Singular-spectrum time series analysis for oil and gas industry production forecasting

Aleksandr Zhukov[1,2], *Denis* Zhurenkov[3], *Yuliya* Basalaeva[3], *Victor* Gedzyun[1], Vladimir Skripachev[1], and *Igor* Kartsan[4,5,*]

[1]«The Federal Center of Expertize and Analyzis», Moscow, Russia
[2]Institute of Astronomy of the Russian Academy of Sciences, Moscow, Russia
[3]Federal state unitary enterprise «All-Russia scientific and research institute «Center», Moscow, Russia
[4]Marine Hydrophysical Institute, Russian Academy of Sciences, 2, Sevastopol, Russia
[5]Reshetnev Siberian State University of Science and Technology, Krasnoyarsk, Russia

Abstract. The article considers theoretical aspects of using the model of singular-spectral analysis of time series with the use of decision trees, and also provides justification of the feasibility of using this model for forecasting the production of products for both oil and gas industry and dual-use products. To reduce the risks of distortion of aggregated forecast values when working with large data sets, the expert should carry out preliminary manual selection and exclusion of products, the release of which has been completed or suspended.

1 Introduction

Long-term and medium-term planning of the development of the defense-industrial complex as a set of individual enterprises and organizations is currently based not only on mobilization plans and the state armament program, but also on a systemic complex of all interconnected works necessary for the production of dual-use and civilian products.

Dual-use products play an increasing role in the structure of output of defense industry organizations. Firstly, this is due to the fact that a significant part of the defense industry production falls on components, raw materials, and materials that can be used in finished products for various purposes. For example, the same products of the radio-electronic industry are used in both military and civilian airplanes, ships, missiles, wheeled and tracked vehicles. Secondly, after the imposition of sanctions by foreign countries, the need for domestic components for finished products has increased even more. The transfer of technologies successfully used in the production of military equipment to civilian samples can partially solve the problem of component shortage. In this regard, there is a need for a more detailed analysis of dual-use products output, as well as long-term forecasting of their production in order to make balanced management decisions. At the same time, there is currently a need to create mechanisms for automated forecasting to minimize the negative impact of the human factor.

* Corresponding author: kartsan2003@mail.ru

2 Theoretical aspects of using the model of singular-spectral analysis of time series with application of decision trees

A time series (dynamic series) is understood as a sequence of observations of a random variable (attribute) ordered in time.

Let us present two main types of time series:

- Stationary time series are time series describing random processes that proceed approximately homogeneously in time and have the form of continuous random fluctuations around some average value, and neither the average amplitude nor the characteristic of these fluctuations reveal significant changes over time. However, in most cases, the real economic situation can be represented as a set of non-stationary (random) processes with a definite directional tendency of development;
- Non-stationary (random) time series are time series that may have trend, seasonality and cyclicality.

Currently, there are several methods for analyzing non-stationary time series, the main among them are:

- wavelet transform;
- Huang-Hilbert transform;
- singular spectral analysis method.

The choice of the singular spectral method for the analysis of dual-use production by the organizations of the defense industry is due to the fact that this method allows not only to analyze non-stationary time series, but also solves the most urgent problem - data forecasting.

The singular-spectral method of analysis is a method of analyzing non-stationary time series, which is a variant of the principal component method and is based on the transformation of a univariate time series into a multivariate one (i.e., the singular-spectral method of analysis decomposes a time series into groups of components, and then, by combining them in a special way, allows to obtain a new time series free of non-deterministic components).

Here is a description of the mechanism of the singular-spectral method of time series analysis.

