DOI: 10.5604/01.3001.0010.8639

GENERALIZED APPROACH TO HURST EXPONENT ESTIMATING BY TIME SERIES

Lyudmyla Kirichenko[1], Tamara Radivilova[2], Vitalii Bulakh[1]
[1]Kharkiv National University of Radioelectronics, Deptartment of Applied Mathematics, [2]Kharkiv National University of Radioelectronics, Deptartment of Infocommunication Engineering

Abstract. *This paper presents a generalized approach to the fractal analysis of self-similar random processes by short time series. Several stages of the fractal analysis are proposed. Preliminary time series analysis includes the removal of short-term dependence, the identification of true long-term dependence and hypothesis test on the existence of a self-similarity property. Methods of unbiased interval estimation of the Hurst exponent in cases of stationary and non-stationary time series are discussed. Methods of estimate refinement are proposed. This approach is applicable to the study of self-similar time series of different nature.*

Keywords: self-similar stochastic process, time series, Hurst exponent

UOGÓLNIONE PODEJŚCIE DO ESTYMACJI WYKŁADNIKA HURSTA NA PODSTAWIE SZEREGÓW CZASOWYCH

Streszczenie. *W pracy przedstawiono uogólnione podejście do analizy fraktalnej samopodobnych procesów losowych przedstawianych w krótkich szeregach czasowych. Zaproponowano kilka etapów analizy fraktalnej. Wstępna analiza szeregów czasowych obejmuje eliminację krótkoterminowej zależności, identyfikację prawdziwej długoterminowej zależności oraz weryfikację hipotezy o istnieniu własności samopodobieństwa. Uwzględniono metody bezstronnej oceny przedziału czasowego wykładnika Hursta w przypadku stacjonarnych i niestacjonarnych szeregów czasowych. Zaproponowano metody walidacji uzyskanego oszacowania wykładnika Hursta. To podejście ma zastosowanie do badania samopodobnych szeregów czasowych o różnym charakterze.*

Słowa kluczowe: samopodobny proces stochastyczny, szeregi czasowe, wykładnik Hursta

Introduction

The tasks of modern nonlinear physics, radio electronics, control theory and image processing require the development and application of new mathematical models, methods and algorithmic support for data analysis. In recent decades, it has been discovered that many stochastic processes in nature and technology have long-term dependence and fractal structure. The most suitable mathematical apparatus for studying the dynamics and structure of such processes is fractal analysis.

Stochastic process $X(t)$ with continuous real time is called to be self-similar of parameter *H*, 0 <*H*<1, if for any value $a > 0$ processes $X(at)$ and $a^H X(t)$ have same finite-dimensional distributions:

$$\text{Law}\{X(at)\} = \text{Law}\{a^H X(t)\}. \tag{1}$$

The notation Law{*} means finite distribution laws of the random process. Parameter *H* is called Hurst exponent. It is a measure of self-similarity. Along with this property, Hurst exponent characterizes the measure of the long-term dependence of stochastic process, i.e. the decrease of the autocorrelation function $r(k)$ in accordance with the power law: $r(k) = k^{-\beta}$, $H = 1 - \beta/2$.

For values 0.5 <*H*< 1 the time series demonstrates persistent behaviour. In other words, if the time series increases (decreases) in a prior period of time, then this trend will be continued for the same time in future. The value *H* = 0.5 indicates the independence (the absence of any memory about the past) of time series values. The interval 0 <*H*< 0.5 corresponds to antipersistent time series: if a system demonstrates growth in a prior period of time, then it is likely to fall in the next period.

Information data flows in telecommunication networks were among the first real stochastic processes, where self-similar properties were discovered. For self-similar traffic, calculating methods of computer network characteristics (channel capacity, buffer capacity, etc.) based on classical models do not conform to necessary requirements and do not adequately assess the network load. Many publications are devoted to the analysis of the fractal traffic properties and their impact on the functioning and quality of service of the telecommunications network [7, 21–24].

Typical examples of fractal stochastic structures are the modern financial markets. The hypothesis of fractality of financial series assumes that the market is a self-regulating macroeconomic system with feedback, using information about past events that affect decisions in the present, and containing long-term correlations and trends. The market remains stable as long as remains its fractal structure. Analyzing the occurrence of time intervals with the different fractal structure, it is possible to diagnose and predict unstable market conditions (crises) [6, 18, 20].

