Fractal Methods in Information Technologies of Processing and Analysis of Parametrically Related Data Streams

© Alexei V. Myshev   © Mikhail I. Turitsyn

National Research Nuclear University MEPhI (IATE),
Obninsk, Russia

mishev@iate.obninsk.ru

Abstract. In this work a new approach to the construction of models and logic circuits of algorithms and procedures for information technology processing and analysis of parametrically related and unrelated data streams within the fractal paradigm is describes. In this case, data streams are defined as information objects whose physical nature can be arbitrary. The information object is investigated apart from any model or scheme, the logical scheme of intellectual technology is built in the form of: facts, regularities and reality. Fractal methods form the framework of the logical, algorithmic and content essence of the approach. The basic premise of the approach is as follows. First, the processing and analysis of parametrically related or unrelated data stream to determine whether it forms a fractal structure and construct a phase portrait of the data stream as an information object. Second, to distinguish the areas of fractal percolation and aggregation in the multifractal structure of the stream, the phase portrait is used. Third, need to estimate the spatial and temporal scales of fractal percolation and aggregation processes.

Keywords: data stream, percolation function, fractal dimension, percolating fractal, fractal aggregate.

1 Introduction

In the paper a new approach to the construction of models and logic circuits of algorithms and procedures of information technology processing and analysis of parametrically related data streams within the fractal paradigmis is described. The methodology for constructing such models and schemes is based on the construction of the percolation function of a parametrically bound data stream and its information phase portrait. In this case, data streams are defined as information objects whose physical nature can be arbitrary. The information object is not investigated within any model or scheme, but the logical scheme of intellectual technology is built in the form of: facts, regularities and reality. Fractal methods form the framework of the logical, algorithmic and content essence of the approach.

The main premise of the content-semantic essence of fractal methods in information technology processing and analysis of large data streams is as follows. The first stage of data stream processing, is to calculate fractal dimension (geometric and universal) to determine whether it forms a multifractal structure. Second, if the initial data stream is a fractal object, then the stream of integer values of the percolation function is formed. Third, construct and analyze a phase portrait of the data stream; highlight in its structure the areas of fractal percolation and aggregation; estimate the degree of discrepancy between geometric and information fractal dimensions as an indicator of the unity of quantitative and qualitative characteristics of the stream. Fourth, estimate the spatial and temporal scales of fractal percolation and aggregation processes.

The fractal paradigm in the methodology of development and implementation of information technologies for processing, analysis and classification of large data streams, in contrast to traditional methods and methods [1,2,3] allows to take into account both the properties of regularity and irregularity of the structure of the state space of the information stream data scale, and their dynamic and information.

The paper is structured based on the following assumptions. First, identify and show the fractal properties of information objects. Second, formulate the task of processing and analyzing parametrically coupled data flows based on the fractal paradigm. Third, on the real results show the importance of this work for the development and implementation of IT–technologies for large data flows processing and analysis.

2 Data streams (processing, analysis and classification)

2.1 Fractal properties of data stream

The main and important point in the formulation of data flow fractal properties is the introduction of percolation function concept into the designated subject domain. This concept in this paper is defined as an attribute by means of which data stream fractal properties are denoted and described. It does not in any way correspond to the known probabilistic definition of the percolation function from percolation theory.

To obtain numerical estimates of the geometric measure of the fractal dimension of a parametrically related information object as a spatial structure, we used the well-known Hausdorff–Bezekovich formula [4,5], the classical definition of which is as follows.

Let the initial data stream form a metric set M, in which the λ-dimensional outer measure $l_\lambda(M)$ is defined as follows. Considered the ρ-covering of the set M, which is a countable covering of this set with S_i sphere of diameter $d_i < \rho$, introduce a measure

$$l_\lambda(M,\rho) = inf \sum_i d_i^\lambda, \qquad (1)$$

where the lower face is taken over all ρ–covers of the set M. There is a limit

$$l_\lambda(M) = \lim_{\rho \to 0} l_\lambda(M, \rho), \qquad (2)$$

finite or infinite, which, as a function of M, is an external measure.

