International Journal of Academic Information Systems Research (IJAISR)
ISSN: 2643-9026
Vol. 5 Issue 9, September - 2021, Pages: 25-30

Wavelet ideology as a universal tool for data processing and analysis: some application examples

Vyacheslav Lyashenko[1], Zhanna Deineko[2], Oleksandr Zeleniy[3], Iryna Tabakova[4]

Department of Media Systems and Technology, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine
e-mail: lyashenko.vyacheslav@gmail.com[1], zhanna.deineko@nure.ua[2], oleksandr.zeleniy@nure.ua[3],iryna.tabakova@nure.ua[4]

Abstract: *Data analysis is an important stage in the study of various processes, phenomena, objects. These data can have a different physical nature and different structure. Therefore, it is important to have a set of versatile tools for the appropriate analysis. We single out wavelet analysis among such tools. The paper shows various directions for using wavelet analysis, shows a formalized description of such directions. Specific examples are also given for each direction of wavelet analysis.*

Keywords — wavelet ideology, data analysis, time series, image processing, wavelet coherence.

1. Introduction

Currently, information is one of the main resources. Information plays an important role in any scientific research, production activity, and human life. Timely and reliable information is the basis for the development of the economy, the diagnosis of human diseases, the creation of new materials and technologies.

Information can be presented in any form [1], [2]. It can be a dataset in the form of a time series, a dataset in the form of a table, a description of semantic relationships between objects, various types of digital images. In general, the information is presented as a primary dataset.

Various approaches, methods, procedures can be used to analyze information [3]-[5]. The choice of such information analysis tools depends on the problem that needs to be solved and on the form of information presentation. However, it is important not only to analyze the primary information. It is also necessary to be able to obtain additional information from the primary data. Ultimately, this is the main factor that determines the choice of tools for analyzing primary data. Among such universal tools, it is necessary to indicate the methods of wavelet ideology. It is the confirmation of this fact that is considered further in the article.

2. Materials and Methods

2.1 Related Work

There are many works that deal with various issues of primary data analysis, information processing, obtaining additional information and new knowledge.

As a rule, among the data analysis tools, statistical methods occupy a special place.

For example, in [6], attention is paid to descriptive statistics. The authors then build various tests against the normalized raw data. To do this, the authors use concepts such as a measure of mean values, variance. This approach to data analysis allows you to visualize the data processing process. It helps to make informed and correct decisions based on the data that is being researched.

In work [7] preliminary methods of data analysis are considered. These methods allow you to select the necessary statistical approaches for data analysis. The authors also substantiate the need to present data in the desired form for their analysis using statistical methods. This is due to the fact that you need to know the assumptions and conditions for using statistical methods. This is important for getting the right conclusions and decisions.

T. Górecki, M. Krzyśko, Ł. Waszak and W. Wołyński analyze special statistical methods for the study of multivariate data [8]. For this, the authors preliminarily define functional dependencies in the form of a data vector. This allows you to pre-isolate groups of related data. Further, such groups are considered as realizations of multidimensional random variables.

In the work of S. Washington, M. Karlaftis, F. Mannering and P. Anastasopoulos, various methods that are used in the field of statistical analysis are considered and summarized [9]. The authors consider both simple methods of statistical analysis and econometric methods.

The work [10] summarizes the methods of statistical analysis that are used to study medical and biomedical data. At the same time, the authors pay attention to the general procedure for data analysis, the correspondence of these data to the chosen procedure of statistical analysis. The authors also consider the issues of choosing statistical hypotheses and procedures for testing such hypotheses.

Other approaches are used to analyze data. So in work [11] methods of graphical analysis are considered. These methods allow you to visualize the data analysis process. The decision-making process is also visualized as a result of the corresponding analysis.

J. Ho, T. Tumkaya, S. Aryal, H. Choi and A. Claridge-Chang also pay attention to the graphical component in data analysis [12]. The authors consider various statistical hypotheses, taking into account the estimated graphs.