Numerous studies have shown that many bioelectrical signals have a fractal structure. Distinct changes of the fractal characteristics of cardio- and encephalograms appear in various diseases, with changes in mental and physical stress on the body. Fractal analysis of bioelectric signals can be the basis for conducting statistical studies, what will allow to formulate methods that will be significant for clinical practice [3, 4, 10].

Obviously, that the evaluation of the Hurst exponent for experimental data plays an important role in the study of processes having self-similarity properties. There are many methods for estimating the self-similarity parameter, each of which bears the imprint of that area of scientific applications where it was originally developed [5, 13, 21, 23]. In practice, the methods of rescaled range, variance-time analysis and the detrended fluctuation analysis are most often used to estimate the Hurst exponent. Methods based on wavelet transform are particular important among the research methods of fractal non-stationary processes. The basic ideas of wavelet-fractal methods of analysis are formulated in [1, 9].

In recent years, fractal time series analysis has become very popular. However, at the present time, there is no universal approach to estimating fractal characteristics based on a preliminary study of the time series structure. The major drawbacks in the application of fractal analysis methods are absence of a preliminary study of the correlation structure of the process, the use of only one method of analysis, a weak study of the statistical properties of estimates obtained from time series of short length.

The goal of the work is to present generalized use of fractal analysis techniques to study the time series of small length, using special methods of preliminary data research.

1. The main methods of estimating the Hurst exponent

Moments of the q-th order of the self-similar random process (1) can be expressed as follows:

$$\mathrm{M}\left[\left|X(t)\right|^q\right] \propto t^{qH}. \quad (2)$$

In fact, all methods of estimating the self-similarity parameter at the time series, are based on relation (2) with the value $q = 2$. The method of rescaled range (R/S-analysis) was proposed by H.Hurst [11] and it is still one of the most popular in the study of fractal series of different nature [8]. It is based on scaling relationship

$$M\left[R(\tau)/S(\tau)\right] \propto \tau^H,$$

where $R(\tau)$ is the range of the cumulative deviate series $x^{cum}(t,\tau)$, $S(\tau)$ is standard deviation of the initial series.

Variance-time analysis is most often used to processes researches in telecommunication networks [21, 22, 24]. It is based on the fact that the variance of the aggregated time series

$$x_k^{(m)} = \frac{1}{m}\sum_{t=km-m+1}^{km} x(t), \; k = 1,...,N/m \quad (3)$$

satisfies scaling relation

$$Var(x^{(m)}) \propto \frac{Var(x)}{m^\beta},$$

where $H = 1 - \frac{\beta}{2}$.

The method of detrended fluctuation analysis (DFA), originally proposed in [19], is currently the main method for determining self-similarity by non-stationary time series [12, 13]. In the DFA method, the fluctuation function $F(\tau)$ is calculated:

$$F^2(\tau) = \frac{1}{\tau}\sum_{t=1}^{\tau}(y(t) - Y_m(t))^2,$$

where $Y_m(t)$ is the local m-polynomial trend. The averaged on the whole of the time series function $F(\tau)$ have scaling dependence on the length of the segment:

$$F(\tau) \propto \tau^H.$$

The wavelet estimation of the Hurst exponent is based on the properties of the detailing wavelet coefficients, which at each decomposition level j also have self-similarity. The method of wavelet estimation [2] is based on the fact that the wavelet energy E can be written as the scaling relation.

$$E_j \propto 2^{(2H+1)j}.$$

2. Generalized approach to Hurst exponent estimation

In [15, 16] a comparative analysis of the statistical characteristics of the Hurst parameter estimates by time series of short length was performed. Summarizing the results of the research, we can suggest the following scheme for fractal analysis of some random process, represented by a time series of length N. In the main stages of fractal analysis, the methods of the rescaled range, DFA, and wavelet estimation are used. Since the application of the wavelet transform requires the appropriate software and work experience, the use of wavelet estimation methods is desirable, but not an obligatory element. However, the use of the DFA method is necessary for two reasons: this method has sufficient accuracy and it is designed to non-stationary time series. Consider step-by-step implementation of generalized approach to estimating the fractal properties of self-similar time series.

