Journal of Engineering and Applied Sciences 14 (2): 348-355, 2019
ISSN: 1816-949X

Various Processes Through the Hidden Stable Dependencies

Alexandr Baldin, Vladimir Burmistrov and Ivan Eroshok
Department of IU-5, Moscow State Technical University, ul. Baumanskaya 2-ya 5,
Moscow, Russia

Abstract: In light of increasing amount of data such challenges as the restoration of lost information fragments and its forecasting are becoming relevant and attractive. Mathematical description of studied processes is one of the main factors for solving such problems. Most of current methods are based on the search for a model that can accurately repeat the geometric form of the process. However, such an approach does not provide an understanding of how much the hidden dependencies can affect the desired result. This study is devoted to various methods of searching and describing processes through such dependencies.

Key words: Time series, signal processing, predictive analytics, big data, hidden dependencies, mathematical description

INTRODUCTION

The study and analysis of time series allows solving a variety of applied problems. These problems include: electricity price forecasting (Weron, 2014), social process study and analysis (Kawash *et al*., 2017), risk assessment in financial activity (Schuermann, 2014), economic forecasting (Alexandridis *et al*., 2017), weather forecasting (Ericsson, 2017) and many others.

Currently, there are many methods (classification of prediction methods and models, 2016) for solving these problems. However, despite this fact such problems as forecasting and restoring the missing values do not have a single-valued solution. A reliable result directly depends on the way how the target process was described. When it comes to a successful process description one understands it as the best process approximation and evaluates the model with standard measures (MAE, MAPE, MASE). Strictly speaking, this attitude is not always correct (Davydenko and Fildes, 2016) because there are more complex relationships than those described by the geometric interpretation.

In solving practical problems, Autoregressive (AR) methods (Filik and Kurban, 2007) and methods based on Artificial Neural Networks (ANN) were most widely used (Morariu *et al*., 2009). In most classical methods if one looks for choosing a prediction model, he/she has to filter the studied process, to allocate the trend (Baheti and Toshniwal, 2014), to determine the periodic and seasonal components (Jiann, 2005), to conduct a stationarity test (Petriea *et al*., 2017) and to make the process stationary if necessary. This approach comes with a lot of difficulties, and in most cases is not fully automated.

The purpose of this study is to propose an alternative approach based on the search for and identification of objective factors that affect the image of a process with hidden stable dependencies. This should fundamentally change the approach to the problems of research and prediction.

MATERIALS AND METHODS

In this research, we have used methods of mathematical statistics, regression and autoregressive analysis, numerical methods for solving systems of linear algebraic equations and correlation analysis.

Studied processes and their main characteristics: We have considered the physical processes that reflect the Mean Square Errors (MSE) of object positioning in space. Let us preliminary analyze them.

Figure 1 presents a Mean Square Deviation (MSD) graph of positional error σ^2 information of a space object in time t (in 360 sec). Figure 2 shows a histogram of distributed values of this series. This process has no explicit trend (y = -0.038x+437). The histogram shows that the process is close to a stationary one and fluctuates around the mean (430) m^2 the maximum and minimum values are 759 and 169 m^2.

Figure 3 shows another process its length is n = 360 values. The process has a small trend (y = 0.5109x+789.29). The histogram is shown in Fig. 4.

The next process is shown in Fig. 5. According to its graph, the process probably has periodicity and a property of stationarity. The histogram of this process is shown in Fig. 6.

Corresponding Author: Alexandr Baldin, Department of IU-5, Moscow State Technical University, ul. Baumanskaya 2-ya 5, Moscow, Russia