Let there be a non-zero (i.e., there exists at least one i such that $f_i \neq 0$) time series $F = (f_0,\dots,f_{N-1})$ of length N (let $N > 2$). A typical analysis mechanism consists of two stages [4]:

Decomposition. Stage 1: Nesting is a translation of the original time series into a sequence of multivariate vectors. Let L be the window length (some integer), $1 < L < N$. The embedding procedure forms $K = N - L + 1$ embedding vectors having dimensionality L:

$$X_i = (f_{i-1},\dots,f_{i+L-2})^T, 1 \leq i \leq K \tag{1}$$

where: X is the trajectory matrix of the time series F consists of embedding vectors as columns:

$$X = [X_i : \dots : X_K] \tag{2}$$

trajectory matrix X is a matrix of the form:

$$X = (x_{ij})_{i,j=1}^{L,K} = \begin{bmatrix} f_0 & f_1 & f_2 & \cdots & f_{K-1} \\ f_1 & f_2 & f_3 & \cdots & f_K \\ f_2 & f_3 & f_4 & \cdots & f_{K+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ f_{L-1} & L & f_{L+1} & \cdots & f_{N-1} \end{bmatrix} \tag{3}$$

So, $x_{ij} = f_{i+j-2}$ and matrix X has equal elements on the "diagonals" $i+j = const$. Thus, the trajectory matrix X is a gankel matrix (i.e., a matrix with equal elements on all diagonals perpendicular to the main diagonal). There is a one-to-one correspondence between gankel matrices of dimension LxK and the time series $N = L+K - 1$.

Stage 2: Singular Decomposition. The result of this stage is the singular decomposition of the time series trajectory matrix.

Let $S = XX^T$. Let $\lambda_1, \ldots, \lambda_L$ denote the eigenvalues of the matrix S taken in non-decreasing order ($\lambda_1 \geq \ldots \lambda_L \geq 0$), and $U_1, \ldots, U_L$ denote the orthonormalized system of eigenvectors of the matrix S corresponding to the singular numbers $\lambda_1, \ldots, \lambda_L$.

Let $d = max\{i: \lambda_i > 0\}$. If we denote $V_i = X^T U_i / \sqrt{\lambda_i}, i = 1, \ldots d$, then the singular expansion of the matrix X can be written as:

$$X = X_1 + \ldots + X_d \tag{4}$$

where $X_i = \sqrt{\lambda_i} U_i V_i^T$. Each of the matrices X_i has rank 1, hence they can be called elementary matrices. The set $(\sqrt{\lambda_i} U_i V_i^T)$ is called the i-th singular triple of the expansion.

Recovery. Step 3: Grouping. Based on the decomposition obtained in formula (4), the grouping procedure divides the entire set of indices $\{1, \ldots, d\}$ into m non-overlapping subsets $I_1, \ldots, I_m$.

Let $I = \{i_1, \ldots, i_p\}$. Then the resulting matrix X_1, corresponding to the group I, is defined as:

$$X_1 = X_{11} + \ldots + X_{1p} \tag{5}$$

Such matrices are computed for $I = I_1, \ldots, I_m$, thus the decomposition given in formula (4) can be represented in a grouped form:

$$X = X_{i1} + \ldots + X_{ip} \tag{6}$$

The procedure for selecting sets $I = I_1, \ldots, I_m$ is called grouping of singular triples.

Stage 4: Diagonal averaging. At this stage, each matrix of the grouped decomposition specified in formula (6) is translated into a new time series of length N.

Let Y be some LxK matrix with elements y_{ij}, where $1 \leq i \leq L$, $1 \leq j \leq K$. Let $L^* = min(L, K), K^* = max(L, K)$ and $N = L + K - 1$. Let $y_{ij}^* = y_{ij}$ if $L < K$. Diagonal averaging translates the matrix Y into the time series $g_0, \ldots, g_{N-1}$ by the formula:

$$g_k = \begin{cases} \frac{1}{k+1} \sum_{m=1}^{k+1} y_{m,k-m+2}^* & \text{for} \quad 0 \leq k < L^* - 1, \\ \frac{1}{L^*} \sum_{m=1}^{L^*} y_{m,k-m+2}^* & \text{for} \quad L^* - 1 \leq k < K^*, \\ \frac{1}{N-k} \sum_{m=k-K^*+2}^{N-K^*+1} y_{m,k-m+2}^* & \text{for} \quad K^* \leq k < N. \end{cases} \tag{7}$$

Formula (7) represents the averaging of matrix elements along the "diagonals" $i+j=k+2$: choosing $k=0$ gives $g_0 = y_{11}$, for $k=1$ we obtain $g_1 = (y_{12} + y_{21})/2$, etc.

