ISSN 2228-9860 eISSN 1906-9642 CODEN: ITJEA8

International Transaction Journal of Engineering, Management, & Applied Sciences & Technologies

http://TuEngr.com

Dynamics Data Prediction based on Time Series Decomposition: The Case of Tourist Flow Data

Alfira Kumratova[1*], Elena Popova[1], Sergey Kurnosov[2], Valery Kondratiev[1], Natalia Kurnosova[1]

[1] *Department of Information Systems, Kuban State Agrarian University named after I.T. Trubilin, Krasnodar, RUSSIA.*

[2] *Department of Computer Technologies and Systems, Kuban State Agrarian University named after I.T. Trubilin, Krasnodar, RUSSIA.*

**Corresponding Author (Email: alfa05@yandex.ru).*

Paper ID: 12A8P

Volume 12 Issue 8

Received 19 April 2021
Received in revised form 02 June 2021
Accepted 08 June 2021
Available online15 June 2021

Keywords:
Tourism economics; Tourist management; Phase analysis; Decomposed time series; R/S-analysis; Tourist forecast; Linear cellular automaton; Russian tourism.

Abstract

This article adopts a complex methodology for predicting the dynamics of the decomposition time series of the tourist flow. Its peculiarities lie in the joint use of both classical and new "nonlinear" statistics. The proposed and tested methods are presented in the form of a pre-forecast and forecast model for assessing the trend stability of time series of the tourist flow and obtaining a forecast. The methods of nonlinear dynamics have been tested including the Hurst normalized range method, phase analysis, and a linear cellular automaton. The results of the analysis and forecast on real data of the tourist flow are presented in the form of the values of the lower level of modeling of tourist and recreational activities, which are input data for the models of the upper level, the level of management of tourist and recreational activities. A quantitative forecast of the magnitude of the tourist flow allows solving the issues of managing tourist and recreational activities, such as planning the employment of rooms.

Disciplinary: Computer Technologies & Information Systems, Applied Mathematics; Hospitality and Tourism Management.

Cite This Article:

Kumratova, A., Popova, E., Kurnosov, S., Kondratiev, V., Kurnosova, N. (2021). Dynamics Data Prediction based on Time Series Decomposition: The Case of Tourist Flow Data. *International Transaction Journal of Engineering, Management, & Applied Sciences & Technologies, 12*(8), 12A8P, 1-10. http://TUENGR.COM/V12/12A8P.pdf DOI: 10.14456/ITJEMAST.2021.163

1 Introduction

The relevance of this study is due to the government's need to create favorable conditions for the dynamic development of the tourism market. By government decree No. 434 of 04/06/2020, tourism has become one of the sectors of the Russian economy most affected by the coronavirus.

The Federal Agency for Tourism website of the Russian Federation presents current and adopted support measures for small and medium-sized businesses in the tourism industry.

In the context of the epidemiological situation in the world, much attention is paid to the development of domestic tourism, in connection with which a quantitative forecast of the size of the tourist flow will make it possible to control one of the main points - the issue of managing tourist and recreational activities, for example, in planning the occupancy of the room stock as a single hotel, and the tourist and recreational complex of the region as a whole.

In the current reality, the tourist and recreational sphere of each region of Russia is strengthening its positions in the domestic market, which affects the socio-economic development of both the region itself and the country as a whole. Furthermore, the announced hard lockdown, fixed record incidence rates in Europe for March 2021, and booking statistics suggest that today the main population of Russia is planning a vacation within the country and even within the region of residence. It is this fact that, in practice, will enable regional tourist and recreational complexes to demonstrate their capabilities in providing quality services, improve the investment climate in the tourism sector, increase the number of communal accommodation facilities, and develop a range of medical, recreational, business and professional services. All of the above contributes to the development of new directions in the tourism industry of each region individually and of the country as a whole, which will determine the problems of the development of the regional tourism industry, allowing to highlight the priority areas of sustainable development of Russian tourism.