2.1. Preliminary study of the time series

1. Before starting the fractal analysis of time series, it is necessary to find out from a priori known information whether the series is cumulative $X_k^{cum} = \sum_{i=1}^{k} x_i$, for example, the currency rate in Fig. 1 (top) or it is a series of increments, for example, data traffic in Fig. 1 (bottom). If, by its nature, the series is cumulative, the following stages of fractal analysis refer to the corresponding series of increments $x_i = X_{k+1}^{cum} - X_k^{cum}$.

2. Determination of intervals of different scaling.

If the self-similar process has several scales that depend on time intervals, then on each such interval the series dynamics is determined by the corresponding Hurst exponent. To determine such intervals, it is necessary to consider the Hurst exponent as a function of time $H(\tau) = f[\log\frac{R}{S}(\tau)]$. This approach is possible in applying the rescaled range method, when time intervals change in small increments [17, 20].

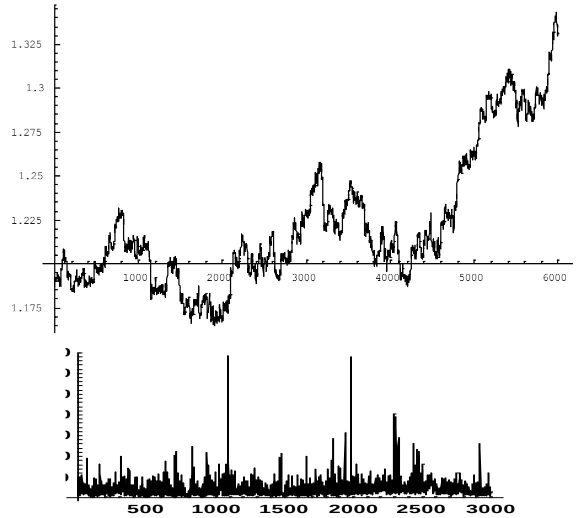

Fig. 1. Cumulative time series: currency rate (top), increments series: data traffic in telecommunication network (bottom)

Fig. 2 shows the dependence $H(\tau)$ for hourly data of the exchange rate. The time intervals which have presence ($0.5 < H < 1$) and absence ($H = 0.5$) of long-term dependence are distinctly different.

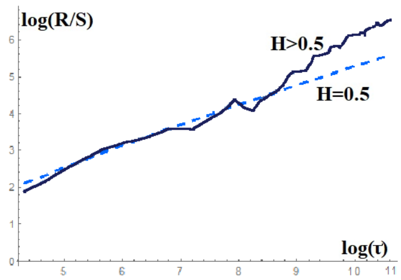

Fig. 2. Hourly data of the exchange rate: time intervals, where there are presence and absence of long-term dependence

In addition to the rescaled range method, it is possible to use the DFA method to get the fluctuation function $F(\tau)$ to determine the intervals of different scales. If there are several scales, the function changes the angle of slope. However, unlike the rescaled range method that investigates the long-term dependence along the entire length of the time, the fluctuation function can be correctly constructed only on the interval of values $N/4$ [12].

3. Identification and removal of short-term autoregressive dependence.

R/S-analysis and DFA allow to discover and eliminate the short-term dependence which is characteristic of autoregressive processes. The autoregressive dependence increases value of the Hurst exponent and demonstrates a false long-term memory [17, 20]. Therefore, when clarifying the fractal structure of the time series, it is first necessary to find out the existence of a short-term correlation. To do this, we need to regress the values $x(t)$ as the dependent variable against $x(t-1)$ and find linear dependence between them: $S(t) = x(t) - (a + b \cdot x(t-1))$. The significance of the coefficient b indicates the presence of short-term dependence. To resolve it, the residual series is determined: $S(t) = x(t) - (a + b \cdot x(t-1))$.

After this, fractal analysis of the residual series $S(t)$ is carried out. If the initial series $x(t)$ has a long-term dependence, then it remains, while the short-term dependence is eliminated. If the autoregressive correlation is significant, then all of the above fractal analysis steps relate to the residual time series.

The fluctuation functions $F(\tau)$ of the EEG realization obtained before and after the removal of the autoregressive dependence are shown in Fig. 3. In this case $H_1 = 0.78,\ H_2 = 0.63$.