Applying diagonal averaging (7) to the resulting matrices, we obtain the time series $\tilde{F}^{(k)} = (\tilde{f}_0^{(k)}, \ldots, \tilde{f}_{N-1}^{(k)})$, and, as a consequence, the original time series $(f_0, \ldots, f_{N-1})$ decomposes into the sum of m time series:

$$f_n = \sum_{k=1}^{m} \tilde{f}_n^{(k)} \tag{8}$$

The main advantages of applying the singular-spectral method to time series analysis are:

- absence of a strict requirement for stationarity of the time series;
- possibility of application to noisy and short series;
- the possibility to identify both periodic and complex non-stationary components of time series.

The weakness of the singular-spectral method is the absence of an analytical model representation of the series, for example, in the form of a sum of simple functions, a compact analytical representation of which could be clearer and more accessible for interpretation than an aggregate of a large number of components [1].

The singular-spectral method of analysis forms a trend of the initial data series development, even if it is not linear. At the same time, for the purpose of further training, a time series of trend deviations from real values, which is commonly referred to as "noise", is identified. It should be noted that focusing on trend and noise improves the accuracy and generalizability of models due to the fact that fewer factors and data are needed to identify patterns.

Training of the noise component can be accomplished by various methods, including the use of a decision tree model.

A decision tree is a data analysis tool that is a hierarchical tree structure that consists of a set of rules that solve problems like "If ..., then ... .".

The general scheme of operation and the main elements of a decision tree are presented in Figure 1.

The method of decision trees consists in the following mechanism: the set of feature vector values is divided into non-overlapping subsets, after which a separate model is built for each such subset. It is necessary to construct a function $f: X \rightarrow Y$ from the set $D = \{x_i, y_i\}_{i=1}^{l}$, called training sample with unknown distribution $P(x,y) = P(x)P(y|x)$. Where x_i, $i = 1,2,...,l$ are feature vectors or priors [3].
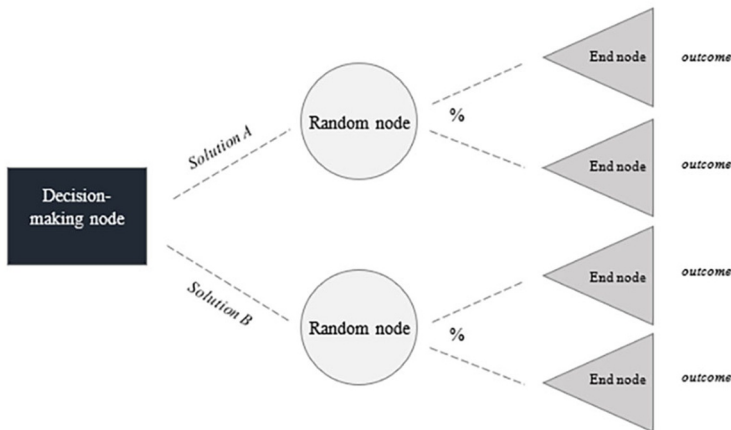

Fig. 1. General scheme of operation of the decision tree.

At the same time, there are several scales of feature measurement: numerical, ordinal, nominal. If y_i (dependent variable) can take only a finite number of values, i.e., $y_i \in \{w_1, w_2, \ldots, w_c\}, c \geq 2$, then it is necessary to solve the classification problem, y_i is called a class label and determines the belonging of the corresponding object to one of c classes, and the feature itself is called a class label. If y_i (dependent variable) is measured on a numerical scale, then it is necessary to solve the problem of regression, and the feature itself is called response.

Among the advantages of using this method of data analysis are the following:
- decision tree algorithms are similar to human decision making;
- the decision tree mechanism is simple enough to understand and does not require special knowledge;
- relatively small amount of calculations and fast training;
- the possibility of working with discrete data.

3 Analysis of the application of singular-spectral analysis of time series with the use of decision trees for forecasting the production of products, including dual-use products

To analyze the adequacy of applying the model of singular-spectral analysis of time series with the application of decision trees to forecasting the production of dual-use products, the expert forecasts of defense industry organizations by industry were compared with the

forecast of artificial intelligence according to the model and the corresponding deviations were identified.