International Journal of Academic Information Systems Research (IJAISR)
ISSN: 2643-9026
Vol. 5 Issue 9, September - 2021, Pages: 25-30

In [13], the issues of machine learning for data analysis are considered. The authors consider procedures that automate the data analysis process. Moreover, such data can be obtained in real time from the world around us.

The authors of the study [14] analyze in detail various methods that allow for qualitative analysis. As a result, it allows for a better understanding of the phenomenon or process that is being investigated. However, at the same time, the authors also note the need to apply methods of descriptive statistical analysis.

M. Salmona, E. Lieber and D. Kaczynski emphasize mixed methods of analysis [15]. The authors consider both quantitative and qualitative methods of analysis. At the same time, the authors talk about the possibility and necessity of combining different methods for data analysis.

We see that various methods of statistical analysis are used to analyze data. Moreover, such data refer to different scientific areas, spheres of activity, branches of science and production. Various methods and approaches are also used for data analysis. Recently, among such methods and approaches, the use of the ideology of wavelets is widespread. This ideology has found application both in technical applications [15]-[17] and in applications related to medicine [18] and economics [19]-[21].

2.2 Wavelet ideology as an analysis tool

The ideology of wavelets is based on the use of special functions – wavelet functions. These functions allow you to make a wavelet transform on the data that is being examined. Such data, as a rule, are presented in the form of a time series or a sequence of data that describe some pattern.

A wavelet function is a function that describes oscillation and scaling [22]. Then, if there is a sharp change in values in the data being examined, the wavelet function can detect this. This can be estimated by the value of the first derivative [22], [23]. In this case, the first derivative tends to zero. Smooth transitions will have small derivative values.

The basis of the formal characteristics of the wavelet ideology for continuous wavelet transform is the use of two functions [22]-[24]:

– wavelet – $\phi(t)$ functions with zero integral value

$$\int_{-\infty}^{+\infty} \phi(t)dt = 0\,, \tag{1}$$

that determines signal details and generates the detail coefficients;

– scaling functions $\varphi(t)$ with a single integral value

$$\int_{-\infty}^{+\infty} \varphi(t)dt = 0\,, \tag{2}$$

that determines rough approximation of signal and generates approximation coefficients.

Then we can describe the wavelet transform as follows [23], [24]:

$$W(g(t)) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} g(t)\phi\left(\frac{t-b}{a}\right)dt\,, \tag{3}$$

where

$\phi(\frac{t-b}{a})$ – is a mother wavelet satisfying condition (1);

a , b – scale and center of time localization, that determine the scale and $\varphi(t)$ function offset in accordance with the scaling conditions (2);

$g(t)$ – is a dataset or time series that is being explored, t – determines the order of the function data $g(t)$.

If we have two time series ($g1(t)$, $g2(t)$), then we can investigate the mutual influence of such data. For this, wavelet coherence is used. Wavelet coherence is analogous to multiple correlations. For analysis, you can use the following expression [25], [26]:

$$R^2(a,b) = \frac{\left|\Omega(a^{-1}W_{g1g2}(a,b))\right|^2}{\Omega(a^{-1}\left|W_{g1}(a,b)\right|^2)\Omega(a^{-1}\left|W_{g2}(a,b)\right|^2)}\,, \tag{4}$$

where:

$W(a,b)$ – values of cross wavelet spectra;

Ω – is a smoothing operator;

$R^2(a,b)$ – the squared wavelet coherency coefficient. $0 \le R^2(a,b) \le 1$. If these values tend to zero, then we have a weak correlation. Otherwise, we have a strong correlation [25], [26].

3. Several Examples of the Application of the Ideology of Wavelets for Data Analysis

One of the indicators of time series analysis is the Hurst exponent. The Hurst exponent is a measure of the self-similarity of a time series. The Hurst exponent also reflects the complexity of the dynamics of the time series, the properties of its correlation structure.