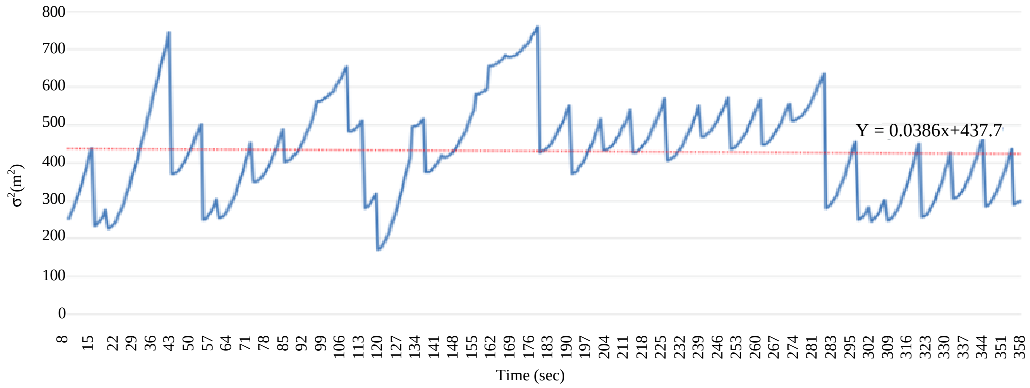

Fig. 1: MSD graph of σ^2 (m^2) according to t (sec), series No.1

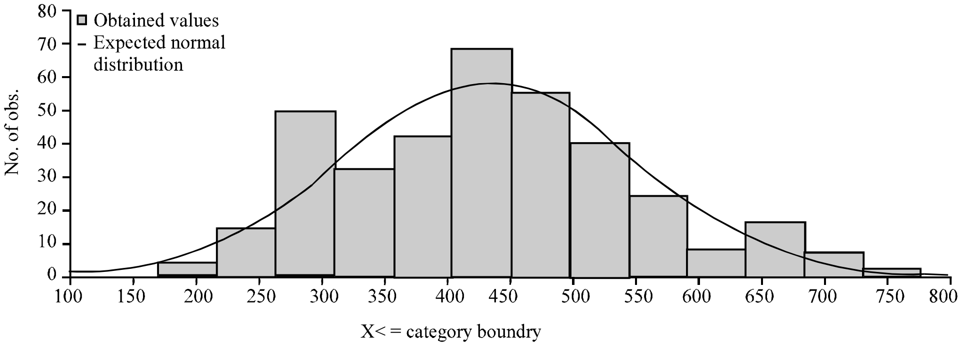

Fig. 2: Historgam of values (Series No. 1) K-s d = 05706, p<20; Lilliefors p<01

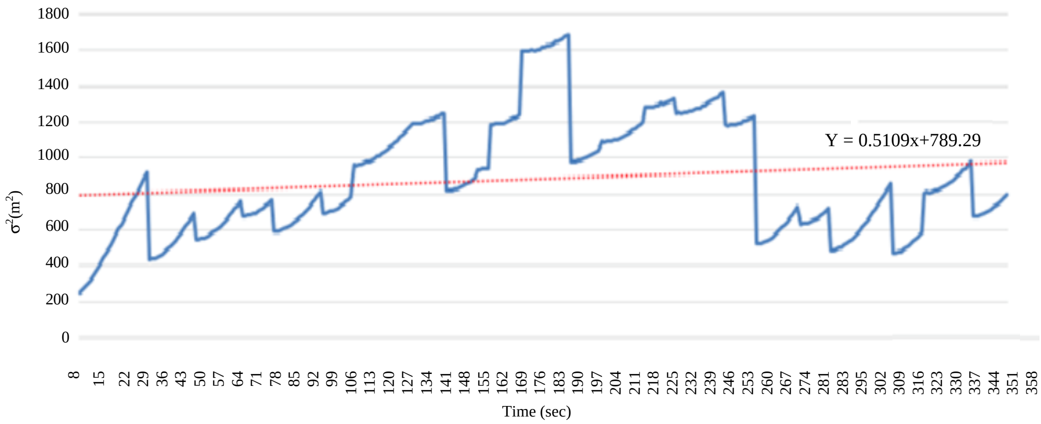

Fig. 3: MSD graph of σ^2 (m^2) according to t (sec), series No. 2

Preliminary analysis shows that all 3 processes differ not only in their physical origin and form but also in characteristic statistical values.

Stability of dependencies: In research we have found that many time series of data possess fractal properties (self-similarity, fractal dimension) (Schroeder, 2005). This allows predicting their changes, determining hidden dependencies, periodicity, etc.