To conduct the study, a sample of 46 nomenclature items of dual-use products was made, which most fully reflect the general picture of dual-use production by defense industries, of which:

- 15 nomenclature items produced by aviation industry organizations, including: navigation, meteorological and geophysical equipment, systems and complexes of airborne radio electronic equipment, autopilot equipment, fire alarm systems and other high-tech products;
- 21 nomenclature items produced by organizations of radio electronic industry, including: radio measuring equipment, microwave electronics and related components, components of electrical machines, vacuum devices, complexes of devices for pre-flight inspection of onboard radio electronic equipment and other high-tech products;
- 10 nomenclature items produced by oil and gas industry organizations, including: oil and gas instrumentation products (engines, communication equipment, electromagnets, electromagnetic contactors and starters, navigation, meteorological, geophysical devices and similar instruments), related products (lighting fixtures) and other high-tech products.

The inputs for the forecast were the actual dual-use output data for 2018-2022, as well as the estimated value for 2023. The forecast values were constructed for the period up to 2033.

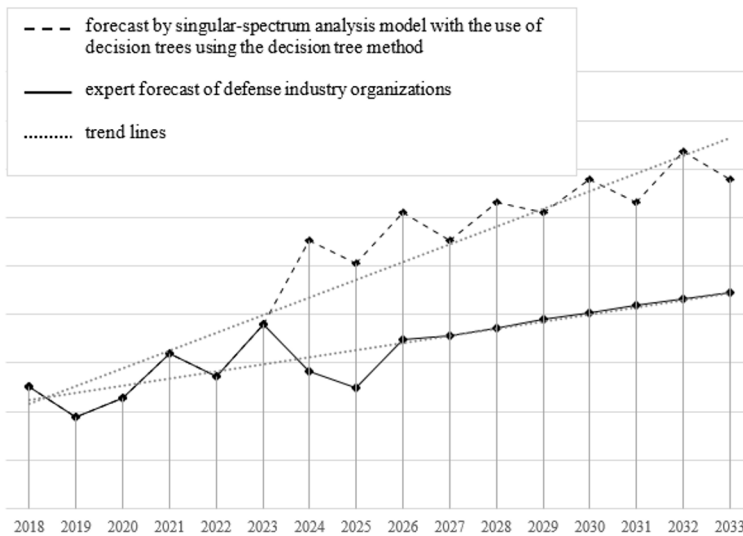To summarize the situation, we present a comparison of aggregated data by industry.

Fig. 2. Comparison of aggregated forecasts for dual-use products of aviation industry organizations.

3.1 Aviation industry

Based on the comparison of forecast data for 15 key dual-use products of the aviation industry to which the artificial intelligence forecasting mechanism was applied, the following conclusions can be drawn (Figure 2):

- linear trends of the forecast for the period up to 2033 have the same direction and do not diverge in the long-term period;

- in the short-term period, the model forecast may significantly deviate from the expert forecast of defense industry organizations due to the fact that the model does not take into account the existing plans of the defense industry organization, as well as political and economic risks;
- the expert forecast of the DIC organizations is more conservative compared to the forecast modeling, which may be due to intentional underestimation of values in order to achieve (exceed) the indicators in the future;
- on average, the model forecast deviates by 67% upwards from the expert forecast of organizations;
- the average growth rate for the period 2018-2033 according to the expert forecast of organizations was 104.2%, and according to the forecast modeling - 107.3%. The divergence of 3.1 p.p. is not so significant in the fifteen-year interval of values.

