15

Singular-spectrum analysis and the normalized range method as a method for differentiation noisy experimental curves

© M.L. Bakhmutsky

Scientific Research Institute for System Analysis of the
National Research Centre „Kurchatov Institute“, Moscow, Russia
E-mail: mbakhmut@mail.ru

Received July 28, 2025
Revised August 19, 2025
Accepted September 25, 2025

It is proposed to differentiate the experimental curve by finite differences in physical space. In this case, the amplitude of the noise increases sharply, but this does not change the random nature of the noise. Using the singular-spectrum analysis, the derivative is presented as a sum of additive components. The fractal dimension (Hurst index) of these components is estimated by the normalized range method. When synthesizing the derivative, the components are summed until the Hurst index of the derivative begins to decrease.

Keywords: differentiation, singular-spectrum analysis, Hurst index.

DOI: 10.61011/TPL.2026.02.63030.20454

The need to differentiate a noisy experimental curve arises in various applications ranging from the analysis of results of hydrodynamic formation surveys to plasma tomography and spectrometry. In the present study, we propose a robust method for solving this ill-posed problem. This method is free from the ambiguity of almost universally accepted techniques and does not involve preliminary smoothing of the curve or increasing the distance between experimental points in calculation of the difference approximation of the derivative. The method of differentiating an experimental function consists in applying singular spectrum analysis (SSA) [1,2] to the difference approximation of the derivative in order to present it as a sum of additive adaptive components and calculating the fractal dimension of this sum using the normalized range method (R/S) [3]. Noise components are thus excluded from the sum, and only the physically significant part remains.

Let us assume that N values of function $\tilde{f}_i = f_i + \delta_i$ were measured at time points t_i ($i = 1{-}N$). The approximation of derivative of a second order in Δt at point t_i takes the form

$$\tilde{g}_i = \frac{\tilde{f}_{i+1} - \tilde{f}_{i-1}}{2\Delta t} = \frac{f_{i+1} - f_{i-1}}{2\Delta t} + \frac{\delta_{i+1} - \delta_{i-1}}{2\Delta t}, \qquad (1)$$

if

$$\Delta t \equiv t_{i+1} - t_i = \text{const}.$$

If $|\delta_i| \leqslant A$, the estimate of the derivative calculation error is $\frac{A}{\Delta t} \to \infty$ when $\Delta t \to 0$; i.e., difference differentiation leads to an enhancement of noise amplitude and jaggedness of the curve. The jaggedness of the curve may be estimated by the R/S method [3] with calculation of the Hurst index.

Step 1. Let us assume that we have a series of $\tilde{g}_i$ values calculated by formula (1): $\tilde{g}_1, \tilde{g}_2, \ldots, \tilde{g}_N$. Using a window m samples in length, we put this series into correspondence with the so-called trajectory matrix $\tilde{A} = \{\tilde{a}_{i,j}\}$, where $\tilde{a}_{i,j} = \tilde{g}_{i+j-1}$, $i = 1{-}m$, $j = 1{-}n$, $m \leqslant \left[\frac{N+1}{2}\right]$, $n = N + 1 - m$. In the present case, window $m = \left[\frac{N+1}{2}\right]$ is taken for definiteness,

$$\tilde{A} = \begin{pmatrix} \tilde{g}_1 & \tilde{g}_2 & \cdots & \tilde{g}_n \\ \tilde{g}_2 & \tilde{g}_3 & & \tilde{g}_{n+1} \\ & \vdots & \ddots & \vdots \\ \tilde{g}_m & \tilde{g}_{m+1} & \cdots & \tilde{g}_N \end{pmatrix}. \qquad (2)$$

This matrix allows one to reconstruct unambiguously the $\tilde{g}_j$ series by averaging using the formula

$$\tilde{g}_j = \begin{cases} \frac{1}{j}\sum\limits_{i=1}^{j} \tilde{a}_{i,j-i+1}, & 1 \leqslant j \leqslant m, \\ \frac{1}{m}\sum\limits_{i=1}^{m} \tilde{a}_{i,j-i+1}, & m \leqslant j \leqslant n, \\ \frac{1}{N-j+1}\sum\limits_{i=1}^{N-j+1} \tilde{a}_{i+j-n,n-i+1}, & n \leqslant j \leqslant N. \end{cases} \qquad (3)$$

We then calculate the singular value decomposition of matrix $\tilde{A} = U\Sigma V^T$, where U is a matrix with orthogonal columns of dimension m (left singular vectors), V is a matrix of right orthogonal singular vectors of dimension n, $\Sigma = \text{diag}\{\sigma_i\}$, $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_n \geqslant 0$ [4].

