOI: https://doi.org/10.60797/BMED.2026.9.7.7

Having previously discussed anomaly detection in the context of wavelet scalograms, it is appropriate to introduce an additional visualization tool for identifying non-regular behavior. Epidemiological data, when viewed as a complex dynamical system, evolve in a high-dimensional phase space that is inherently difficult to visualize. However, as demonstrated in the preceding analysis, different regions often exhibit similar long-term dynamics, suggesting that the evolution curves of these systems possess identifiable patterns.

To reconstruct the behavior of this multi-component system, we employ Principal Component Analysis (PCA) [11] combined with dimension reduction based on Takens' embedding theorem [12]. By projecting the system's trajectory onto the plane of the two primary principal components, we can represent the global dynamics compactly and identify deviations from typical patterns.

Figure 8 illustrates the evolution of new case data for Moscow, with the trajectory colored by time. The Omicron outbreak is clearly visible as a distinct deviation from the primary trajectory (indicated in light green, corresponding to January 2022). Outside of this period, the system exhibits quasi-cyclic behavior, highlighting the stability of the long-term epidemic trends. Note that axes PC_1 and PC_2 do not have the direct sense because they are the main directions of the PCA which the phase space is projected on. However, it does not interfere the visualisation.

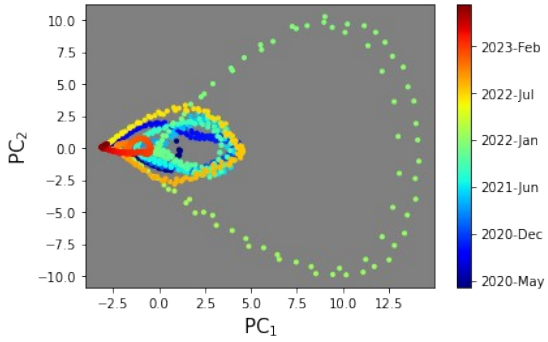

Figure 8 - System behaviour dynamic in Moscow visualization with PCA
DOI: https://doi.org/10.60797/BMED.2026.9.7.8

Figure 9 illustrates the overall evolution of the system. The first significant anomaly is observed during the first half of 2021, which coincides with the emergence of the Delta variant. Additionally, the Omicron outbreak is clearly identifiable as a distinct light-green loop in the phase space.

 © Authors of the article / Authors of the article

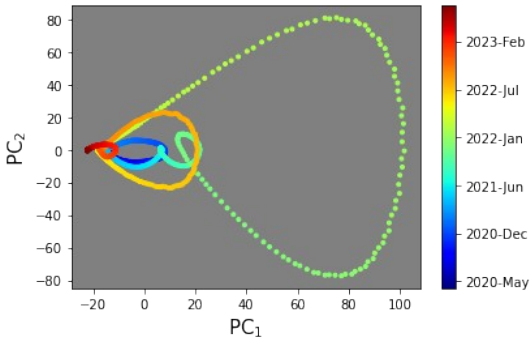

Figure 9 - Overall behaviour dynamic visualization with PCA
DOI: https://doi.org/10.60797/BMED.2026.9.7.9

Conclusion

This study successfully leveraged the power of wavelet analysis to unravel the complex, multi-scale dynamics of COVID-19 epidemiological data. By applying wavelet transforms, we were able to decompose the temporal series of infection rates and other relevant metrics into their constituent frequency components, revealing patterns that operate across distinct temporal scales.

A key contribution of this research lies in the application of wavelet coherence analysis. This powerful technique allowed us to not only identify the presence of correlations between different epidemiological signals (e.g., case counts across regions, or case counts and intervention timelines) but also to pinpoint the specific ranges of temporal scales where these correlations are maximized. This finding is crucial, as it suggests that epidemiological processes, driven by factors such as the emergence and spread of specific viral strains, tend to manifest with similar characteristics at certain scales.