3.2 Radioelectronic industry

Based on the comparison of forecast data for 21 key dual-use products of the radio electronics industry, to which the forecasting mechanism with the help of artificial intelligence was applied, the following conclusions can be drawn (Figure 3):
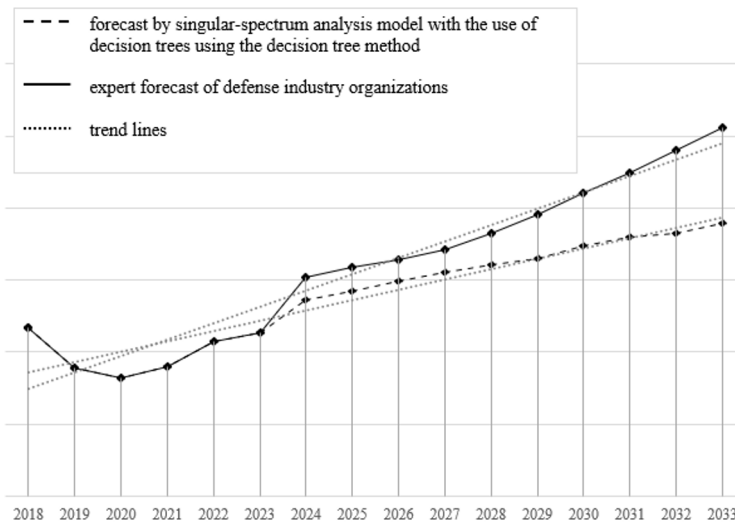

Fig. 3. Comparison of aggregated forecasts for dual-use products of radioelectronic industry organizations.

- linear trends of forecasts in the short-term and long-term period have the same direction and do not diverge;
- the expert forecast of the defense industry organizations is more optimistic
- compared to forecast modeling;
- on average, the model forecast deviates by 16% downward from the expert forecast of the DIC organizations;
- the average growth rate for the period 2018-2033 according to the expert forecast of the DIC organizations was 105.7%, and according to the forecast modeling - 103.5%. The discrepancy of 2.2 p.p. can be considered insignificant.

3.3 Oil and gas industry

Based on the comparison of forecast data for 10 key dual-use products of the oil and gas industry to which the artificial intelligence forecasting mechanism was applied, the following conclusions can be drawn (Figure 4):

- – in general, the linear trends of forecasts have the same direction and do not diverge over the long-term period;
- – in the short-term period, the model forecast may significantly deviate from the expert forecasts of the defense industry organizations, which may be due to the cyclical nature of shipbuilding production;
- – the expert forecast of the defense industry organizations is more optimistic compared to the forecast modeling;
- – on average, over the entire period, the model forecast deviates by 6% downward from the expert forecast of the defense industry organizations;
- – the average growth rate for the period 2018-2033 according to the expert forecast of DIC organizations was 123.8%, and according to the forecast modeling - 119.8%. The discrepancy of 4 p.p. in the long-term period can be considered insignificant.
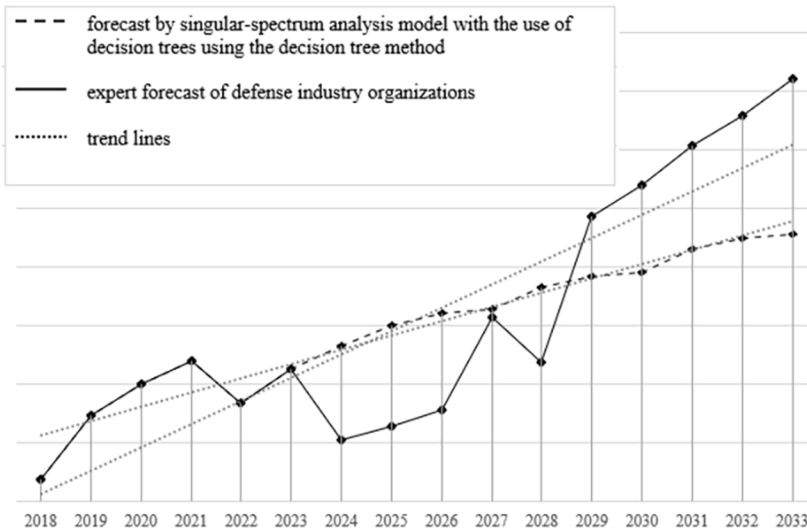

Fig. 4. Comparison of aggregated forecasts for dual-use products of oil and gas industry organizations

It is important to note the aspect that forecasting based on singular-spectral analysis of time series using decision trees cannot be used for products that are in the phase-out stage, as the model cannot predict the suspension or termination of production, which in turn can be caused by both random risks and plans of the organization. Let us illustrate this shortcoming with an example.

