Comparative Analysis of Data Dynamics Based on Wavelet Coherence Using Higher-Order Moments

Yasser Al-Sharo
Faculty of Information Technology, Department of Cyber Security
Ajloun National University
Ajloun, Jordan
Yaser.Shrah@anu.edu.jo

Amer Abu-Jassar
Faculty of Information Technology, Department of Computer Science
Ajloun National University
Ajloun, Jordan
amer_abu_jassar@hotmail.com

Oleg Vasiurenko
Department of Cybersecurity, IT and Economics
Kyiv University of Intellectual Property and Law
Kyiv, Ukraine
vasurenko@i.ua

Vyacheslav Lyashenko
Department of Media Systems and Technology
Kharkiv National University of Radio Electronics
Kharkiv, Ukraine
viacheslav.liashenko@nure.ua

***Abstract*—The article discusses the issues of analyzing data presented in the form of time series and reflecting the dynamics of the development of a certain process, phenomenon or object. Particular attention is paid to conducting a comparative analysis of the dynamics of various data sets. For these purposes, a technique for estimating wavelet coherence was used. The approach also involves transforming the original data based on a sliding sampling window using higher order moments as an implementation of the memory effect. This makes it possible to assess the dynamics of the large data sets and increase the efficiency of their processing and analysis. Results for real data are presented, which make it possible to obtain an additional assessment of their mutual dynamics. The obtained results show the reliability and effectiveness of such analysis.**

Keywords—analysis, dynamics, comparison, wavelet coherence, moments of higher orders.

I. Introduction

Analysis is one of the stages in the research and study of various phenomena, objects, processes. At the same time, it is also a research tool that helps to obtain additional data and to transform it into knowledge [1]-[3]. This is fully consistent with the general concept of data science [4], [5].

Thus, a comprehensive analysis of available data helps to make the most effective and well-founded decisions. Ultimately, this affects the development or successful functioning of what is being studied. This interest in analysis is primarily due to the fact that any process, phenomenon or object is described by a certain set of data. In this case, it is necessary to consider various sources of obtaining such data: technical applications, economic dynamics, data on natural and man-made disasters, etc. [6], [7].

Primary data can be presented in the form of some structured set: tables, databases, graphs, etc. However, the most common presentation of data is its dynamic structuring in the form of some time series [2], [5], [7]. This presentation of data allows them to be analyzed and evaluated both from the point of view of the general time interval and at its individual intervals. It also simplifies the comparative analysis of data obtained either from different sources (i.e., from different technical devices or processes of economic dynamics), or from the same source (for different indicators from one technical device or some economic dynamics). Considering the time structure of data, it is essential to take into account the dynamics of their change, which is crucial for conducting comparative analysis [6]. In this case, it is advisable to talk about big data analysis.

Various methods and approaches can be used to analyze data. Among such works, it is necessary to mention, for example, the study by H. Wickham and H. Wickham, who examined various methods of analysis and data processing systems [8]. In this case, special attention is paid to statistical analysis as one of the leading research tools. Introduction to data mining is widely discussed in the work of such authors as M. R. Berthold and D. J. Hand [9]. At the same time, the study [10] is devoted to the consideration of issues of comparative analysis using the example of collisions of heavy ions. This study uses descriptive statistics and inferential statistical methods. In "A comparative study between physics, electrical and data driven lithium-ion battery voltage modeling approaches" [11] the authors draw attention to the comparison of physical and electrical characteristics when modeling the voltage of lithium-ion batteries. J. Mattke, C. Maier, T. Weitzel, J. E. Gerow and J. B. Thatcher conduct a comparative analysis in information systems research in order to select the most effective ones for further work [12]. Herein both quantitative and qualitative characteristics of the systems that are compared are used. In their study, D. Jelonek, N. H. Tien, M. T. H. Dao and D. T. Minh, based on various data sets, conduct a comparative analysis of the business strategies of Vietnamese developers using the Hoffer matrix [13].