Problem setting: The initial process is presented as a time series in the range from t_1-t_n:

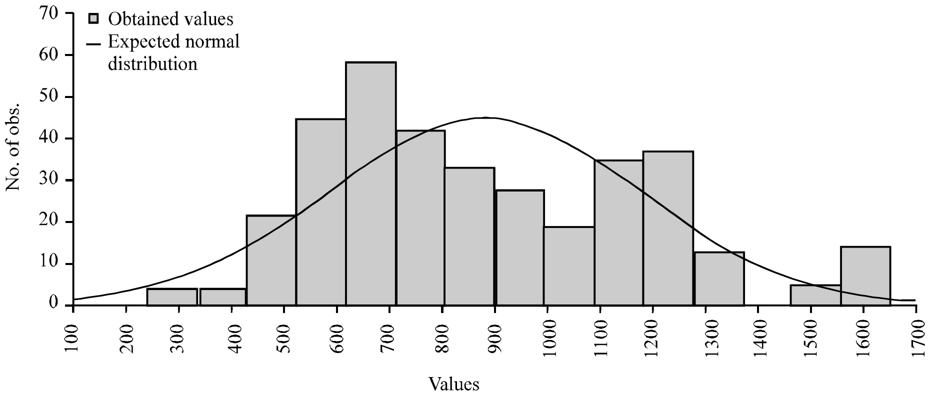

Fig. 4: Historgam of values (Series No. 2) K-s d = 11148 p<01; Lilliefors p<01

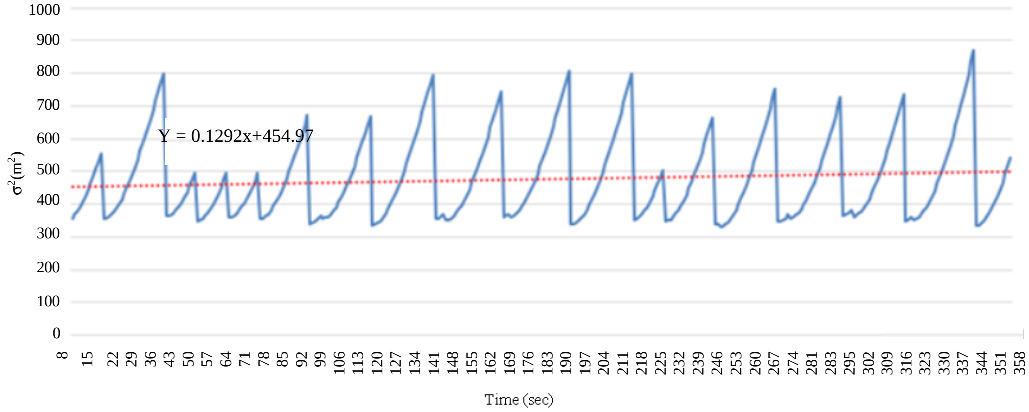

Fig. 5: MSD graph of $\sigma^2(m^2)$ according to t (sec), series No. 3

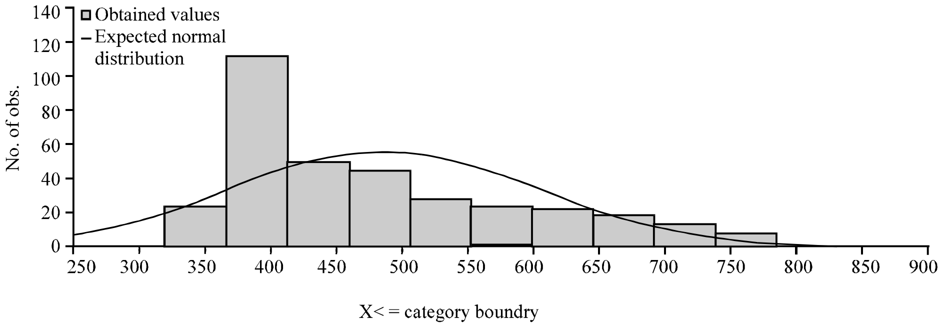

Fig. 6: Historgam of values (Series No. 3)

$$y_1, y_2, y_3, \ldots, y_k, y_n \quad (1)$$

Let us divide this time span into m equal intervals with the length of $\tau = k$ and represent the series Eq. 1 as follows:

$$y_1, \ldots, y_k, y_{k+1}, \ldots, y_{2k}, y_{2k+1}, \ldots, y_{3k}, \ldots, y(m-1)\times k+1, \ldots, y_{m\times k}$$