The Hausdorff dimension *dim M* of the set *M* is determined by the behavior of $l_\lambda(M)$ not as a function of M, but as a function of λ:

$$dim\ M = sup\{\lambda: l_\lambda(M) = \infty\} = inf\{\lambda: l_\lambda(M) = 0\}, (3)$$

that is *dimM* – is the «transition point»: for *λ> dim M,* the value $l_\lambda(M) = 0$, and for *λ < dim M,* the value $l_\lambda(M)$ –is infinitely large.

Unfortunately, the fractals theory mathematical apparatus based on the fractal dimension of Hausdorff is little applicable to the description of time series and parametrically related information objects. Therefore, to identify patterns due to the properties of the time aspect and the parametric connection of the world events in the information space of the data stream, it is necessary, first of all, to determine the measure of fractal dimension parametric or temporal structure.

For these purposes, the formula for estimating the measure of time structures fractal dimension and Hurst statistics is used.

To estimate parametrically related data stream fractal dimension measure the following empirical law takes place [6]:

$$\frac{\langle R(m;\tau)\rangle_T}{\langle R(1;\tau)\rangle_T} = m^F, \qquad (4)$$

where$\langle R(m;\tau)\rangle_T$ – is a measure of the time or parametrically related information structure on the interval of time parameter change or the connection T_m, $\langle R(1;\tau)\rangle_T$ – for the interval of length *τ, m*– integer number, τ – the duration of the link of the time structure, *T* – the considered period of time. *F*–fractal dimension temporary or parametrically associated structures.

On the other hand, Hurst found that the normalized scope R / S for time dependences is well described by the empirical relation:

$$R/S = cN^H, \qquad (5)$$

where*R* – the scope of the change in the values of the data stream elements over the entire interval *T*, *S* –standard deviation, *N* – the multiplicity factor of the period *T*in standard units, *H* – Hurst index, *c* – constant.

The Hurst empirical law can be considered as a special case of the formula (4) for a parametrically related data stream structure. In this case, the following analogy is valid: the Hurst exponent can be considered as an analogue of the fractal dimension estimation *F* for *S* = 1. It should be noted that the fractal dimension *F* and Hurst index *H* are not sensitive to such artifacts and

phenomena as intermittency in a random medium and singular errors. This is largely due to the fact that the above approaches to the data streams fractal structure analysis do not reflect the information nature of the objects investigated. Based on these assumptions, we have obtained a formula for calculating the estimation of the measure of universal fractal dimension, which is a synergy of geometric and information dimension [7,8]:

$$d_b = \lim_{\varepsilon \to 0} \frac{\sum_{i=1}^{K} p_i \log \sum_{j=1}^{K}(1 - \rho_{ij})p_j}{\log \varepsilon}, \qquad (6)$$

where p_i – the probability of the *i–th* data stream element r_i hitting in the *i–th* subinterval of $\Delta = |r_{max} - r_{min}|$; ε – length of the subinterval for a given partition of the intervalΔ; ρ_{ij} – randomized metric between the centers of the *j–th* and *i–th* subintervals; r_{max} and r_{min} – the maximum and minimum values of the stream elements.

In the work two types of randomized metrics were considered in the work: geometric and informational. To calculate the geometric metric of the ρ_{ij} the following formula was used [7]:

$$\rho_{ij} = \frac{|r_i - r_j|}{|r|}, \qquad (7)$$

where $|r_i - r_j|$ – is the geometric distance between the *i–th* and *j–th* under the intervals; $|r|$ – is the length of the interval Δ. To calculate the information metric we used the ratio below:

$$\rho_{ij} = |\rho_i - \rho_j|. \qquad (8)$$

On the one hand, the above–described formulas for obtaining various estimates of measures of fractal dimensions are integral estimates of fractal properties of the stream of parametrically coupled data. On the other hand, they allow us to formulate and describe the problem of processing, analysis and classification of parametrically linked data stream within the fractal paradigm.