Despite such variety of analysis and tools areas (methods, theories, algorithms) that allow to conduct relevant research, the expansion of data volumes and their significant variability require the development of new approaches. A particularly important point is the development of new approaches when it comes to mutual analysis of data dynamics, when it is necessary to obtain additional data at each interval from the studied interval [6].

Thus, the main purpose of this paper is to consider the approach to conducting a comparative analysis of data that can be obtained from various sources and describe any processes, phenomena, objects. In other words, a new generalized approach to comparative data analysis should be considered, which describes the study over time and can be presented in the form of a time series. For its implementation, it is necessary to consider the possibility of using wavelet coherence estimates and transforming primary data to increase the efficiency of their processing and to obtain additional information. The following are specific examples that confirm the feasibility of such processing of primary data.

II. General Preliminary Questions for Comparative Analysis of Data Dynamics

A. *Wavelet Coherence as a Comparative Analysis Tool*

One of the tools for conducting comparative analysis is the use of wavelet coherence estimates, the possibility of constructing which is based on wavelet ideologists. The method for constructing wavelet coherence estimates allows for comparison of various data presented in the form of time series. In this case, it is advisable to consider the dynamics of data from the point of view of different time intervals, and this does not violate the generalization of such dynamics as a whole. This is the main difference between wavelet coherence estimates and classical methods of statistical analysis [14], [15].

To construct such estimates, the following generalized formula is used, taking into account two series of data ($f(t)$ and $m(t)$), reflecting the change in a particular indicator over time [16]-[18]:

$$R^2(a,b) = \frac{\left| \Lambda(a^{-1} W_{f(t)m(t)}(a,b)) \right|^2}{\Lambda(a^{-1} \left| W_{f(t)}(a,b) \right|^2) \Lambda(a^{-1} \left| W_{m(t)}(a,b) \right|^2)}, \quad (1)$$

where:

$W(a,b)$ – transverse wavelet spectra,

a,b – scale and center of time localization that determine the scale of the wavelet transform,

Λ – smoothing operator,

$R^2(a,b)$ – wavelet coherence coefficient. $0 \le R^2 \le 1$. If these values tend to zero, then a weak correlation is observed. Otherwise, there is a strong correlation [16]-[18].

Initial data $f(t)$ and $m(t)$, as a rule, represent some primary indicators of the process, phenomenon or object being studied. Therefore, the value of wavelet coherence estimates according to formula (1) can be called primary estimates. However, there are no restrictions, from the point of view of the general ideology of wavelets, regarding the presentation of the original data. In this case, the important point is the substantiation of such a representation.

B. *Transformation of Primary Data to Construct Wavelet Coherence Estimates*

The main idea of the corresponding transformation is to implement the memory effect that is inherent in time series data [19], [20]. Its essence lies in the fact that subsequent data values do not occur spontaneously, but change considering previous values. As a result, these changes should be taken into account when generating some transformation of the original time series.

Such transformation is possible due to a sliding window of sampling from the original data series, the formation of the length of such a window and the lag of its sliding according to the original data. A schematic explanation of this transformation is presented in Fig. 1.

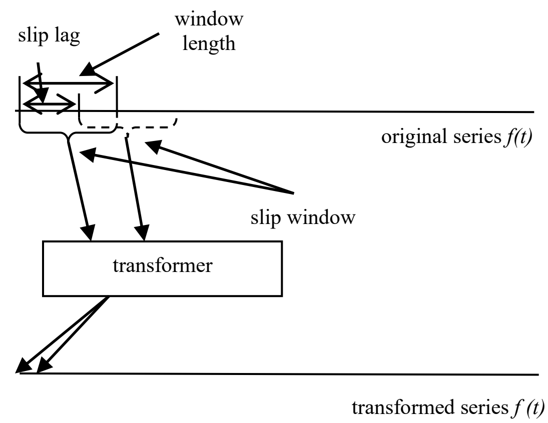

Fig. 1: Schematic representation of the transformation of the original time series using a sliding window

Thus, in this case, a specific set of data in the selected window is considered. This makes it possible to enhance the memory effect of a time series in the case of unforeseen failures (during analyzing data on technical devices) or the occurrence of non-standard situations (in the economy, in the case of disasters). Then this effect should be modeled both by the size of the sliding window and by changing the delay value for such a window. This is where additional information appears about the relationship between the data being studied.