where:

$$n = m\times k + v \text{and } 0 \leq v \leq k$$

or represent the process as intervals $y(\tau_1)$, $y(\tau_2)$, ..., $y(\tau_m)$, We will allocate four variables for each interval:

$$y_i, y_j, y_p, y_q$$

That will fit the conditions $i \neq j$ and $p \neq q$:

$$y(t) = y\begin{pmatrix} \tau_1 \\ \tau_2 \\ \dots \\ \tau_m \end{pmatrix} = \begin{pmatrix} y_1 & y_2 & \dots & y_k \\ y_{k+1} & y_{k+2} & \dots & y_{2k} \\ y_{2k+1} & y_{2k+2} & \dots & y_{3k} \\ \dots & \dots & \dots & \dots \\ y_{(m-1)k+1} & y_{(m-1)k+2} & \dots & y_{mk} \end{pmatrix} \Rightarrow$$

$$\begin{pmatrix} y_i & y_j & y_p & y_q \\ y_{k+1} & y_{k+j} & y_{k+p} & y_{k+q} \\ y_{2k+i} & y_{2k+j} & y_{2k+p} & y_{2k+q} \\ \dots & \dots & \dots & \dots \\ y_{(m-1)k+i} & y_{(m-1)k+j} & y_{(m-1)k+p} & y_{(m-1)k+q} \end{pmatrix}$$

Let us form two functional dependencies for each interval –$F_1 = f(y_i, y_j)$ and $F_2 = f(y_p, y_q)$– and compose two numerical sequences out of them:

$$F_{11}, F_{12}, F_{13}, \dots, F_{1m}$$

And

$$F_{21}, F_{22}, F_{23}, \dots, F_{2m}$$

We will find the mean value of time-series data for each interval:

$$F_{mean} = \frac{\sum_{i=1}^{k} y_i}{k}$$

Then, we will compose a numerical sequence out of mean values:

$$F_{mean1}, F_{mean2}, F_{mean3}, \dots, F_{mean\_m}$$

Now we can transform the initial process Eq. 1:

$$y(t) \Rightarrow Y(F_1, F_2, F_{cp}) =$$

$$\begin{pmatrix} F_{11} = f(y_i, y_j) & F_{12} = f(y_p, y_q) & F_{1cp} \\ F_{21} = f(y_{k+i}, y_{k+j}) & F_{22} = (y_{k+p}, y_{k+q}) & F_{2cp} \\ F_{31} = f(y_{2k+i}, y_{2k+j}) & F_{32} = f(y_{2k+p}, y_{2k+q}) & F_{3cp} \\ \dots & \dots & \dots \\ F_{m1} = f(y_{(m-1)k+i}, y_{(m-1)k+j}) & F_{m2} = f(y_{(m-1)k+p}, y_{(m-1)k+q}) & F_{mcp} \end{pmatrix}$$

We propose two hypotheses about the stable dependencies on the time-series interval. The first one stands for the equality between the dependencies $F_1 = f(y_i, y_j)$ and $F_2 = f(y_p, y_q)$ on the considered time interval:

$$F_1 = F_2 \tag{2}$$

The second for the equality between the mean value of variables on the interval with the dependencies $F_1 = f(y_i, y_j)$ and $F_2 = f(y_p, y_q)$ on the interval:

$$F_1 = F_2 = F_{mean} \tag{3}$$

Dependencies F_1, F_2 and f_{mean} on the intervals. Let us consider the processes represented as 3 time series. Their preliminary analysis is presented in the previous study. We will check how hypothesis are fulfilled in each of them.

The considered period of each process is alternately divided into intervals from k = 4 to k = 50. We form the series of F_1, F_2 and f_{mean} dependencies for each process of each sequence of intervals. Table 1 shows the correlations between these dependencies over all the studied intervals for each time series. Then we choose one interval with the less fulfilled correlation between the F_1, F_2 and f_{mean} dependencies in each of three time series. These dependencies will be checked for conformity with the hypotheses put forward.