Let the organization produced some product in the period from 2018 to 2022, and in 2023 the production ended (the production volume amounted to 0 rubles). Forecast modeling cannot determine the zero value as the end of production, therefore it builds a forecast with increasing values (Figure 5).
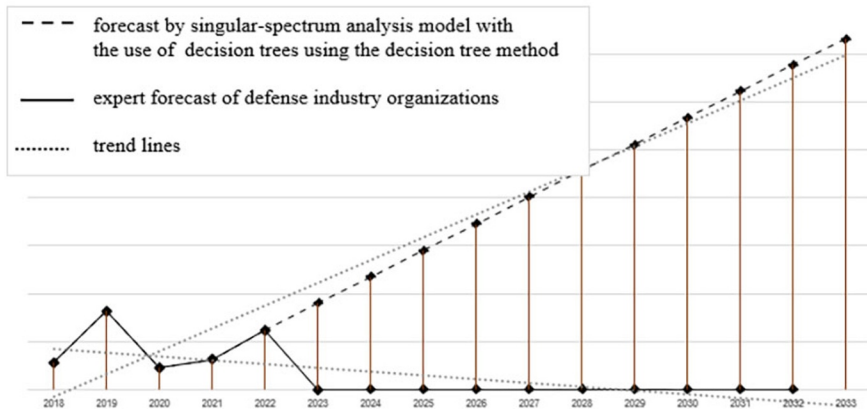
Fig. 5. Prediction errors.

In order to reduce the risks of artificial intelligence distorting the forecast when working with large data sets, the expert should perform a preliminary manual selection and exclusion of products that are being finalized. For the above analysis of 46 dual-use products in three industries, such data were excluded.

4 Conclusion

The study, based on the analysis of a sample of 46 key products to which the forecasting mechanism with the help of artificial intelligence was applied, allowed us to draw the following conclusions.

Forecasting based on the model of singular-spectral analysis of time series with the use of decision trees shows adequate results and can be used to forecast aggregate values of dual-use products output by the organizations of the defense industry in the long-term period in the context of industries and the defense industry as a whole.

Forecasting based on the model of singular-spectral analysis of time series with the use of decision trees does not take into account possible political, economic and other risks, so expert adjustments should be made (if necessary) to the final machine forecast.

References

1. N.V. Abalov, V.V. Gubarev, O.K. Alsova, Proceedings of SPIIRAN **4(35)**, 49-63 (2014)

2. A.M. Batkovskiy, M.A. Batkovskiy, P.V. Kravchuk, Actual issues of modern economics **3**, 267-279 (2020)

3. S.A. Tolkachev, Economic Revival of Russia **3(61)**, 127-145 (2019)

4. I.N. Kartsan, S.V. Efremova, V.V. Khrapunova, M.I. Tolstopiatov, IOP Conference Series: Materials Science and Engineering, **450(2)**, 022015 (2018). https://www.doi.org/10.1088/1757-899X/450/2/022015

5. V.P. Bauer, S.N. Silvestrov, P.Yu. Baryshnikov, Information Society **3**, 30-40 (2017)

6. A.A. Grishin, S.P. Stroev, Vestnik of Technological University **9**, 180-186 (2018)

7. H. Hassani, Journal of Data Science **5**, 239-257 (2007)

8. S.I. Dovguchits, D.A. Zhurenkov, Scientific Bulletin of the defense-industrial complex of Russia **3**, 16-24 (2022). https://www.doi.org/10.52135/2410-4124_2022_3_16

9. T.B. Timofeeva, E.A. Ozdoeva, Management **8(3)**, 112-122 (2020). https://www.doi.org/10.26425/2309-3633-2020-8-3-112-122

10. S.V. Porshnev, F. Rabaya, *Study of the peculiarities of the singular spectral analysis method application in the task of time series analysis and forecasting* (Zebra, Ulyanovsk, 2016)