In order to estimate the Hurst exponent, one can use the apparatus of the wavelet theory. In general, you can write [24]:

$$E_j \approx 2^{(2H-1)j}\,, \tag{5}$$

where

H – is the Hurst exponent,

E_j – the energy value at a given level of the wavelet expansion with a given number (N_j) of detailing wavelet coefficients ($d_{g(t)}$).

Then the general formula for estimating the Hurst exponent is as follows:

$$\log_2 E_j = F(d_{g(t)}) \approx (2H-1)j + \text{const}\,. \qquad (6)$$

Or

$$\begin{aligned} &(2H-1)j + \text{const} = \log_2 E_j, \\ &2Hj - j = \log_2 E_j - \text{const}, \qquad (7) \\ &H = \frac{\log_2 E_j - \text{const} + j}{2j} = \frac{\log_2 E_j - \text{const}}{2j} + \frac{1}{2}. \end{aligned}$$

Then, by scanning the initial series ($g(t)$) in accordance with formula (7), it is possible to obtain the dynamics of the change in the Hurst exponent for such a series of data.

In fig. 1 shows a row, let's call it "Fishing 1".

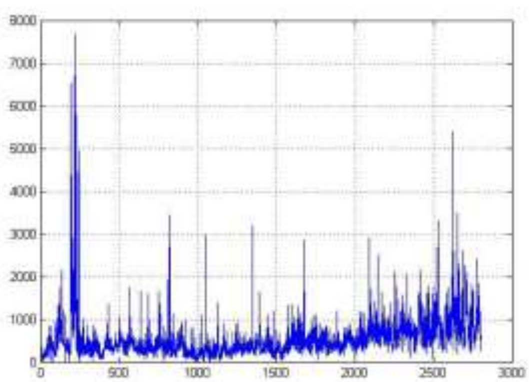

Figure 1. Some row of "Fishing 1"

In fig. 2 shows a row, let's call it "Fishing 2".

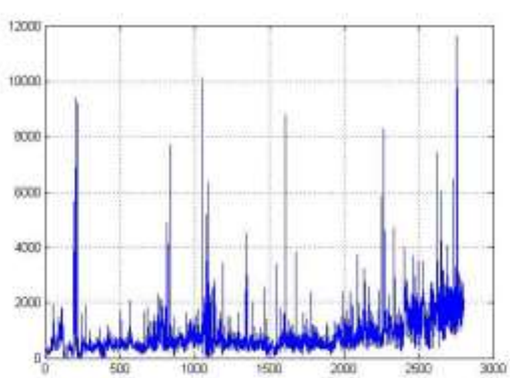

Figure 2. Some row of "Fishing 2"

In fig. 3 shows the dynamics of the Hurst exponent for the data in fig. 1 and fig. 2.

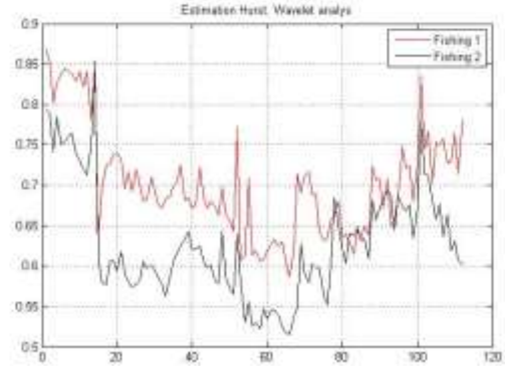

Figure 3. Dynamics of the Hurst exponent

We see a similar, but different dynamics of the Hurst exponent for the data in fig. 1 and fig. 2. This allows us to speak about a different correlation structure of the initial data, as well as about the difference in the dynamics of the studied data series. It is important to know this when analyzing the self-similarity of series, in studying their properties, and predicting the development of such series.