Strategy of criterion determination for assessing the priority of each interval, the subject of individual studies: In this study, we have used the following formula as the K_{cr} criterion:

$$K_{cr} = K_{general} - \sigma_{error}$$

$$K_{general} = Ks_1 \times Ks_2 \times K_{Fmean}$$

$$\sigma_{error} = \frac{1}{\left[\frac{n}{k}\right]} \sum_{m=1}^{\left[\frac{n}{k}\right]} \frac{1}{F_{mean_m}} \sqrt{D\left[\sum_{i=1.m}^{k.m} F_{mean_m} - y_i\right]}$$

where, Ks_1, ks_2 and Kf_{mean} are dependency correlations, respectively.

Proof of hypothesis: We will prove the hypothesis by working with the sequences, composed of dependences F_1, F_2 and f_{mean} for the intervals k = 7 (series No. 1 and 2) and k = 9 (series No. 3) that were considered as the worst ones in terms of correlation fulfillment according to our criterion.

Since, hypothesis 2 is a special case of hypothesis 3, then only one hypothesis 3 is subject to verification and evaluation. If we have the equality Eq. 3, then the condition Eq. 4 must be satisfied-the equality between arithmetic mean and geometric mean values:

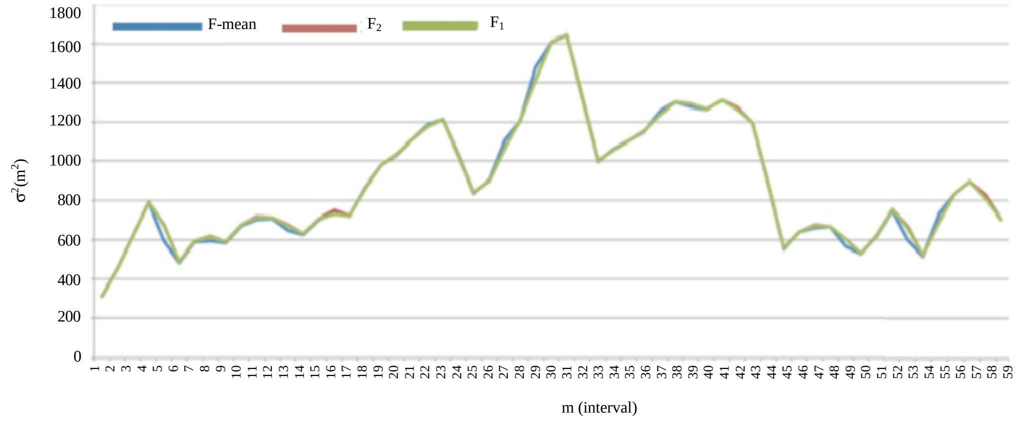

Fig. 7: Correlation between F_1, F_2 and f_{mean} in the intervals of series No. 2, k = 6

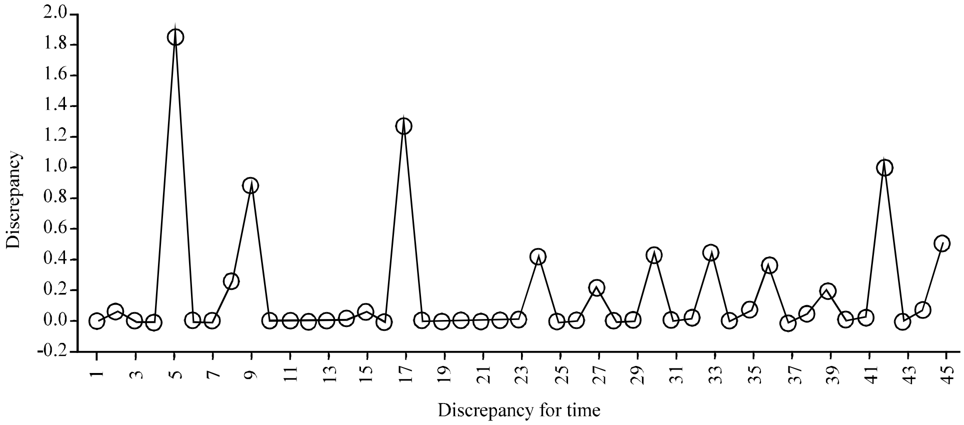

Fig. 8: Deviation (residual $D_{residual}$) of series No. 1

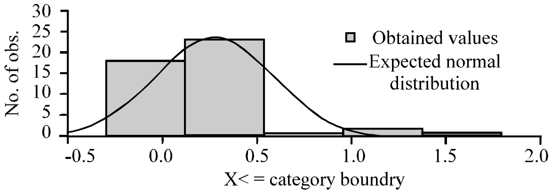

Fig. 9: Histogram of deviations $D_{residual}$ (Series No. 1) K-s d = 34906, p<01; Lilliefors p<01