To enhance the memory effect of a time series, it is proposed to consider transforming its original data using moments of higher orders, where skewness and kurtosis are highlighted. In the first case, the skewness of the data distribution is taken into account, in the second a measure of the sharpness of the peak of such a distribution is considered. Thus, the shape of the data distribution is included. In general, the above mentioned helps to better account for the reciprocity between the data series that are being examined.

It should be noted that formula (1) for estimating wavelet coherence does not change, but only the series of original data changes to the transformed series.

III. Results and Discussion

A. Original Time Series Data and their Transformation

As it was mentioned earlier, the proposed approach to comparative analysis of data dynamics can be considered for any set of indicators – from technical devices to economic data. Furthermore, time series are selected to describe some of the dynamics of economic data. In particular, these are some types of financial futures. These data series will not be specified, but are proveded as an example for comparing such data based on wavelet coherence using moments of higher orders (skewness and kurtosis).

Fig. 2 presents two series of data (series 1 and series 2), which describe the dynamics of quotes for two types of financial futures.

Note, that the relatively similar dynamics between such data series, where the correlation coefficient is 0.97. But local changes in data dynamics should also be emphasized.

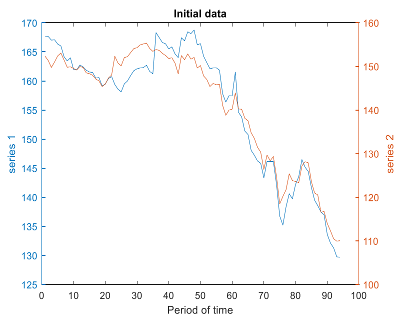

Fig. 2: Dynamics of primary data (series 1 and series 2)

It is these local changes that create special differences in the comparative aspect of the dynamics of such data series. Therefore, it is important to emphasize these differences and highlight them against the general background of data changes. For these purposes, it is proposed to use moments of higher orders, which are formed from primary data based on a sliding window.

Fig. 3 shows transformed data series (for series 1 and series 2), which reflect the skewness with a sliding window of 8 samples and a lag of one sample.

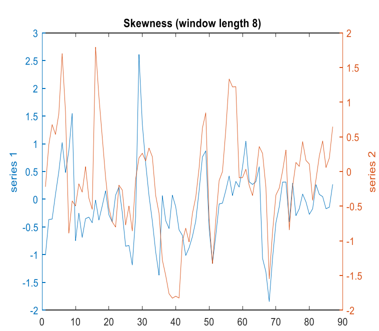

Fig. 3: Transformed data (skewness) for the original series 1 and series 2 with a sliding window of 8 samples and a lag of 1 sample

Herein, significant differences between the data series under study are already visible. Such differences are manifested both in the dynamics and in the variability of data changes. This is an important aspect in carrying out further comparative analysis.

The figure below shows the transformed data displays kurtosis with an 8-sample sliding window and a one-sample lag for the same pair of original run 1 and run 2 data.

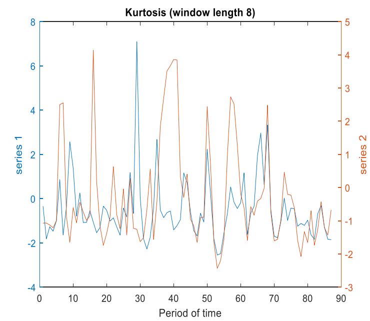

Fig. 4: Transformed data (kurtosis) for the original series 1 and series 2 with a sliding window of 8 samples and a lag of 1 sample

This figure also highlights the significant differences in the dynamics of the data series examined in our research paper. Moreover, such differences are distinctive from those presented in Fig. 3. This allows a detailed comparison of the dynamics of the two data series to be conducted.