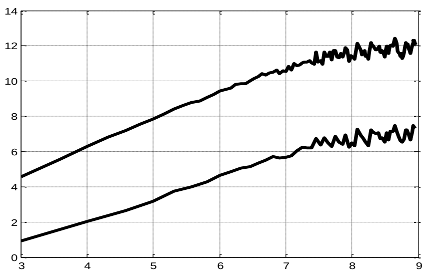

Fig. 3. Fluctuation functions before the removal of the autoregressive dependence (on top) and after the removal (bottom)

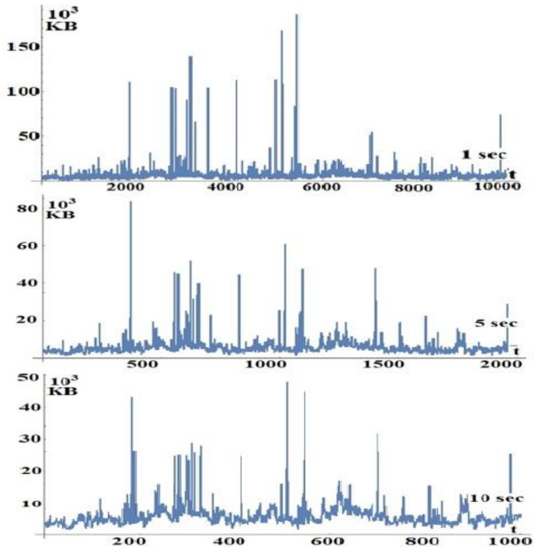

Fig. 4. Aggregated time series of telecommunication traffic

4. Testing the hypothesis about the presence of self-similarity.

If the Hurst parameter *H* is close to 0.5, it is necessary to test the self-similarity hypothesis. As a null hypothesis, it is usually postulate that random increments are independent. In [20], the criteria and areas for accepting this hypothesis are presented.

A qualitative test of the existence of the properties of statistical self-similarity is the construction of aggregated time series (3), for which sample distribution functions are calculated. In the case of the self-similarity of the time series $x(t)$ the aggregated series have the same distribution, confirmed by statistical criteria. Fig. 4 demonstrates the aggregated time series of telecommunication traffic. It is obvious that at all levels of aggregation the distribution functions of time series have heavy tails.

2.2. Estimation of the Hurst exponent of a stationary series

To estimate the Hurst exponent it is necessary to determine whether the series $x(t)$ is stationary by known statistical methods. If the series is stationary, then the Hurst exponent and confidence interval estimation of the *H* can be determined by the above or others methods.

The results of the researches [2, 12, 15] showed that the estimates of the Hurst parameter are biased. For each method, the bias and the mean square deviations of estimates depend on the time series length and decrease with increasing series length. Fig. 5 shows the values of the mean estimate bias for each method, depending on the length of the time series. Estimates calculated by the DFA and wavelet transform methods have the minimum bias. The estimates obtained by the wavelet method have a much smaller deviations than others. Wavelet estimates significantly depend on the mother wavelet [2, 14].

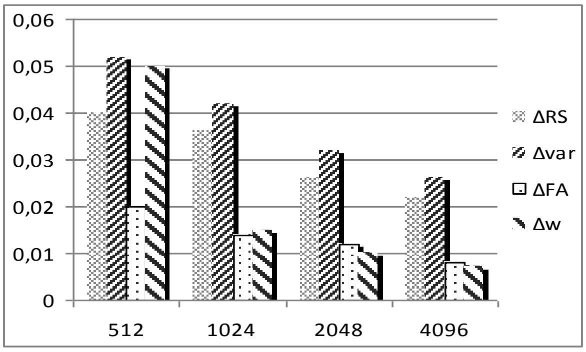

Fig. 5. Mean estimate bias for each method

The sample distribution laws of estimates of the Hurst parameter were investigated and it was shown that they have a normal distribution [2, 12, 15, 16, 20]. In this way, estimate of the Hurst exponent can be represented by confidence interval within which the true value *H* is found:

$$\hat{H} + \Delta - t_\alpha S < H < \hat{H} + \Delta + t_\alpha S\,, \tag{4}$$

where $\hat{H} = \hat{H}(N,\ method)$ is obtained evaluate of *H*; *N* is the time series length; *method* is the chosen method of estimation; $\Delta = \Delta(N,\ method)$ is the calculated mean bias of the estimate, $S = S(N, method)$ is the calculated standard deviation; α is required significance level; t_α is the quantile of the simple normal distribution [15].