We can also use the ideology of wavelets in image processing. For this we will use formulas (1-3). The corresponding analysis is carried out for all lines of the original image, then for all columns of the image. Further, the results are combined. As a result, we get a processed image, where you can see the outlines of the selected objects. By setting different parameters for combining the results of processing the original image in rows and columns, we can get different levels of detail for image processing. This makes it possible to increase the efficiency of the analysis of the original image, to increase the accuracy of identifying objects in the image. Below are some of the results of such image processing, where the parameter of the wavelet transform scale (a) and the parameters for the key rows (ST) and columns (SC) play an important role in obtaining the final image.

In fig. 4 shows the original image of cytological preparations.

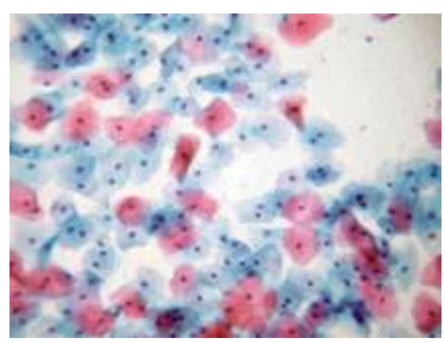

Figure 4. The original image of cytological preparations

In fig. 5-fig. 8 shows the results of the wavelet processing of the original image with different parameters.

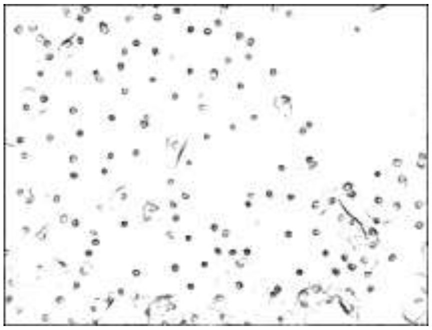

Figure 5. Results of the wavelet processing of the original image (a=20, ST=10, SC=12)

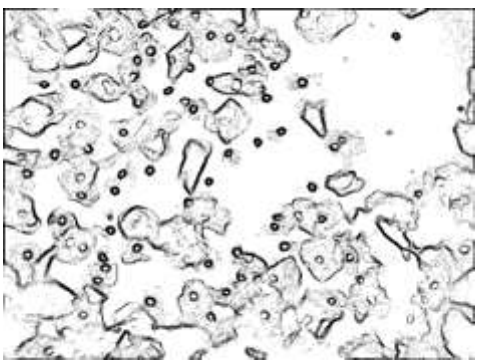

Figure 6. Results of the wavelet processing of the original image (a=50, ST=15, SC=15)

Figure 7. Results of the wavelet processing of the original image (a=50, ST=40, SC=45)

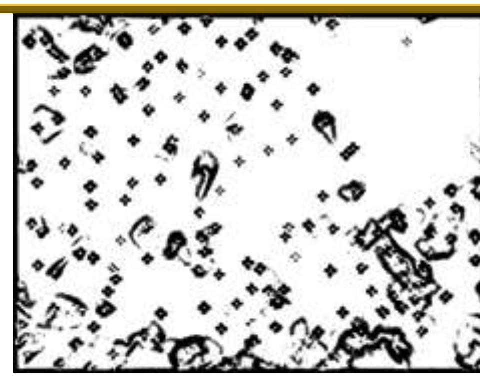

Figure 8. Results of the wavelet processing of the original image (a=30, ST=20, SC=25)

We see different results of processing the original image. These results are highly dependent on the choice of input parameters for the appropriate wavelet transform. We can only identify the nucleus of the cell or the cell as a whole. This allows you to apply different analysis algorithms and identify different objects in the image. This is an important point for the diagnosis of diseases.

Wavelet analysis also allows you to analyze the mutual dynamics of different time series. This can be done on the basis of a formalized dependence, which is represented by expression (4).

In fig. 9 shows a plot of some time series 1.

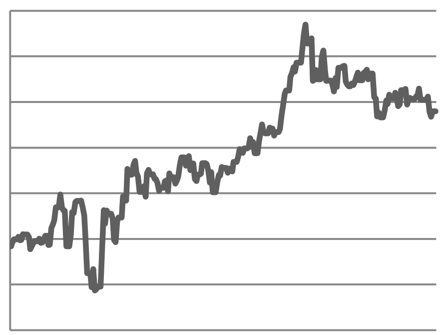

Figure 9. Time series plot 1.