11. S.A. Prokhorov, *Prikladnoy analiz neekvidistantnykh vremennykh ryadov* (SGAU Publ., Samara, 2001)

12. I. Kartsan, A. Zhukov, S. Pronichkin, *E3S Web of Conferences* **389**, 090056 (2023). https://www.doi.org/10.1051/e3sconf/202338909056

13. Ch. Chui, *An Introduction to Wavelets* (Mir Publ., Moscow, 2001)

14. A.O. Zhukov, K.S. Khachaturyan, S.A. Khachaturyan, Modern Science: Actual Problems of Theory and Practice. Series: Economics and Law **4**, 18-22 (2023). https://www.doi.org/10.37882/2223-2974.2023.04.11

15. S.I. Dovguchits, Scientific Bulletin of the Russian Defense Industry Complex **4**, 39-42 (2020)

16. S. Malla, *A Wavelet Tour of Signal Processing* (Mir Publ., Moscow, 2005)

17. S.I. Dovguchits, D.A. Zhurenkov, Scientific Bulletin of the defense-industrial complex of Russia **4**, 5-12 (2022)

18. Z. Peng, G. Wang, BioMed Research Int. 2017, 12 (2017)

19. I. Kartsan, A. Zhukov, S. Pronichkin, *E3S Web of Conferences* **376**, 05013 (2023). https://www.doi.org/10.1051/e3sconf/202337605013

20. A.I. Stankevich, Volga region. Physico-mathematical sciences **1(9)**, 134-138 (2009)

21. K. Liao, R. Xiao, J. Gonzalez, L. Ding, PLOS ONE **9(1)**, e85192 (2014)

22. R.M. Baevsky, G.G. Ivanov, L.V. Chireykin, A.P. Gavrilushkin, P.Ya. Dovgalevsky, Yu.A. Kukushkin, T.F. Mironova, D.A. Prilutsky, A.V. Semenov, V.F. Fedorov, A.N. Fleyshman, M.M. Medvedev, Vestnik aritmologii **24**, 65-87 (2002)

23. O.V. Mandrikova, T.L. Zalyaev, Digital Signal Processing **1**, 11-16 (2015)

24. V.V. Vityazev, *Analiz neravnomernykh vremennykh ryadov* (SPb Univ. Publ., St. Petersburg, 2001)

25. D.K. Galyagin, P.G. Frik, Matematicheskoe modelirovanie sistem i protsessov **4**, 20-28 (1996)

26. S.A. Prokhorov, *Approximative analysis of random processes* (SGAU Publ., Samara, 2001)

27. B.N. Avdonin, A.M. Batkovskiy, M.A. Batkovskiy, Voprosy radioelectroniki. **4(2)**, 35-47 (2014)

28. B.N. Avdonin, A.M. Batkovsky, M.A. Batkovsky, Economics, Statistics and Informatics, Bulletin of UMO **2**, 147-149 (2014)

29. O.V. Glebova, O.V. Gracheva, A.V. Simonov, Modern Economy Success **4**, 29-36 (2019)

30. S.S. Chebotarev, V.A. Elshin, Economics and Entrepreneurship **5(106)**, 55-60 (2019)

31. A.M. Batkovskiy, Voprosy innovatsionnoy ekonomiki **3(3)**, 36-46 (2011)

32. I.A. Kolesnikov, V.B. Konovalov, V.V. Korzo, Voennaya Mysl **10**, 48-52 (2018)

33. G.N. Chernysheva, L.V. Shchupak, Ekonominfo **16(2-3)**, 77-80 (2019)

34. A.V. Boykova, A. Tokarev, Financial Economics **9**, 445-447 (2019)

35. A.M. Batkovskiy, Economics, entrepreneurship and law **2**, 38-54 (2011)

36. A.M. Batkovskiy, Economics, entrepreneurship and law **1**, 30-44 (2011)

37. L.A. Sazanova, International journal of social and humanitarian sciences **1(1)**, 202-204 (2016)

38. A.N. Gorban, P.A. Gorban, G. Judge, Entropy 12(5), 1145-1193 (2010)

39. G. Valenduc, P. Vendramin, Transfer **23(2)**, 121–134 (2017)

40. K.I. Dementiev, Management consulting **1**, 107-120 (2023)