$$D_{residual} = \frac{F_1+F_2+F_{mean}}{3} - 3\sqrt{F_1+F+F_{mean}}$$

$$\frac{F_1+F_2+F_{mean}}{3} = 3\sqrt{F_1+F+F_{mean}} \quad (4)$$

To check the Eq. 4 we will evaluate the residual $D_{residual}$, found by the formula: In this case we form two sequences of arithmetic mean and geometric mean values, composed of F_1, F_2 and f_{mean} dependencies according to, chosen intervals of each physical process. Unlike the white noise, a stationary process with finite variance which random variables are not correlated and their mean is zero, the studied process has insignificant differences. Its parameters are shown in Fig. 7-9.

$D_{residual}$ is a stationary process with a finite variance (0.619) which random variables are not correlated and their mean (0.341) tends to zero, namely-it is close to white noise. Dependencies F_1, F_2 and f_{mean} after re-transformation.

Let us check the presence of dependencies F_1, F_2 and f_{mean} on the time series formed on the basis of mean values of the interval. Let us carry out a research similar to the one in the previous section but with a time series composed of f_{mean} for each interval as initial data. Thus, we will compress the original process by k times. It is worth noting that we have taken the best intervals according to the K_{cr} criterion from Table 1.

Table 1: Criteria and correlations between dependencies F_1, F_2 and f_{mean}

k	Ks_1	Ks_2	$K_{F\text{-}mean}$	$K_{general}$	σ_{error}	K_{cr}
4	0.991780	0.994097	0.999288	0.985224	0.061441	0.923783
6	0.977762	0.992330	0.994713	0.965133	0.071426	0.893706
5	0.981974	0.991980	0.996362	0.970555	0.087301	0.883254
8	0.983985	0.991514	0.993743	0.969530	0.110885	0.858645
7	0.970966	0.992697	0.987945	0.952254	0.114687	0.837568
9	0.968474	0.992387	0.981984	0.943786	0.122173	0.821612
4	0.995103	0.995798	0.999931	0.990853	0.032268	0.958585
5	0.997159	0.997932	0.999835	0.994933	0.045035	0.949898
6	0.992674	0.995730	0.999129	0.987574	0.042049	0.945525
8	0.995860	0.997659	0.998823	0.992359	0.066711	0.925648
7	0.993059	0.994422	0.999652	0.987176	0.063955	0.923221
12	0.998374	0.998685	0.998801	0.995866	0.084494	0.911372
4	0.975602	0.984157	0.998649	0.958849	0.069029	0.889820
5	0.950794	0.978854	0.993289	0.924443	0.085033	0.839410
6	0.936486	0.977358	0.985549	0.902056	0.099062	0.802994
8	0.909088	0.976108	0.966894	0.857990	0.132679	0.725311
7	0.911237	0.945609	0.986189	0.849773	0.135793	0.713980
9	0.870839	0.971169	0.943857	0.798249	0.145262	0.652987

Table 2:Criteria and correlations between dependencies F_1, F_2 and f_{mean}

k	Ks_1	Ks_2	$K_{F\text{-}mean}$	$K_{general}$	σ_{error}	K_{cr}
5	0.966255	0.993619	0.977030	0.938036	0.169755	0.768281
4	0.946915	0.989261	0.965179	0.904128	0.160517	0.743611
8	0.980900	0.971254	0.966336	0.920632	0.184829	0.735803
4	0.935289	0.992767	0.943770	0.876314	0.150005	0.726309
4	0.984391	0.989566	0.998379	0.972541	0.110054	0.862487
5	0.980355	0.992146	0.993775	0.966601	0.140807	0.825793
6	0.972957	0.995370	0.978815	0.947934	0.126091	0.821844
9	0.978469	0.994141	0.986729	0.959827	0.148901	0.810926
4	0.676350	0.932902	0.805230	0.508075	0.218119	0.289956
9	0.764952	0.919753	0.726896	0.511420	0.223537	0.287883
10	0.758457	0.898424	0.594711	0.405245	0.237445	0.167800
8	0.884147	0.689431	0.606565	0.369736	0.246280	0.123456