2.3. Estimation of the Hurst exponent of a nonstationary series

If the series $x(t)$ is non-stationary, then the correct estimate of the Hurst parameter can be determined by the DFA method or by wavelet estimation.

1) First it is necessary to study the structure of the series using the correlation function, spectral density or spectrum of wavelet energy, which allow to identify the trend and cyclic components of the initial series.
2) When evaluating the Hurst exponent by method of DFA, it is necessary at first to make rough estimate using local polynomial trends of increasing degree and determine the smallest polynomial degree from which the Hurst parameter

estimate stops to change [12]. After this, to evaluate the self-similarity of the time series it is necessary to delete the local polynomial trend of the found degree. Fig. 6 presents the fluctuation functions $F(\tau)$ for a model fractal series with quadratic trend. After removing the local polynomial trend of the order of the larger two, the values of the Hurst index estimates cease to change.

3) The wavelet estimation of a non-stationary time series can be carried out according to the methods presented in [1, 2]. In this case, the evaluation of the exponent H depends essentially on the chosen mother wavelet.

2.4. Refinement of the evaluate of Hurst parameter

The analysis of the correlation dependence between the Hurst parameter estimates obtained by different methods showed that the sample correlation coefficients have absolute values less than 0.5. The correlation of wavelet estimates with ones calculated by other methods is insignificant. Therefore, the arithmetic mean of unbiased estimates obtained by several estimation methods can be used to increase the accuracy of the estimation.

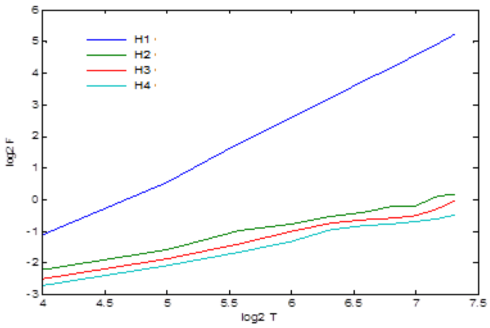

Fig. 6. Fluctuation functions with different local polynomial trends

To improve the accuracy of wavelet estimation, the comparative analysis of the statistical characteristics of the estimates calculated using different wavelet functions was carried out [2, 14]. The correlation analysis of the wavelet estimates of the Hurst exponent, obtained using different wavelets, showed that the more accurate estimation of the Hurst parameter is the arithmetic average of the estimates calculated through several different wavelet functions.

3. Conclusion

The paper offers the generalized approach to the analysis of fractal properties of time series. The proposed method involves a preliminary study of the structure of the time series, unbiased interval estimation of the self-similarity parameter and the joint use of several methods of fractal analysis. This makes it possible to increase the reliability of the Hurst exponent estimates. This approach is applicable to the study of self-similar time series of different nature: telecommunication traffic, financial indicators, biomedical signals, etc.