Thus, defining the role and place of the tourism industry in the regional economy in the realities of today, the authors note a significant strengthening of the position and influence both on the socio-economic development of the region and on the structure of its economy. In 2020, the Russian Government allocated 15 billion rubles to stimulate and develop the dynamics of domestic tourist flows. In 2021, Rosturizm allocated 2 billion rubles. For the resumption of the program to promote domestic tourism with the possibility of receiving cashback to purchase tours in Russia. According to Doguzova [5], within the first stage of the program was used by 60 thousand Russian citizens, spending more than 1.4 billion rubles on tours and accommodation. The most popular destinations within the cashback program are Krasnodar Territory, Stavropol Territory, and Crimea. The flow of tourists in the winter period also increased significantly to Karachay-Cherkessia, Kabardino-Balkaria - ski resorts. The ski season in these regions is still going on. It is pretty long, including the May holidays. On the May holidays, the most significant internal (Russian) tourist flow is expected to these regions, the intensity of which is one of the indicators of the state and assessment of the tourism market.

An assessment of the general prospects for the development of tourism for any region, as well as for the country and the world as a whole from the standpoint of economic efficiency and social effect, is based on the totality of the listed indicators, including the hand - the tourist flow [10].

The forecasting methods proposed by the authors make it possible to quantify the intensity of the tourist flow as a complex indicator reflecting the level of the tourist market in a region or an individual object (hotel complex, sanatorium, tourist base, etc.).

2 Materials and methods

Using the analysis, generalized conclusions, and description of the results, the researcher forms "structured" information about the time series. This structured knowledge allows the analyst to develop sound predictive analysis methods [9, 11]. The article uses the values of the decomposed daily time series of the tourist flow to the Dombay ski village by days of the week from 2015–2017 as a calculation base. Note that the quantitative data on the volume of the tourist flow is a primary indicator of the tourist and recreational activities of a country, region, and individual organization. Therefore, the authors investigated: the initial time series (TS) of the daily tourist flow and decomposition TS (by days of the week, separately TS of weekends (Saturday, Sunday), weekdays). The practical importance of studying separately TS of weekends is because the most significant part of tourists is the neighboring regions for the Karachay-Cherkess Republic: Stavropol Territory, Rostov Region, Krasnodar Territory.

Separately, we note that for conclusions about the quantitative measurement of the dynamics of a process of any nature, the study of averaged (typical) values is not practical [8, 11]. This is because the accuracy of the quantitative prediction determines the sequence of data that allows you to identify the possibility of the next value in time.

In classical statistics, we distinguish three leading indicators for the temporal sequence of data: kurtosis, asymmetry, and variation. It is believed that these three coefficients act as a triad for assessing the trend stability of the process [7, 8, 11, 14].

The study of the time series of the tourist flow of different decompositions is presented in Table 1.

Table 1: Statistical indicators for decomposed time series of tourist flows

Stat. Indicators / Days of the week	MX	DX	σ	V	A	E	Median
Monday	63.16	2082.4	45.63	0.72	2.2	**6.4**	49.5
Tuesday	61.4	2154.2	46.4	0.75	2.03	**5.6**	48
Wednesday	57.4	1184.5	34.4	0.6	1.8	**5.21**	48.5
Thursday	91.8	5098.5	71.4	0.77	1.6	2.28	73
Friday	92.57	6962.9	83.4	0.9	2.24	**5.45**	69
Saturday	136	10116.2	100.6	0.74	1.45	2.43	108
Sunday	207.3	30179.8	173.7	0.84	1.33	1.39	151
Weekdays	370.8	29158.8	170.76	0.46	1.18	1.63	333
Weekend	344.7	43774	209.2	0.61	0.91	0.25	291.5
Original TS "Tourists"	1021.1	11195.5	105.81	1.03	2.74	**9.65**	67

Based on the visualization of the graphical representation of the initial data of decomposition TS by days of the week, the authors proposed a mechanism for carrying out the data

normalization procedure for the further forecasting procedure: isolated cases of an outlier in the time series "Monday," "Thursday," "Friday" were removed. It can be noted that outliers correspond to single event components and create a "blurry" picture (dynamics).