B. Comparative Dynamics of the Studied Data Based on Wavelet Coherence Estimates

First of all, it is crucial to consider wavelet coherence estimates in order to compare the mutual dynamics of the initial data between series 1 and series 2, which are presented in Fig. 2.

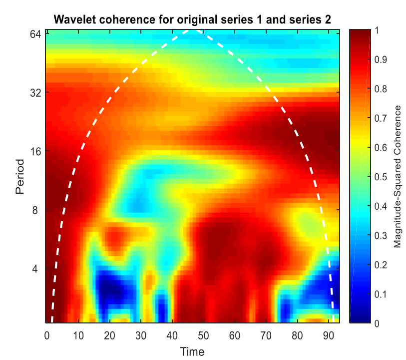

Fig. 5: Wavelet coherence between original data series 1 and series 2

It can be seen that series 1 and 2 are the most consistent at the beginning and after the first half of the time period under study. This consistency is also the most significant during this period (consistency scores are highlighted in red). Also noteworthy is this consistency in the short and long term(s).

This helps to better understand the dynamics of the data being studied and, in particular, to make effective decisions about entering the securities market.

However, the data in Fig. 3 and Fig. 4 show the presence of certain differences in the dynamics of the primary data. Therefore, consideration of wavelet coherence estimates occurs from the point of view of transforming such data using higher order moments.

Fig. 6 shows the results of wavelet coherence estimates for data that are transformed using a sliding window of 8 samples with a lag of 1 sample based on the skewness indicator (these data were displayed in Fig. 3). Fig. 7 displays similar data, but for the kurtosis indicator (these data were displayed in Fig. 4).

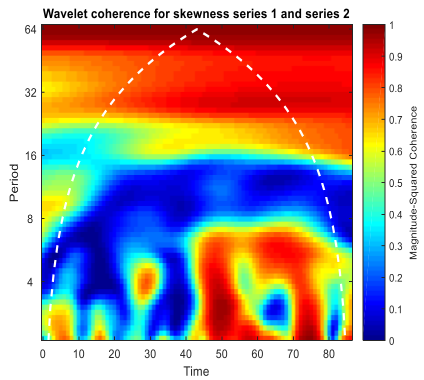

Fig. 6: Wavelet coherence for the data shown in Fig. 3 (conversion method – skewness)

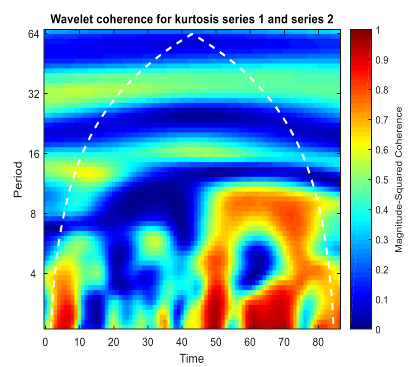

Fig. 7: Wavelet coherence for the data shown in Fig. 4 (conversion method – kurtosis)

It can be seen that the dynamics of wavelet coherence estimates presented in Fig. 6 and Fig. 7, in general, inherits the dynamics of estimates, which is reflected in Fig. 5. Moreover, the data in Fig. 6 and fig. 7 clarify the primary estimates, since the shape of the distribution of the values of the primary data of series 1 and series 2 is taken into account. This allows for more accurate forecasts and more informed decisions.

For comparison, wavelet coherence estimates made for skewness and kurtosis measures are also considered, but for a sliding window of 12 samples and with a delay of 1 sample. The corresponding presentation of estimates is shown in Fig. 8 and Fig. 9.

It can be said that the presented estimates are within the range of previous data and do not contradict them. In this case, the primary estimates are refined and the possibilities for varying the memory effect of time series are expanded.