2.2 Problem statement

For processing parametrically connected data streams it is proposed to use mathematical and logical apparatus of fractal theory [5,7,8]. As a criterion of regularity of the data stream, the above-defined quantitative estimates of the geometric and universal fractal dimension are used. The main premise and the meaning of the criterion used is that the values of the estimates of the fractal dimension measure reflect the degree of "hole" of the initial data stream with respect to the information scale of measurement.

The primary procedure for processing the original data stream explained by the following logic diagram and algorithm.

First, the information scale for measuring the values of the elements of the original data stream is determined. In the channels of storage and transmission of information of information–measuring or computing

system any data stream is defined as an information object, i.e. binary set. In the language of digital technology, this means that the elements of this set can take two values: one or zero. On the information scale indicated above, either a numerical or information metric can be determined.

Secondly, digitization of the information scale procedure is implemented. The range and price of the scale division are digitized by elements of the natural series.

Third, a stream of integer values of the percolation function of the original data stream is formed. The percolation function reflects and describes the geometric structure of the "leaky information space" of the original data stream.

Fourth, the procedure of constructing a phase portrait of the data stream is implemented. The phase space is a plane on which the following coordinate system is determined, namely: the abscissa axis - percolation function values, the ordinate axis – digitized information scale values.

The relationship between the scale integer values and another scale of measurement of the source data elements stream is carried out by following attributes:

- general scale of variability of the values of the elements of the original array in any other non-integer algebraic system of their measurement;
- common scale division price in a non-integer algebraic system;
- number of significant digits.

The integer values of percolation function are determined by following ratio:

$$h_i = r_i - r_{i-1}, \tag{9}$$

where h_i – values of percolation function, r_i – the number of the interval in which the corresponding element of the original data stream falls.

The number of partitions L of interval $\Delta = |r_{max} - r_{min}|$ into subintervals is determined based on the following ratio:

$$L = \Delta / \varepsilon. \tag{10}$$

The main premise of the problem statement is to determine whether the processed and analyzed data stream is a fractal? If so, describe its fractal structure and calculate the integral estimates of fractal dimensionmeasures. The solution of problem is to obtain above estimates of fractal dimension measures and construct the phase portrait.

2.3 Fragments of processing and analysis

Two data sets for processing and analysis using fractal methodswere used. **In the first case**, data were presented by the stream of regular harmonic function values. **In the second case**, were used empirical data presented by the icm–20608 quadcopter gyroscope measurements.

To solve this problem, a software component that implements the above algorithm of data stream

processing was developed. The function y = 2.5sin(2πt) was used as a real harmonic function that induces the data stream. The volume of data stream was not less than ten thousand elements.

Total scale for the harmonic function varied within (–2.5;2.5).Two different scale division intervals *ε=0.01* and *ε=0.001* were taken. The number of partitions *L=500* and *L=5000* respectively. Gyroscope data stream values were changed in the range (–840;1270). Scale division intervals *ε=0.01* and *ε=0.001*.The number of partitions *L=211000* and *L=2110000* respectively.

As a calculations result, the following fractal dimension measure estimates were obtained, which was determined by the formula (6) for various metrics ρ_{ij}, which were determined by the formulas (7) and (8).

For the harmonic function (figure 1) the following numerical estimates of universal fractal dimension for various metrics ρ_{ij} were obtained:

1. Geometric metric:
   d_b = 0,950 for ε =0,01;
   d_b = 0,827 for ε =0,001;
2. Information metric:
   d_b = 1,060 for ε =0,01;
   d_b = 0,837 for ε =0,001.

The above measure of universal fractal dimension values d_b show that the harmonic function reflects the information set, which has a regular structure (the condition of regularity is $d_b \to 1$). Here the geometry of information set and its information compendency are in good agreement. This is well confirmed by the fact that the d_b values for geometric and information metrics are close for different ε values.
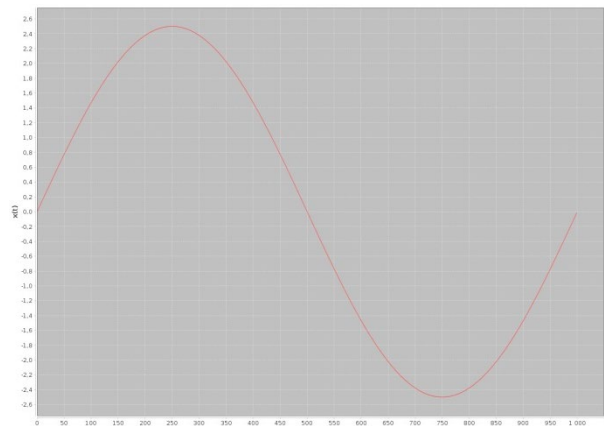