Nobel laureate G. Markovitz defined two leading risk indicators: mathematical expectation and variance or standard deviation. In later works, asymmetry coefficients were also described as indicators of the risk measure $A = \sigma^{-3} \sum_{s=1}^{n} (W_s - M)^3 P_s$ and excess $E = \sigma^{-4} \sum_{s=1}^{n} (W_s - M)^4 P_s$, where P_s - this is the probability (relative frequency) of the value of the random variable W_s, $1 \le s \le n$.

Analysis of the calculated risk indicators (by Table 1) allows us to draw several conclusions:

- for decomposition TS weekdays, on average, tourists are expected to arrive in 2-3 excursion buses. A sharp increase in the tourist flow is observed on weekends, which is confirmed by practice;

- with relatively equal values of the mathematical expectation for TS "Thursday" and "Friday," there is a significant difference in the variance index. This is a consequence of the difference in the form of their distribution;

- coefficient of variation for all investigated TS V> 0.33. The latter means that the degree of data dispersion is considered significant;

- for all considered TS, the asymmetry index is higher than 0.5, i.e., there is significant asymmetry in data distribution;

- highlighted in the table kurtosis index values exceeding 3, characterizing the "normality" of the distribution, for normalized TS "Monday" and "Friday," as well as for TS "Tuesday" and "Wednesday," confirm the fact that their probability density functions do not obey Gauss's law;

- the advantage of the median is that outliers do not influence it. A significant range in the values of the mathematical expectation and the median confirms that the studied TS did not obey the customary distribution law.

3 Result and Discussion

Given the presence of "heavy tails" in the original TS and decomposition TS ("Monday," "Tuesday," "Wednesday," "Friday"), it is proposed to use the apparatus of nonlinear dynamics methods to identify pre-forecast characteristics and calculate predicted values [2, 4, 9, 12], which has proven itself when working not only with long (large) samples but also with short TS [7, 11].

A detailed description of the algorithm for the operation of nonlinear dynamics methods: R / S-analysis, phase analysis, the linear cellular automaton is presented in the sources [5, 6, 7, 11, 13].

Table 2 shows the calculated indicators obtained based on nonlinear dynamics methods for the initial and decomposed time series of tourist flows.

Table 2: Results of the calculated indicators in the study of decomposed time series of the tourist flow

Indicators / Days of the week	Forecast analysis				
	R / S analysis			Phase Analysis	
	Breakpoint	Regression equation	Hurst exponent	The length of the quasi-cycles with the highest frequency	Frequency of length of quasi-cycles
Monday	5	$y = 0.64 \cdot x - 0.42$	0.64	5	10
Tuesday	5	$y = 0.67 \cdot x - 0.46$	0.67	5	7
Wednesday	7	$y = 0.62 \cdot x - 0.4$	0.62	5	7
Thursday	6	$y = 0.65 \cdot x - 0.43$	0.65	4; 5	8
Friday	6	$y = 0.65 \cdot x - 0.43$	0.65	5	9
Saturday	5	$y = 0.65 \cdot x - 0.43$	0.65	4	9
Sunday	5	$y = 0.65 \cdot x - 0.42$	0.65	5	10
Weekdays		$y = 0.64 \cdot x - 0.42$		4	7
Weekends	5	$y = 0.61 \cdot x - 0.39$	0.61	5	10
Original TS "Tourists"	6	$y = 0.63 \cdot x - 0.41$	0.63	5	46

Analysis of the calculated data in Table 2 allows us to draw the following conclusions:

For each studied time series, the fifth or sixth point of separation from the R / S-trajectory is mainly fixed. On average, the duration of 5–6 weeks characterizes a month for decomposition TS, which in turn represents pre-forecast information for determining the memory depth of the time series;