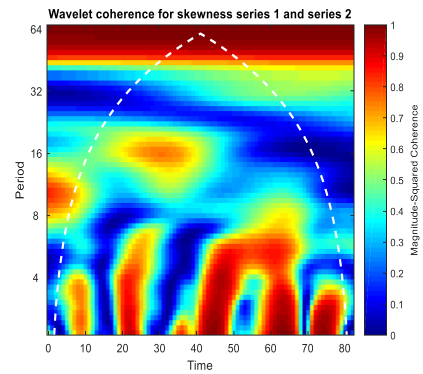

Fig. 8: Wavelet coherence for a sliding window of 12 samples and a lag of 1 sample for the original data series 1 and series 2 (conversion method – skewness)

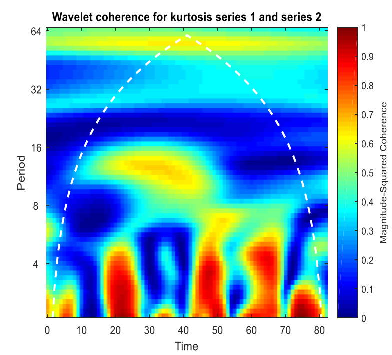

Fig. 9: Wavelet coherence for a sliding window of 12 samples and a lag of 1 sample for the original data series 1 and series 2 (conversion method – kurtosis)

It should be noted that the change in the size of the sliding window and its sliding lag is completely determined by the analysis task and the size of the initial sample. Although the size of such a window also correspondents with the accuracy of determining the shape of the distribution of primary data over the period under study. The approach discussed above allows a more accurate assessment of the mutual dynamics of data in general, taking into account the memory effect of time series.

IV. Conclusion

The paper discusses the issues of conducting a comparative analysis of the studied data dynamics. Based on a brief review of literary sources, the relevance of this topic is shown, regardless of the object of research. In this case, special attention is paid to data presented in the form of time series, which allows the appropriate analysis both between individual series and with even different fragments of the same series. As a methodology for conducting comparative data analysis, the approach of obtaining wavelet coherence estimates is considered. To enhance the memory effect inherent in time series it is proposed to transform primary data using moments of higher orders through updating a sliding window. Specific examples, provided in this research paper, show the effectiveness of the proposed approach, the feasibility of its use to enhance the memory effect of a time series, and clarify the results of primary estimates.

Among the conditions for using this approach, one should indicate a sufficient length of the original data series. This length is determined by the ability to use a sliding window to obtain the converted data. This should be determined individually in each specific case. At the same time, it is also necessary to pay attention to the possibility of using other methods of converting primary data. Such transformation will be determined by the conditions of comparison and the feasibility of obtaining additional information in new conditions.

REFERENCES

[1] Y. X., Yu, H. P., Gong, H. C., Liu, and X., Mou, "Knowledge representation and reasoning using fuzzy Petri nets: a literature review and bibliometric analysis," Artificial Intelligence Review, vol. 56, no. 7, pp. 6241-6265, 2023, doi:10.1007/s10462-022-10312-3.

[2] V., Savanevych, S., Khlamov, O., Briukhovetskyi, T., Trunova, and I., Tabakova, "Mathematical Methods for an Accurate Navigation of the Robotic Telescopes," Mathematics, vol. 11, no. 10, pp. 2246, 2023, doi: 10.3390/math11102246.

[3] M., Dadkhah, T., Maliszewski, and V. V., Lyashenko, "An approach for preventing the indexing of hijacked journal articles in scientific databases," Behaviour & Information Technology, vol. 35, no. 4, pp. 298-303, 2016, doi:10.1080/0144929X.2015.1128975.

[4] H., Han, and S., Trimi, "Towards a data science platform for improving SME collaboration through Industry 4.0 technologies," Technological Forecasting and Social Change, vol. 174, pp. 121242, 2022, doi: 10.1016/j.techfore.2021.121242.

[5] I. H., Sarker, A. S. M., Kayes, S., Badsha, H., Alqahtani, P., Watters, and A., Ng, "Cybersecurity data science: an overview from machine learning perspective," Journal of Big data, vol. 7, pp. 1-29, 2020, doi: 10.1186/s40537-020-00318-5.