Figure 1 Graph of 2,5sin(2πt) function

Another information picture is observed for the gyroscope values, which are shown in figure 2. The numerical estimates of the universal fractal dimension for the set of gyroscope data using a geometric metric are given below:

d_b = 3.438 for ε =0,01;
d_b = 7.439 for ε =0,001.

The d_b modulus value with decreasing ε increases significantly, which is the criterion and indicator of irregularity. The designated information object is a data

stream with an irregular structure.

The values of percolation function 2,5sin(2πt) for different values of common scale and tick marks are shown in figures 3 and 4. As can be seen from figures 3 and 4 at lower values ε percolation function exhibits pronounced properties of regularity or continuity for 2,5sin(2πt) function calculated values variability.
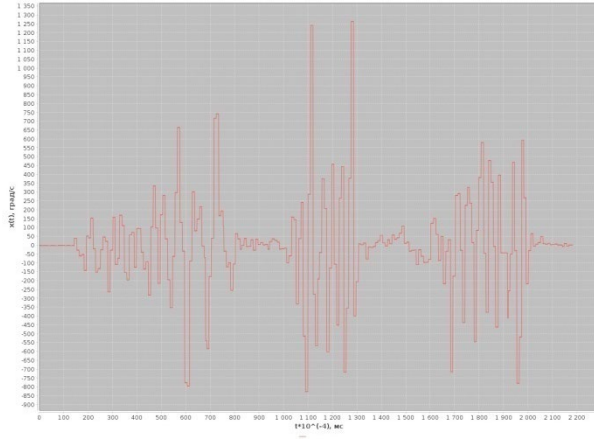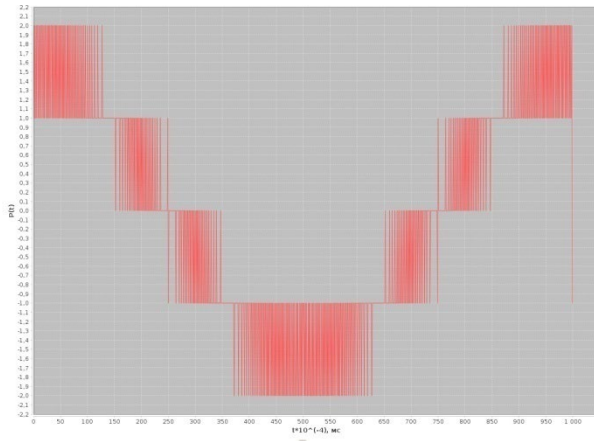

Figure 2 Gyroscope data graph
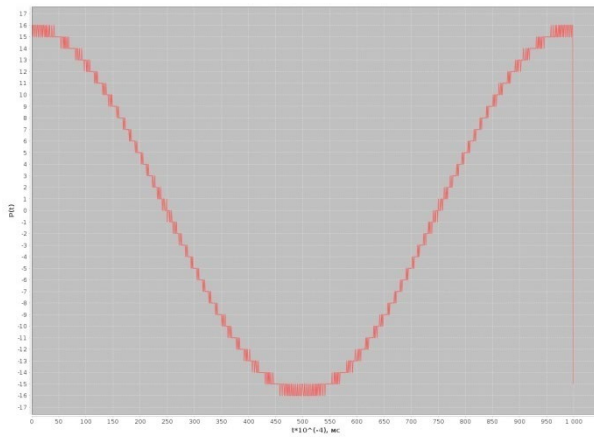

Figure 3 Percolation graph for 2,5sin(2πt), ε =0,01

Figure 4 Percolation graph for 2,5sin(2πt), ε =0,001

A geometric illustration of percolation function for gyroscope data stream is shown in figure 5. The graph of percolation function fully reflects the properties of irregularity and the data stream elements values singularity.