- for the all-time series, the value of the Hurst exponent belongs to the zone of "gray" noise. The latter means that the data series do not follow random walks [8];

- the regression equations are of the same type for all studied TS. This conclusion allows us to analyze coefficient *a*, the value of which ranges from 0.61 to 0.67. those the entire investigated model series TS has the property of fractality [1, 8];

- the lengths of the quasi-cycles with the highest frequency are 4 and 5, respectively, which is equal to the calendar month for decomposition TS by days of the week and the number of weekdays for the original TS "Tourists";

- one of the pre-predictive characteristics is the highest frequency of the length of quasi-cycles, which characterizes the "const" property of the process under study. TS "Tourists" has the highest value of this indicator, which means the presence of frequently encountered 5-day quasi-cycles. In this context, we can talk about the company of the trend stability property of the process under study. For decompositional VR, the highest frequency is 10 (weeks), which on average will make it possible to make a medium-term forecast for two months;

To preserve the dynamics of the increments of the initial weekly time series, the authors carried out a data normalization procedure.

The data normalization algorithm follows

1. Find the minimum value in the original TS increments;

2. Calculate its absolute value;

3. Add the latter to each element of the original TS increments;

4. To avoid zero elements of the series, we add to each value of the obtained TS $\Delta > 0$, in our case $\Delta = 1$.

Thus, the dynamics of the studied normalized series of increments correspond to the dynamics of the base series. The positive values of the elements of the normalized series make it possible to calculate the forecast and apply the linear cellular automaton algorithm to the series.

Table 3: Results of applying the linear cellular automaton algorithm for tourist flow decomposed time series.

Indicators / Days of the week	Original time series				Normalized time series of increments		
	Forecast	Memory depth	Forecast error (%)	MAE (%)	Memory depth	Forecast error (%)	MAE (%)
Monday	61.3	8	41	24.4	8	13.6	29.2
Tuesday	26.4	9	43	26.8	8	15.7	31.6
Wednesday	53.2	8	37	19.4	7	17.1	27.4
Thursday	87	9	43	36.6	8	20	50.2
Friday	75.6	13	50	43.8	8	16.7	58.8
Saturday	118.8	8	48	61.9	8	20	77.4
Sunday	175.5	9	47.6	84.8	9	16.8	97.4
Weekdays	325.5	8	29.4	95.4	8	22	108.6
Weekends	354	7	36.4	104.5	7	20.4	116.5
TS "Tourists"	58.4	16	56.9	58.7	16	10.7	71.5

Analysis of the calculations in Table 3 allows us to draw the following.

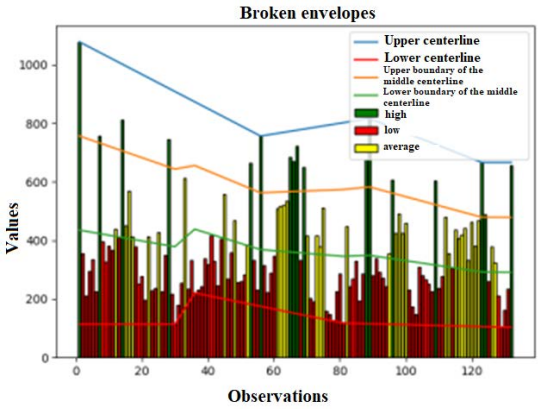
- for the studied initial time series, the value of the memory depth varies in the range from 7 to 16; for normalized decomposition, TS is in the range from 7 to 9. This means that the latter is more trend-stable [8];

- the results obtained based on a linear cellular automaton (column "Forecast error") for the normalized time series of tourist flow values are much better than for the initial TS. The forecast error for the normalized time series of increments does not exceed 22% and is an order of magnitude lower than that of the initial data. Note that for natural processes, the testing of the LCA algorithm shows a forecast error of no more than 25% [7].