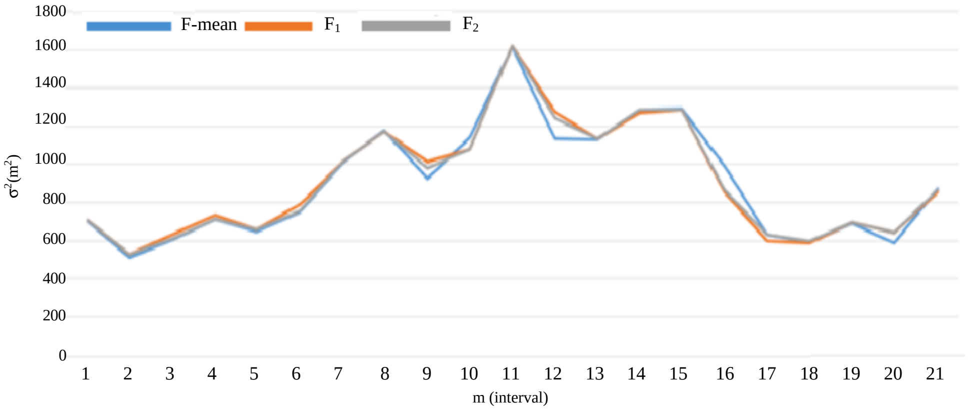

Fig. 10: Autocorrelation function of deviations $D_{residual}$ (Series No. 1) (Standard errors are white-niose estmates)

Table 2 shows the correlations between these dependencies. Figure 10 and 11 shows the relationship between the F_1, F_2 and f_{mean} dependencies in the case of the series No. 2 for k = 4.

Studies have shown that $D_{residual}$ is a stationary process in any of the 4 intervals for all studied processes, since, its mathematical expectation (mean) does not depend on time and has a finite variance. The stationary process $D_{residual}$ is close to white noise but is not a white noise, although its means are close to zero as well as the correlation between its members.

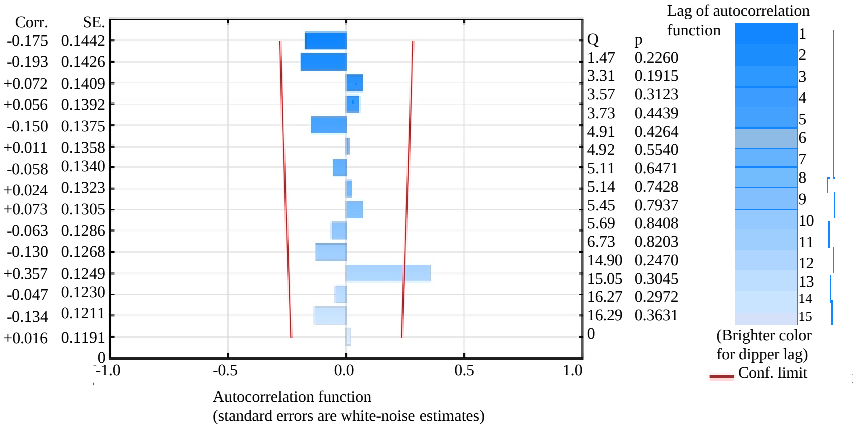

Fig. 11: Correlation between F_1, F_2 and f_{mean} in the intervals of the series No. 2, k = 4 after the secont transformation

RESULTS AND DISCUSSION

Our studies confirm the put forward hypothesis about finding and describing hidden dependencies of considered processes on the intervals. Their stability is confirmed by a high degree of correlation on the each considered interval. Dependencies indicate the presence of internal relations on the existing fragments of the process. Therefore, one can assume that similar relations are preserved on other, unknown intervals.