Step 2. We calculate the sequence of matrices $A^k = \sum\limits_{i=1}^{k} \sigma_i u_i \times v_i^T$, $k = 1{-}k_0$.

Here, u_i and v_i are the ith singular vectors of matrices U and V, $\times$ denotes outer (Kronecker) multiplication, and σ_i is the singular value. In what follows, σ_i, u_i, v_i are referred to as a singular triplet. Using formula (3), we calculate the g^k derivative approximation values corresponding to this matrix. Let us determine the fractal dimension of the series of derivative values calculated using k singular triplets. As

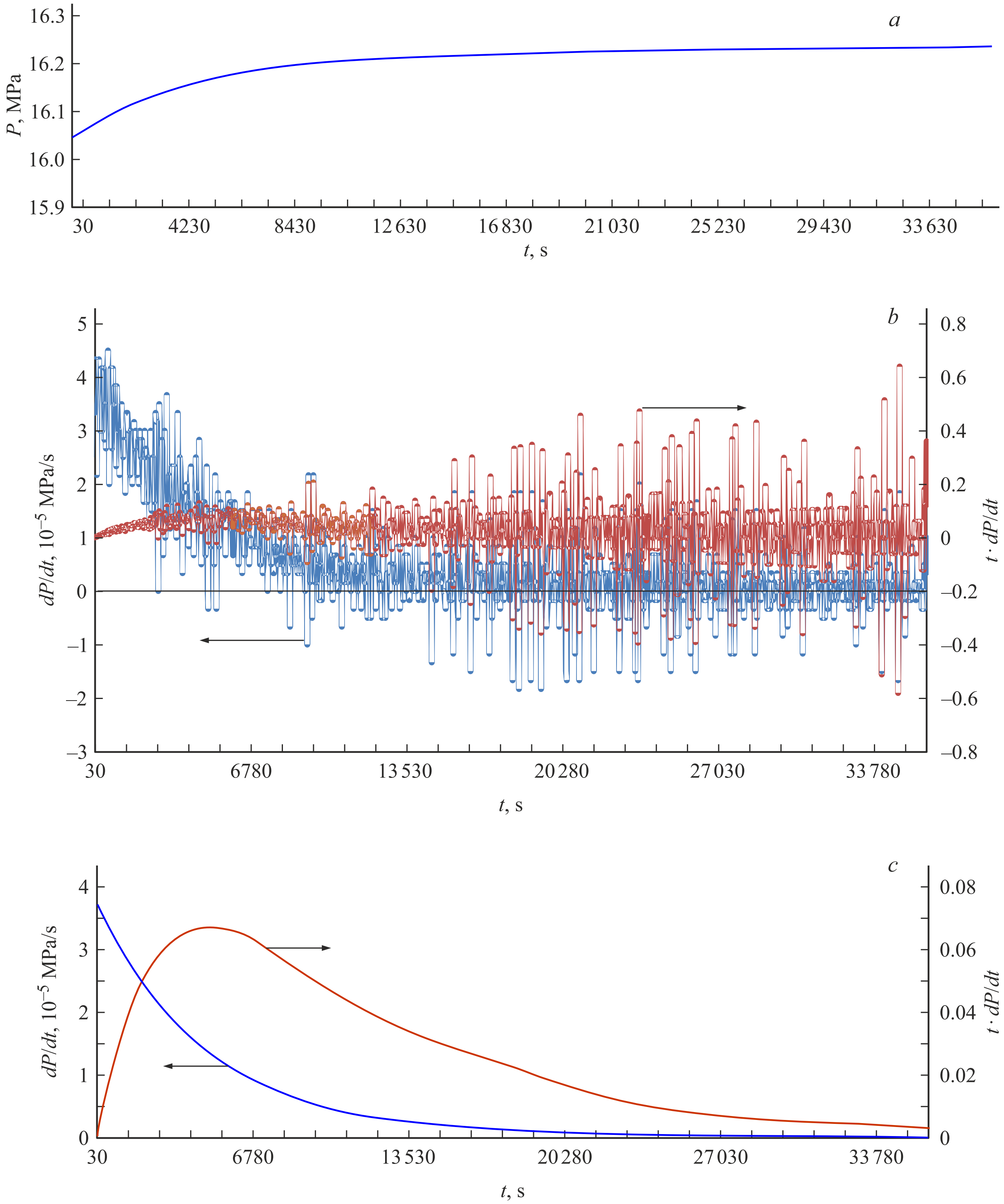

Figure 1. *a* — Pressure recovery curve of a well of one of the fields. Pressure and time are plotted on the ordinate and abscissa axes, respectively. *b* — Plot of the derivative calculated using formula $f'_i = \frac{f_{i+1}-f_{i-1}}{x_{i+1}-x_{i-1}}$. Pressure derivative, time, and logarithmic derivative are plotted on the ordinate, abscissa, and auxiliary ordinate axes, respectively. *c* — Plots of the derivative (primary ordinate axis) and the logarithmic derivative reconstructed from the first singular triplet.

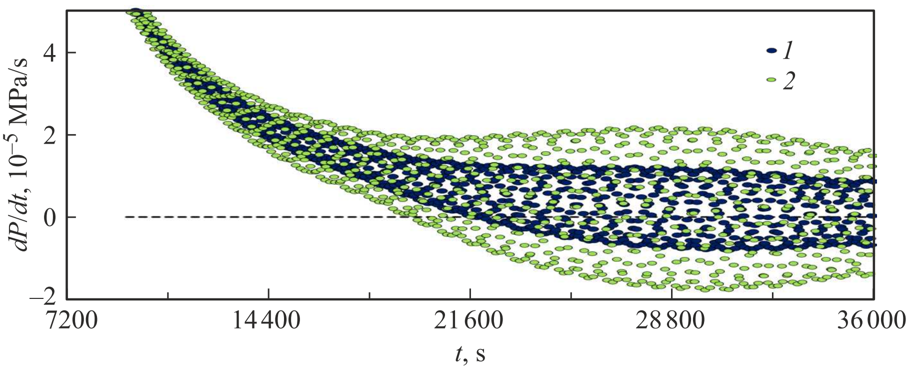

Figure 2. Plot of the derivative reconstructed from two (first and second) and the first three singular triplets.