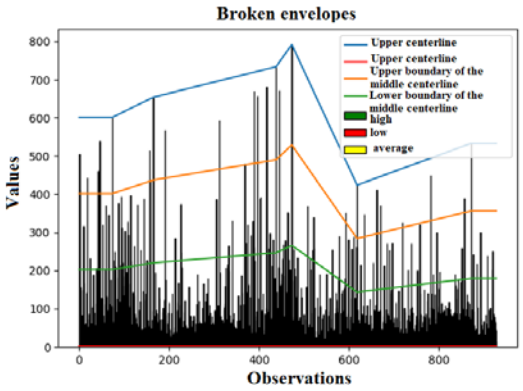
- the average absolute error MAE for the initial decomposition TS varies in the range (19; 85). Therefore, the adequate predictive model is made for TS "Wednesday" since the smallest MAE value was obtained.

Figure 1 shows the author's development "Methods of nonlinear dynamics" [3], which contains the above algorithms for methods of nonlinear dynamics. The result of the operation of the linear cellular automaton algorithm is presented for two initial TS: "Weekdays," for which the

forecast error is less than 30%, and the initial daily TS "Tourists" with the most significant forecast error. Figures e) and f) show three graphs: the dynamics of the TS itself, the result of the forecast model, and the moving average. When conducting a comparative analysis of classical forecasting methods and methods of nonlinear dynamics, it can be concluded that for volatile time series, additional research is required in the context of building formations (aggregation, increment).

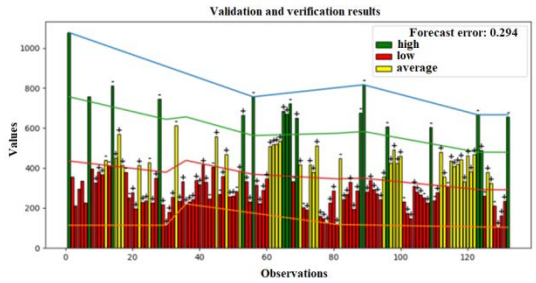


a)

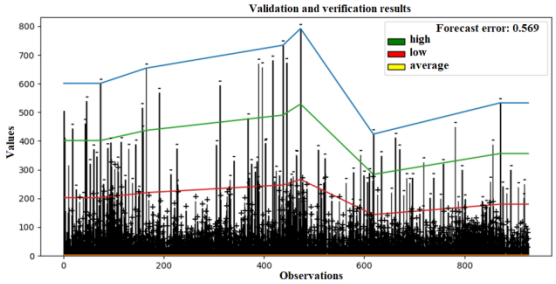


b)

Verification of the predictive model of a linear cellular automaton: a) TS "Weekdays"; b) original TS "Tourists"

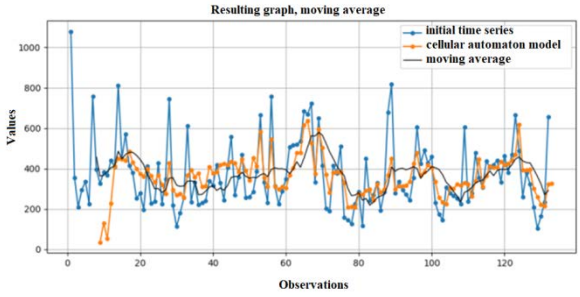


c)

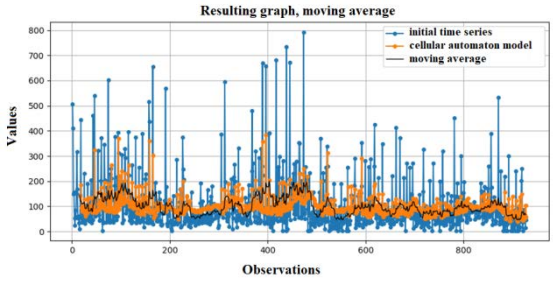


d)