In fig. 10 shows a plot of some time series 2.

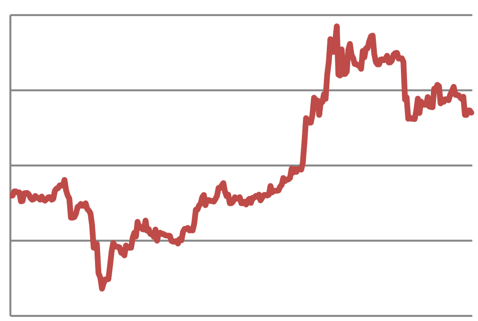

Figure 10. Time series plot 2.

In fig. 11 shows the wavelet coherence values between time series 1 data and time series 2 data.

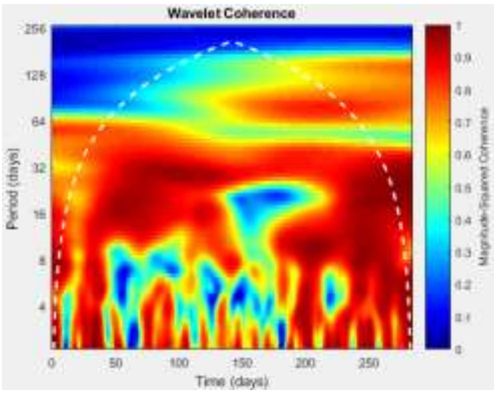

Figure 11. The value of the wavelet coherence between time series 1 and 2

We see a close relationship between time series 1 and 2 both for different time intervals and for different depths of assessing such a relationship. Thus, we can analyze in detail the relationship between the data we are exploring. This helps to make the right decisions and predict the future dynamics of the data that is being explored.

4. Conclusion

We have discussed the importance of data analysis for decision-making in any field of activity or research. The paper notes the possibility of using various methods and approaches for data analysis. Among such methods and approaches, methods of statistical analysis prevail. At the same time, it was noted that wavelet analysis is becoming the tool that has recently been widely used for various studies.

We have identified three areas of using wavelet ideology for data analysis. The paper provides a formalized description of such areas. The work also shows specific examples of the use of wavelet ideologists for analyzing different types of data. This allows us to say that the wavelet ideology is a universal tool for data analysis.

5. References

[1] Rajaonarivo, L., Courgeon, M., Maisel, E., & De Loor, P. (2017, March). Inline co-evolution between users and information presentation for data exploration. In Proceedings of the 22nd International Conference on Intelligent User Interfaces (pp. 215-219).

[2] In, J., & Lee, S. (2017). Statistical data presentation. Korean journal of anesthesiology, 70(3), 267.

[3] Lee, J., & Hwang, D. (2020). Single-cell multiomics: technologies and data analysis methods. Experimental & Molecular Medicine, 52(9), 1428-1442.

[4] Omarov, M., Tikhaya, T., & Lyashenko, V. (2019). Internet marketing metrics visualization methodology for related search queries. International Journal of Advanced Trends in Computer Science and Engineering, 8(5), 2277-2281.

[5] Maione, C., & Barbosa, R. M. (2019). Recent applications of multivariate data analysis methods in the authentication of rice and the most analyzed parameters: A review. Critical reviews in food science and nutrition, 59(12), 1868-1879.

[6] Mishra, P., Pandey, C. M., Singh, U., Gupta, A., Sahu, C., & Keshri, A. (2019). Descriptive statistics and normality tests for statistical data. Annals of cardiac anaesthesia, 22(1), 67-74.

[7] Mishra, P., Pandey, C. M., Singh, U., Keshri, A., & Sabaretnam, M. (2019). Selection of appropriate statistical methods for data analysis. Annals of cardiac anaesthesia, 22(3), 297-302.