is known, $D = 2-H^k$, where H^k is the Hurst index of this series [3].

Step 3. We calculate H^k via

$$\langle g^k \rangle = \frac{1}{N}\sum_{i=1}^{N} g_i^k, \qquad S = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(g_i^k - \langle g^k \rangle\right)^2},$$

$$z_m = \sum_{i=1}^{m}\left(g_i^k - \langle g^k \rangle\right), \quad R = \max_{1\leqslant m\leqslant N}(z_m) - \min_{1\leqslant m\leqslant N}(z_m),$$

$$R/S = \left(\frac{\pi}{2}N\right)^{H^k}, \qquad (4)$$

which yields $H^k = \frac{\lg(R/S)}{\lg(\frac{\pi}{2}N)}$.

The number of singular triplets used to reconstruct the trajectory matrix and derivative runs from 1 to k_0. The value of k_0 is determined from inequality $H^{k_0} > H^{k_0+1}$. The dependence of the Hurst index on the number of samples is of no importance in this case. Let us consider two examples of application of this algorithm for derivative calculation. The first example is the pressure recovery curve measured in the process of hydrodynamic testing of a real well. When one analyzes and interprets the results of hydrodynamic well surveys, the logarithmic derivative of the curve of pressure recovery over time is calculated [5]. Figure 1, *a* shows the pressure recovery curve. Its Hurst index is $H = 0.8059015$.