Validation of the predictive model of a linear cellular automaton: c) TS "Weekdays"; d) original TS "Tourists"



e)

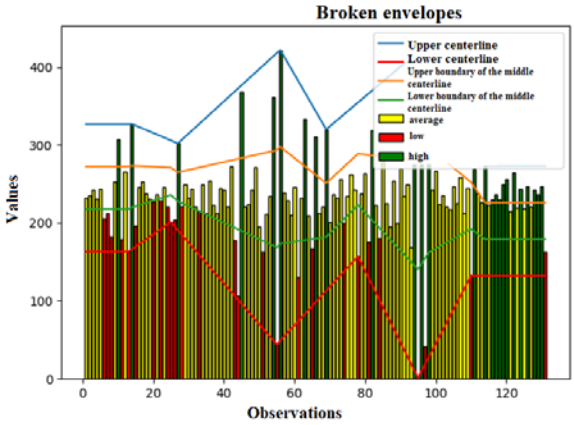


f)

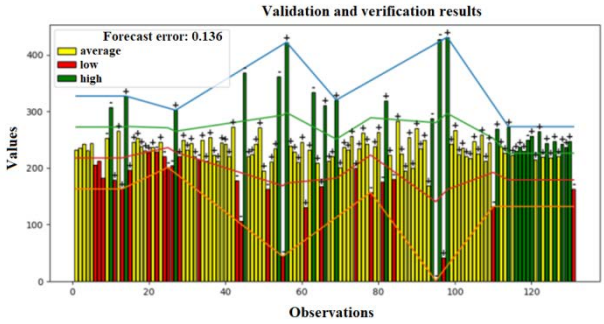
Dephasification of the predictive model of a linear cellular automaton: e) TS "Weekdays"; f) original TS "Tourists»

Figure 1: Interface of the program "Methods of nonlinear dynamics," tab "Linear cellular automaton" (fragment)

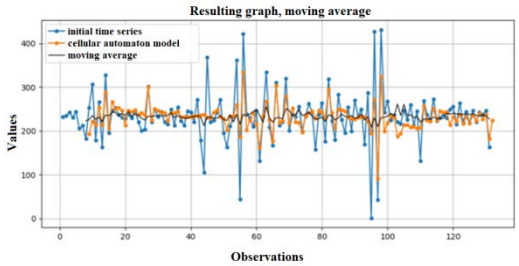
Figure 2 shows the results of applying the algorithm of a linear cellular automaton in the study of the normalized time series of increments "Monday," for which a forecast error of 13.6% was obtained.



a) Verification procedure (three-color coloring: H - low values, C - medium, B - high)



b) Validation procedure



c) Defusification procedure

Figure 2: Linear cellular automaton in the study of the normalized time series of Monday's increments

The mechanism of operation of the listed algorithms allows you to obtain a synergistic effect from the study of complex socio-economic processes when comparing the results of the dyad: the initial time series and a number of its increments.

4 Conclusion

Thus, the work presents the approbation of the developed complex of programs for the automation of nonlinear dynamics methods as applied to the elements of the recreational system on the actual data of the tourist flow. The analysis is carried out, and the procedure for forecasting these data for the lower level of modeling of tourist and recreational activities is carried out. The obtained forecast values, in turn, are the initial data for the models of the upper level of management of tourist and recreational activities (planning the reservation of the room stock), which becomes a significant fact in the context of the development of domestic tourism, within the existing framework of the epidemiological situation.

The proposed methods of multilevel modeling with a predictive result are universal. They can be applied both to the individual enterprise of the tourist complex and to the region. Having in the arsenal of daily data on the tourist flow for an extended period (at least 5-10 years) and using the developed predictive modeling methods, solve tourism management issues from booking rooms to financing and planning investments.

5 Availability of Data and Material

Data can be made available by contacting the corresponding author.

6 Acknowledgement

This reported study was funded by the RFBR according to the research project 19-010-00134A.