Time-series fractal property identification is one of the possible methods of analysis. Fractal time series can be found in various natural and scientific processes: solar activity (Scafetta and West, 2003), human body functioning (Diaz *et al*., 2015), noise of various electronic devices (Wolf, 1978).

The presence and wide prevalence of fractal properties in time series of completely different origin and statistical properties allow suggesting that there are internal laws which presence is not obvious and is not always possible to identify by usual methods. Even an incomplete knowledge of the internal laws of process formation can allow solving many practical problems and developing a unified universal approach for their solution.

The proposed strategy makes it possible to form a method for process description based not on the repetitive geometric form but on its internal relations. This approach is designed to eliminate one of the drawbacks inherent in classical models and to avoid the situation when results of forecasting or restoring the necessary fragments are not correct despite the high accuracy of approximation by the target process model.

The constant nature of these three dependencies and their presence on the intervals of the studied process do not allow drawing the firm conclusions about the presence of analogous dependencies on the intervals of another kind of processes. This issue is the subject of individual research and is beyond the scope of this study.

CONCLUSION

All of the studies have showed a stable pattern on different intervals of the studied processes. We have confirmed that it is possible to find a set of intervals for each studied process that will satisfy all of the presented dependencies. The rest of the process fragments obtained in the course of the research can be classified as stationary ones with a finite variance. Consequently, their contribution can be neglected and considered as "white noise". Stability of dependency fulfillment on the chosen intervals allows forming an analytical description of the process, based on which one could restore a part of intervals or form the basis of forecasting methods.

REFERENCES

Alexandridis, A.K., M. Kampouridis and S. Cramer, 2017. A comparison of wavelet networks and genetic programming in the context of temperature derivatives. Intl. J. Forecasting, 33: 21-47.

Baheti, A. and D. Toshniwal, 2014. Trend analysis of time series data using data mining techniques. Proceedings of the IEEE International Congress on Big Data (BigData Congress), June 27-July 2, 2014, IEEE, Anchorage, Alaska, USA., ISBN:978-1-4799-5057-7, pp: 430-437.

Davydenko, A. and R. Fildes, 2016. Forecast Error Measures: Critical Review and Practical Recommendations. In: Business Forecasting: Practical Problems and Solutions, Gilliland, M., L. Tashman and U. Sglavo (Eds.). John Wiley & Sons, New York, USA., pp: 238-249.

Diaz, M.H., F.M. Cordova, L. Canete, F. Palominos and F. Cifuentes *et al.*, 2015. Order and chaos in the brain: Fractal time series analysis of the EEG activity during a cognitive problem solving task. Procedia Comput. Sci., 55: 1410-1419.

Ericsson, N.R., 2017. Economic forecasting in theory and practice: An interview with David F. Hendry. Intl. J. Forecasting, 33: 523-542.

Filik, U.B. and M. Kurban, 2007. A new approach for the short-term load forecasting with autoregressive and artificial neural network models. Intl. J. Comput. Intell. Res., 3: 66-71.

Jiann, H.C., 2005. Seasonal adjustment of time series. Stat. Singapore Newsl., 1: 11-14.

Kawash, J., N. Agarwal and T. Ozyer, 2017. Prediction and Inference from Social Networks and Social Media. Springer, Berlin, Germany, ISBN:978-3-319-51048-4.

Morariu, N., E. Iancu and S. Vlad, 2009. A neural network model for time-series forecasting. Rom. J. Econ. Forecasting, 12: 213-223.

Petrica, A.C, S. Stancu and V. Ghitulescu, 2017. Stationarity-the central concept in time series analysis. Intl. J. Emerging Res. Manage. Technol., 6: 6-16.

Scafetta, N. and B.J. West, 2003. Solar flare intermittency and the earth's temperature anomalies. Phys. Rev. Lett., Vol. 90,

Schroeder, M., 2005. Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise. Dover Publication, Mineola, New York, USA., Pages: 528.

Schuermann, T., 2014. Stress testing banks. Intl. J. Forecasting, 30: 717-728.

Weron, R., 2014. Electricity price forecasting: A review of the state-of-the-art with a look into the future. Intl. J. Forecasting, 30: 1030-1081.

Wolf, D., 1978. Noise in Physical Systems. Springer, New York, USA., Pages: 337.