Figure 1, *b* shows the plots of the derivative calculated using formula (1) and the logarithmic derivative. The Hurst index of the derivative is $H = 0.783470335$. We have 1203 samples for the measured functions. Using a window 602 samples in size, we construct a trajectory matrix for the difference derivative, find its singular value decomposition, and sequentially reconstruct the derivative using formula (3). Figure 1, *c* shows the plots of the derivative and the logarithmic derivative reconstructed from the first singular triplet. The Hurst index of the reconstructed derivative is $H = 0.804599782$. Figure 2 presents the plots of the derivative reconstructed from two (*1*) and three (*2*) singular triplets. Plot *1* has a visible noise component, and this component is amplified in plot *2*. The Hurst indices of curves *1* and *2* are 0.804431406 and 0.803925077, respectively. Thus, a reduced Hurst index is indicative of noise amplification.

The maximum Hurst index corresponds to the derivative reconstructed from the first singular triplet, and this is the main result (Fig. 1, *c*). Let us consider the second example. Since filtration processes in a formation are characterized by linear or weakly nonlinear parabolic equations, the pressure variation curves are smooth. Therefore, our second example concerns the differentiation of a noisy oscillating model curve taken from [6]. The model problem was differentiation of function

$$f(t) = \sin(t) + \frac{1}{100}\sin(100t) + h(t), \quad t \in [0, \pi], \qquad (5)$$

2001 is the number of samples, and $h(t)$ is noise (random numbers distributed uniformly within the range from -0.01 to 0.01).

Figure 3, *a* shows the plots of the function and its derivative (auxiliary ordinate axis) calculated using formula (1). Differentiating the function with formula (1) and using a window with size $\left[\frac{N+1}{2}\right]$ (i.e., 1001 samples), we find the trajectory matrix and its singular value decomposition. Synthesizing the derivative values according to formula (3) and calculating the Hurst index, we find the function derivative based on four singular triplets. Figure 3, *b* presents the Hurst indices as functions of the number of singular triplets in the approximation of derivatives (upper curve) and the Hurst indices for each series component constructed from a singular triplet. The analytically calculated derivative (without noise) and the difference derivative of the noisy function are plotted in Fig. 3, *c*. Since the analytical and difference derivatives almost merge into a single curve, we perform a quantitative assessment of the closeness of the analytical and difference derivatives.

Let us compare the analytical derivative of a model function without noise and the result of calculating the derivative by the proposed method for the same function with noise. Let us denote the value of analytical derivative at