7 References

[1] Vintizenko, I.G., Yakovenko V.S. (2008). *Economic cyclomatic*. Moscow, publishing house Finance and statistics, 426 p.

[2] Zang, V.B. (1999). *Synergetic economy. Time and changes in nonlinear economic theory*. Moscow, publishing house "Mir", 335 p.

[3] Kumratova, A.M., Sivkov, K.A. (2020). *Nonlinear dynamics methods*. Certificate of registration of a computer program. RU 2016661998.

[4] Kurdyumov, S.P., Malinetskiy, G.G., Potapov, A.B. (1996). *Nonstationary structures, emotional chaos, cellular automata*. New in synergetics. Riddles of the world of nonequilibrium structures. Moscow, publishing house "Nauka", 95–164.

[5] Naumov, L. (2001). How to Increase the Speed of "Life," or Efficient Data Organization to Increase the Speed of Cell Search and Resolving Neighborhood Relations while Implementing John Horton Conway's Cellular Automaton "Life." *Computer science, 33*, 25-31.

[6] Neumann, J. (1971). *The theory of self-replicating automata*. Moscow, publishing house "Mir", 378 p.

[7] Perepelitsa, V.A., Kasaeva, M.D. (2002). Forecasting a natural time series based on a cellular automaton model. *Modern aspects of the economy, 9*(22), 201-208.

[8] Peters, E. (2000). Chaos and order in the capital markets. Moscow, publishing house "Mir", 333 p.

[9] Sergeeva, L.N. (2002). *Modeling the behavior of economic systems by methods of nonlinear dynamics (chaos theory)*. Zaporizhzhia, 227p.

[10] Chudnovsky, A. D., Zhukova, M. A. (2017). *Management of consumer preferences in domestic tourism and hospitality and the main directions of the implementation of the tourist product*. Moscow, Federal Agency for Tourism, 370p.

[11] Kumratova, A., Popova, E., Khudyakova, E., Vasilenko, I., Saykinov, V. (2021). Statistics Application of the Dynamics Socio-Economic Processes: A Case of Russian Insurance Data. *International Transaction Journal of Engineering, Management, & Applied Sciences & Technologies*, 12(3), 12A3I, 1-9.

[12] Mandelbrot, B. (1982). The Fractal Geometry of Nature. New York: W. H. Freeman, 302 p.

[13] Packard, N., Crutchfield, J., Farmer, D., Shaw, R. (1980). Geometry from a Time Series. *Physical Review, 45*, 712-716.

[14] Kumratova, A., Popova E., Tyunin E., Tretyakova N., Murlin A. (2021). Synergistic Effects in The Dynamics of SocioEconomic Systems and Processes. *Bioscience Research*, *18*(2), 1488-1492.

Alfira Kumratova is an Associate Professor of the Department of Information Systems, Kuban State Agrarian University named after I.T. Trubilin. She is a Candidate of Economic Sciences. Her research interests are Applied Computer Science Methodology and Research Methods.

Dr. Elena Popova is Professor and Head of the Department of Information Systems, Kuban State Agrarian University named after I.T. Trubilin. She is a Doctor of Economic Science. Her research interests are Methods and Tools for Designing Information Systems.

Sergey Kurnosov is an Associate Professor of the Department of Computer Technologies and Systems, Kuban State Agrarian University named after I.T. Trubilin. He is Dean of the Faculty of Applied Informatics. He is a Candidate of Economic Sciences. His research interests are Modern Information and Communication Technologies in Scientific Research.

Valery Kondratiev is an Associate Professor of the Department of Information Systems, Kuban State Agrarian University named after I.T. Trubilin. He is a Candidate of Economic Sciences. His research interests are Information Systems and Technologies.

Natalia Kurnosova is an Associate Professor of the Department of Information Systems, Kuban State Agrarian University named after I.T. Trubilin. She is a Candidate of Economic Sciences. Her research interests are Methods of Increasing the Efficiency of Information Systems.