References

[1] Abry P., Flandrin P., Taqqu M.S., Veitch D.: Self-similarity and long-range dependence through the wavelet lens. Theory and applications of long-range dependence, Birkhäuser 2003.
[2] Abry P., Veitch D.: Wavelet analysis of long-range dependent traffic. IEEE/ACM Transactions Information Theory, Vol. 1(44), 1998.
[3] Bassingthwaighte J.B., Liebovitch L.S., West B.J.: Fractal Physiology. Oxford University Press, New York 1994.
[4] Ching E.S.C., Tsang Yue-Kin : Multifractality and scale invariance in human heartbeat dynamics. Physical Review E, Vol. 76, 2007, 041910.
[5] Clegg R.G.: A practical guide to measuring the Hurst parameter. Computing science technical report 2005. Vol. CS–TR–916.
[6] Cont R., Tankov P.: Financial modelling with jump processes. Chapman & Hall CRC Press, 2004.
[7] Czarkowski M., Kaczmarek S., Wolff M.: Influence of Self-Similar Traffic Type on Performance of QoS Routing Algorithms. INTL Journal of electronics and telecommunications, Vol. 62, no. 1, 2016, 81–87.
[8] Feder J.: Fractals. Plenum, New York 1988.
[9] Flandrin P., Gonzalves P., Abry P.: Scaling, Fractals and Wavelets. John Wiley & Sons, London 2009.
[10] Harikrishnan K.P., Misra R., Ambika G: Can the multifractal spectrum be used as a diagnostic tool? Chaotic Modeling and Simulation, Vol. 1, 2013, 51–57.
[11] Hurst H.E., Black R.P., Simaila Y.M.: Long-Term Storage: An Experimental Study. Constable, London 1965.
[12] Kantelhardt J.W., Koscielny-Bunde E., Rego H.H.A., Havlin S., Bunde A.: Detecting long-range correlations with detrended fluctuation analysis. Physica A 295, 2001, 441–454.
[13] Kantelhardt J.W.: Fractal and Multifractal Time Series. 2008. http://arxiv.org/abs/0804.0747 (available: 12.10.2017).
[14] Kirichenko L., Deineko Zh.: Estimation of the self-similarity of stochastic time series by the wavelet analysis method. Radio-electronic and computer systems, Vol. 4 (38), 2009, 99–105.
[15] Kirichenko L., Radivilova T., Deineko Zh.: Comparative Analysis for Estimating of the Hurst Exponent for Stationary and Nonstationary Time Series. Information Technologies & Knowledge, Vol. 5, No 4, 2011, 371–388.
[16] Kirichenko L., Radivilova T.: Comparative analysis of statistical properties of the Hurst exponent estimates obtained by different methods. Information Models of Knowledg. ITHEA, Kiev–Sofia 2010.
[17] Kirichenko L., Radivilova T.: Investigation of long-term dependence of network traffic using R / S-analysis. Automated control systems and automation devices, Vol. 135, 2006, 51–55.
[18] Meléndez G.R.: The fractal nature of bitcoin: evidence from wavelet power spectra. Fundacion Universidad de las Americas Puebla 39, 2014. http://ssrn.com/abstract=2547281 (available: 12.10.2017).
[19] Peng C.-K., Buldyrev S.V., Havlin S., Simons M., Stanley H.E., Goldberger A.L.: Mosaic organization of DNA nucleotides. Phys. Rev. E 49, 1994, 1685–1689.
[20] Peters E.E.: Fractal Market Analysis: applying chaos theory to investment and economics. Wiley, 2003.
[21] Sheluhin O.I., Smolskiy S.M., Osin A.V.: Self-similar processes in telecommunications. JohnWiley & Sons Ltd, Chichester 2007.
[22] Stollings W.: High-speed networks and Internets. Performance and quality of service. New Jersey 2002.
[23] Tsugawa S., Ohsaki H.: Emergence of Fractals in Social Networks: Analysis of Community Structure and Interaction Locality. 38th Annual Computer Software and Applications Conference, 2014
[24] Willinger W., Taqqu M.S., Erramilli A.A.: Bibliographical guide to self-similar traffic and performance modeling for modern high-speed network in ″Stohastic networks: theory and applications″. Claredon Press (Oxford University Press), Oxford 1996.

D.Sc. Lyudmyla Kirichenko
e-mail: lyudmyla.kirichenko@nure.ua

D.Sc., professor of Department of Applied Mathematics in Kharkiv National University of Radioelectronics. Major fields of scientific research: self-similar and multifracta time series analysis, wavelet analysis, chaotic systems. Author of nearly 150 publications in this research area.

Ph.D. Tamara Radivilova
e-mail: tamara.radivilova@gmail.com

Ph.D., associated professor of Department of Infocommunication Engineering in Technical Sciences at Kharkiv National University of Radioelectronics. Major fields of scientific research: self-similar and multifracta time series analysis, telecommunication network, cybersecurity. Author of nearly 100 publications in this research area.

M.Sc. Vitalii Bulakh
e-mail: bulakhvitalii@gmail.com

Graduate student at Kharkiv National University of Radioelectronics. Major fields of scientific research: self-similar and multifractal time series analysis, machine learning, data mining, chaotic systems.

otrzymano/received: 10.10.2017 *przyjęto do druku/accepted: 15.02.2018*