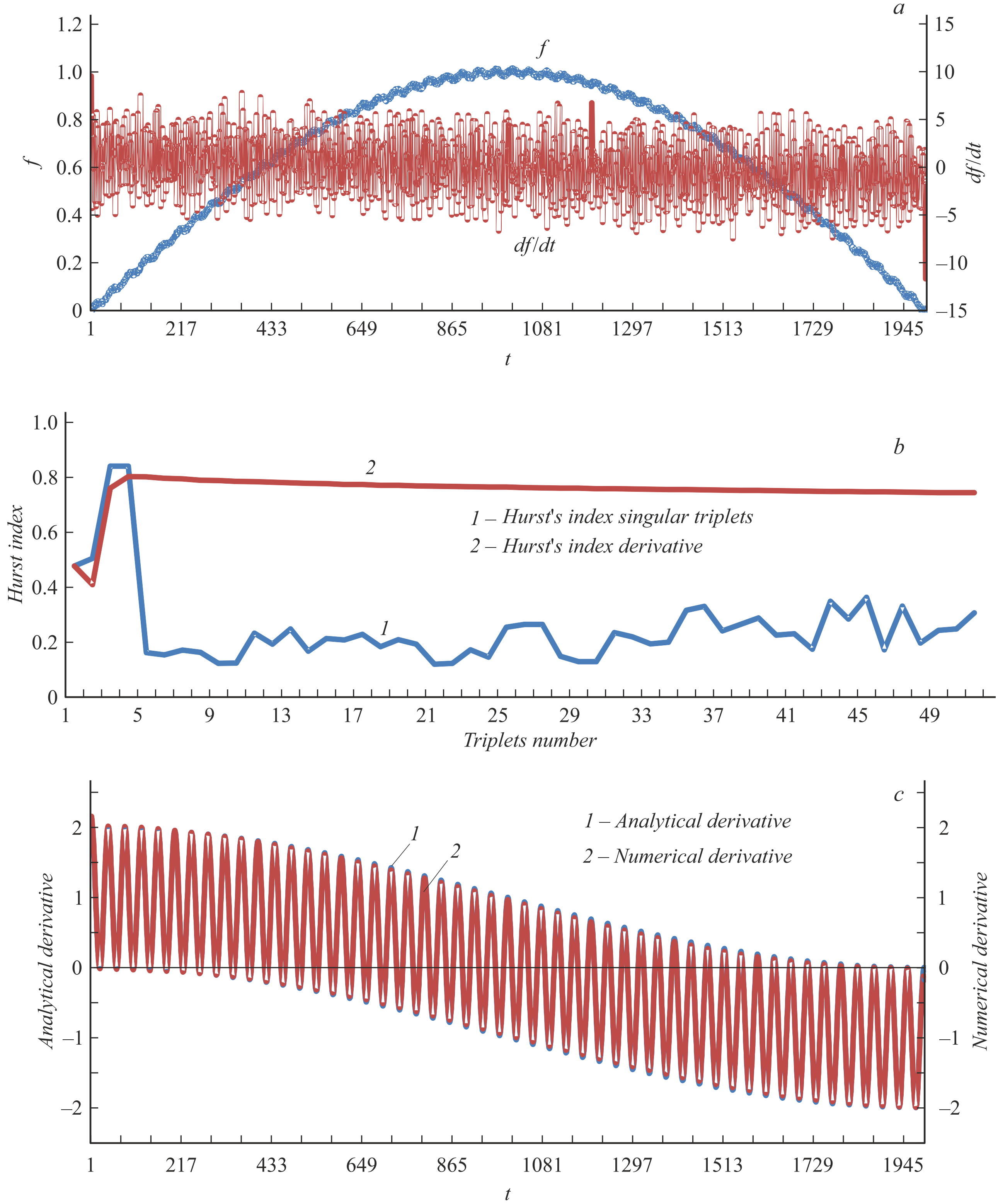

Figure 3. *a* — Model function (primary ordinate axis) and its derivative. *b* — Hurst indices of the synthesized derivative (upper curve) and derivative components calculated from singular triplets (lower curve). The singular triplet number is plotted on the abscissa axis. *c* — Comparison of analytical and difference derivatives. All the presented quantities are dimensionless.

point i (ith vector component) as g_i, the value of difference derivative at point i (ith component of the difference derivative vector) as $\breve{g}_i$, and the $w_i = g_i - \breve{g}_i$ difference vector as w_i. Then, $|g| = 44.74371464$, $|\breve{g}| = 44.26865779$, and $|w| = 1.138447732$. The cosine of the angle between the vectors of derivatives of the analytical and noisy curves is 0.999729803; i.e., the angle itself is 0.023246921 rad or $1.331989725°$.

$$\frac{|g| - |\breve{g}|}{|g|} \cdot 100\,\% = 1.061728687\,\%,$$

$$\frac{|w|}{|g|} \cdot 100\,\% = 2.544374649\,\%.$$

Thus, it is clear that direct difference differentiation with subsequent singular spectrum analysis controlled by the calculated Hurst index allows one to calculate correctly the derivative of a noisy experimental function.

Funding

This study was supported financially as part of state assignment No. FNEF-2024-0001 (Development and Implementation of Trusted Artificial Intelligence Systems Utilizing New Mathematical and Algorithmic Methods and Fast Computing Models Implemented on the Basis of Russian Computing Systems) of the Scientific Research Institute of System Analysis of National Research Center „Kurchatov Institute.“

Conflict of interest

The author declares that he has no conflict of interest.

References

[1] R. Vautard, P. Yiou, M. Ghil, Physica D, **58**, 95 (1992). DOI: 10.1016/0167-2789(92)90103-T
[2] N.E. Golyandina, *Metod „Gusenitsa“-SSA: analiz vremennykh ryadov* (Izd. VVM, SPb., 2004) (in Russian).
[3] J. Feder, *Fractals* (Plenum Press, 1988).
[4] J.W. Demmel, *Applied numerical linear algebra* (SIAM, 1997).
[5] R.N. Horne, *Modern well test analysis. A computer-aided approach* (Petroway, 2002).
[6] A.N. Tikhonov, V.Ya. Arsenin. *Metody resheniya nekorrektnykh zadach* (URSS, M., 2022).

Translated by D.Safin