The identification of these “privileged” scales of correlation has significant implications for epidemiological modeling and intervention strategies. It suggests that models developed for understanding disease dynamics in one region may indeed be generalizable and transferable to other regions, provided they are applied at these identified scales. This is because the underlying drivers of epidemic spread that are most strongly correlated across different locations appear to operate within these specific frequency bands. Consequently, understanding the dynamics at these particular scales can offer a more robust foundation for predictive modeling and the development of effective public health responses that are less geographically constrained.

Additionally, the PCA projection of the dynamic system is applied and used for the visualization of the system's behaviour. It is shown that it can be successfully used for detection of the anomal behaviour in the data alongside the wavelet transform.

Финансирование

Проект 23-71-10068, Российский научный фонд.

Funding

Project 23-71-10068, Russian Science foundation.

Конфликт интересов

Не указан.

Conflict of Interest

None declared.

Рецензия

Все статьи проходят рецензирование. Но рецензент или автор статьи предпочли не публиковать рецензию к этой статье в открытом доступе. Рецензия может быть предоставлена компетентным органам по запросу.

Review

All articles are peer-reviewed. But the reviewer or the author of the article chose not to publish a review of this article in the public domain. The review can be provided to the competent authorities upon request.

Список литературы на английском языке / References in English

1. Shuja J. COVID-19 open source data sets: a comprehensive survey / J. Shuja, E. Alanazi, W. Alasmary // Applied Intelligence. — 2021. — № 51. — P. 1296–1325. — DOI: 10.1007/s10489-020-01862-6.

2. Lilly J.M. On the Analytic Wavelet Transform / J.M. Lilly, S.C. Olhede // IEEE Transactions on Information Theory. — 2010. — № 56 (8). — DOI: 10.1109/TIT.2010.2050935.

3. Du M. Comparison of Long COVID-19 Caused by Different SARS-CoV-2 Strains: A Systematic Review and Meta-Analysis / M. Du, Y. Ma, J. Deng [et al.] // International Journal of Environmental Research and Public Health. — 2022. — № 19 (23). — DOI: 10.3390/ijerph192316010.

4. Yang W.C. Discrete wavelet transform for multiple decomposition of gravity anomalies / W.C. Yang, Z.H. Shi, Z.Z. Hou // Chinese Journal of Geographics. — 2013. — № 44. — P. 529–537. — DOI: 10.1002/cjg2.171.

5. Elistratov S.A. Data discretization in prognostic models for epidemiology / S.A. Elistratov // Cifra. Biomedical Sciences. — 2024. — № 3. — DOI: 10.60797/BMED.2024.3.2.

6. Yves M. Wavelets and Operators / M. Yves. — Cambridge : Cambridge University Press, 1992. — 223 p.

7. Sejdic E. Quantitative Performance Analysis of Scalogram as Instantaneous Frequency Estimator / E. Sejdic, I. Djurovic, L. Stankovic // IEEE Transactions on Signal Processing. — 2008. — № 56 (8). — P. 3837–3845.

8. Andén J. Deep scattering spectrum / J. Andén, S. Mallat // IEEE Trans. Signal Process. — 2014. — № 62. — P. 4114–4128.

 © Authors of the article / Authors of the article

9. Elistratov S. Towards the Salinity Profile Influence on an Internal Wave Attractor Formation / S. Elistratov, I. But // Water waves. — 2024. — № 6. — P. 333–348. — DOI: 10.1007/s42286-024-00094-2.

10. Vacha L. Co-movement of energy commodities revisited: Evidence from wavelet coherence analysis / L. Vacha, J. Barunik // Energy Economics. — 2012. — № 34 (1). — DOI: 10.1016/j.eneco.2011.10.007.

11. Elistratov S.A. COVID-19 Epidemiological Indicators POD Spatial Decomposition / S.A. Elistratov // Proceedings of ISP RAS. — 2024. — № 36 (2). — P. 181–192. — DOI: 10.15514/ISPRAS-2024-36(2)-13.

12. Takens F. Detecting strange attractors in turbulence / F. Takens; edited by D.A. Rand, L.-S. Young // Dynamical Systems and Turbulence, Lecture Notes in Mathematics. — 1981. — Vol. 898. — P. 366–381.