[8] Górecki, T., Krzyśko, M., Waszak, Ł., & Wołyński, W. (2018). Selected statistical methods of data analysis for multivariate functional data. Statistical Papers, 59(1), 153-182.

[9] Washington, S., Karlaftis, M., Mannering, F., & Anastasopoulos, P. (2020). Statistical and econometric methods for transportation data analysis. Chapman and Hall/CRC.

[10] Yan, F., Robert, M., & Li, Y. (2017). Statistical methods and common problems in medical or biomedical science research. International journal of physiology, pathophysiology and pharmacology, 9(5), 157.

[11] Chambers, J. M., Cleveland, W. S., Kleiner, B., & Tukey, P. A. (2018). Graphical methods for data analysis. Chapman and Hall/CRC.

[12] Ho, J., Tumkaya, T., Aryal, S., Choi, H., & Claridge-Chang, A. (2019). Moving beyond P values: data analysis with estimation graphics. Nature methods, 16(7), 565-566.

[13] Mahdavinejad, M. S., & et al.. (2018). Machine learning for Internet of Things data analysis: A survey. Digital Communications and Networks, 4(3), 161-175.

[14] Assarroudi, A., & et al.. (2018). Directed qualitative content analysis: the description and elaboration of its underpinning methods and data analysis process. Journal of Research in Nursing, 23(1), 42-55.

[15] Ahmad, M. A., & et al.. (2016). The Methodology of Image Processing in the Study of the Properties of Fiber as a Reinforcing Agent in Polymer Compositions. International Journal of Advanced Research in Computer Science, 7(1), 15-18.

[16] Zheng, Q., & et al.. (2020). Using continous wavelet analysis for monitoring wheat yellow rust in different infestation stages based on unmanned aerial vehicle hyperspectral images. Applied Optics, 59(26), 8003-8013.

[17] Dadkhah, M., & et al.. (2019). Methodology of wavelet analysis in research of dynamics of phishing attacks.

International Journal of Advanced Intelligence Paradigms, 12(3-4), 220-238.

[18] Lyashenko, V., Mohammed, O. I. A., & Babker, A. Ma. (2020). COVID-19 wavelet coherence data for some Gulf countries, GSC Biological and Pharmaceutical Sciences, 11(2), 166-174.

[19] Baranova, V., & et al.. (2019, October). Wavelet Coherence as a Tool for Studying of Economic Dynamics in Infocommunication Systems. In 2019 IEEE International Scientific-Practical Conference Problems of Infocommunications, Science and Technology (PIC S&T) (pp. 336-340). IEEE.

[20] Lyashenko, V., Bilotserkivskyi, O., Pyvavar, I., & Kots, H. (2021). Mutual Dynamics of Certain Types of Bitcoin: Data from Wavelet Coherence. Journal of Engineering, Technology, and Applied Science, 3(2), 58-65.

[21] Choi, S. Y. (2020). Industry volatility and economic uncertainty due to the COVID-19 pandemic: Evidence from wavelet coherence analysis. Finance Research Letters, 37, 101783.

[22] Kingsbury, N. (1999). Image processing with complex wavelets. Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, 357(1760), 2543-2560.

[23] Heil, C.E., & Walnut, D.F. (1989). Continuous and discrete wavelet transforms. SIAM review, 31(4), 628-666

[24] Lyashenko, V., Deineko, Z., & Ahmad, A. (2015). Properties of wavelet coefficients of self-similar time series. International Journal of Scientific and Engineering Research, 6(1), 1492-1499.

[25] Torrence, C., & Webster, P. J. (1999). Interdecadal changes in the ENSO–monsoon system. Journal of climate, 12(8), 2679-2690.

[26] Vasiurenko, O., & et al.. (2020). Spatial-Temporal Analysis the Dynamics of Changes on the Foreign Exchange Market: an Empirical Estimates from Ukraine. Journal of Asian Multicultural Research for Economy and Management Study, 1(2), 1-6.