[6] O., Kuzomin, V., Lyashenko, M., Tkachenko, M. A., Ahmad, and H. Kots, "Preventing of technogenic risks in the functioning of an industrial enterprise," International Journal of Civil Engineering and Technology, vol. 7, no. 3, pp. 262-270, 2016.

[7] S., Khlamov, O., Briukhovetskyi, V., Savanevych, T., Trunova, and I., Tabakova, "Astronomical Big Data Analysis by the CoLiTec Software," In 2022 IEEE 9th International Conference on Problems of Infocommunications, Science and Technology (PIC S&T), 2022, pp. 391-394, doi:10.1109/PICST57299.2022.10238543.

[8] H. Wickham, "Data analysis," Springer International Publishing, pp. 189-201,2016, doi:10.1007/978-3-319-24277-4_9.

[9] M. R., Berthold, and D. J., Hand, "Intelligent data analysis: an introduction," Springer, 2007.

[10] H., Wolter, and et al., "Transport model comparison studies of intermediate-energy heavy-ion collisions," Progress in Particle and Nuclear Physics, vol. 125, pp. 103962, 2022, doi: 10.1016/j.ppnp.2022.103962.

[11] Y., Liang, and et al., "A comparative study between physics, electrical and data driven lithium-ion battery voltage modeling approaches," SAE Technical Paper, No. 2022-01-0700, 2022, doi:10.4271/2022-01-0700.

[12] J., Mattke, C., Maier, T., Weitzel, J. E., Gerow, and J. B. Thatcher, "Qualitative comparative analysis (QCA) in information systems research: status quo, guidelines, and future directions," Communications of the association for information systems, vol. 50, no. 1, pp. 8, 2022, doi: 10.17705/1CAIS.05008.

[13] D., Jelonek, N. H, Tien, M. T. H., Dao, and D. T. Minh, "Comparative analysis of business strategy of Vietnamese real estate developers: the use of Hoffer matrix," International journal of multidisciplinary research and growth evaluation, vol. 3, no. 1, pp. 197-204, 2022.

[14] A. T., Abu-Jassar, Y. M., Al-Sharo, V., Lyashenko, and S., Sotnik, "Some Features of Classifiers Implementation for Object Recognition in Specialized Computer systems," TEM Journal: Technology, Education, Management, Informatics, vol. 10, no. 4, pp. 1645-1654, 2021, doi: 10.18421/TEM104-21.

[15] A. T., Abu-Jassar, H., Attar, V., Lyashenko, A., Amer, S., Sotnik, and A., Solyman, "Access control to robotic systems based on biometric: the generalized model and its practical implementation," International Journal of Intelligent Engineering and Systems, vol. 16, no. 5, pp. 313-328, 2023, doi: 10.22266/ijies2023.1031.27.

[16] C., Torrence, and P. J., Webster, "Interdecadal changes in the ENSO–monsoon system," Journal of climate, vol. 12, no. 8, pp. 2679-2690, 1999, doi:10.1175/1520-0442(1999)012<2679:ICITEM>2.0.CO;2.

[17] C. E., Heil, and D. F., Walnut, "Continuous and discrete wavelet transforms," SIAM review, vol. 31, no. 4, pp. 628-666, 1989, doi: 10.1137/1031129.

[18] V., Baranova, O., Zeleniy, Z., Deineko, G., Bielcheva, and V., Lyashenko, "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), 2019, pp. 336-340, doi: 10.1109/PICST47496.2019.9061301.

[19] T., Wand, M., Heßler,and O., Kamps, "Memory Effects, Multiple Time Scales and Local Stability in Langevin Models of the S&P500 Market Correlation," Entropy, vol. 25, no. 9, pp. 1257, 2023, doi: 10.3390/e25091257.

[20] S., Kim, and C., Eom, "Long-term memory and volatility clustering in high-frequency price changes," Physica A: Statistical Mechanics and its Applications, vol. 387, no. 5-6, pp. 1247-1254, 2008, doi: 10.1016/j.physa.2007.08.061.