This stage of parametrically related stream data processing and analysis in the information technology logical chain allows us to obtain data stream fractal nature and numerical estimates of fractal geometry measures, as well as to obtain percolation function values stream. The second stage in the logical chain of information technologies for processing and analysis of parametric bound data stream is the construction of its phase portrait.

Let's illustrate the results obtained at this stage, using the example of data streams considered earlier.

For the data stream induced by the harmonic function 2,5sin(2πt) for *ε=0.01* and *ε=0.001*, phase portraits are shown in figures 6 and 7.

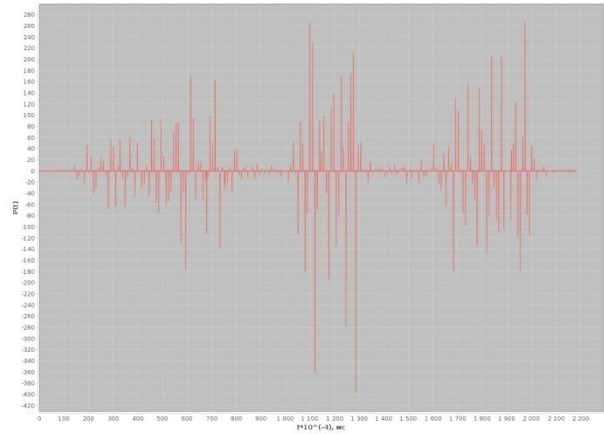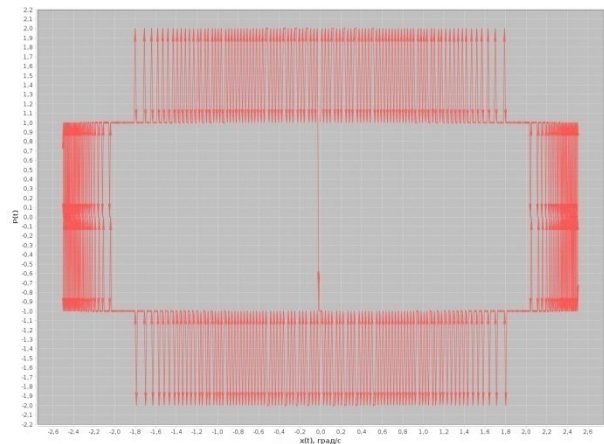**Figure 5** Percolation graph for gyroscope data, ε =0,001

Figure 6 Phase portrait for data stream induced by function 2,5sin(2πt), ε =0,01

Phase portrait shown in figure 6 illustrates following fractal nature properties of analyzed data stream. First, at a given General scale and the scale division intervals, the

distinct fractal properties of stream do not appear, but with an increase in the values of ε, it will already have these properties. Secondly, the presented data stream has the property of regularity, because the framework of the phase portrait has a closed trajectory. Figure 7 shows a phase portrait of the same data stream for a smaller value of ε. The geometric image of this portrait illustrates quite fully regular properties of data stream in question. The framework of the geometric image of the phase portrait forms a pronounced closed trajectory.
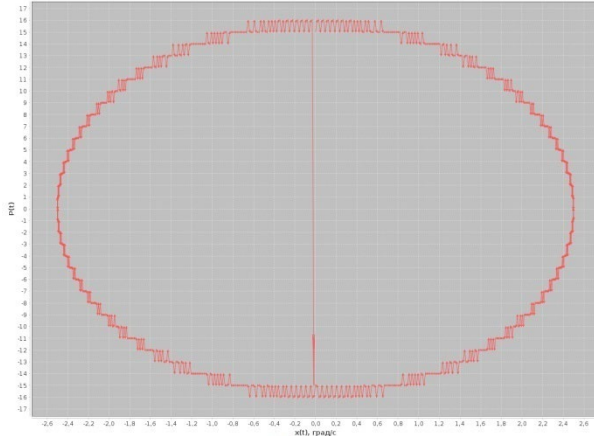

Figure 7 Phase portrait 2,5sin(2πt), *ε* =0,001

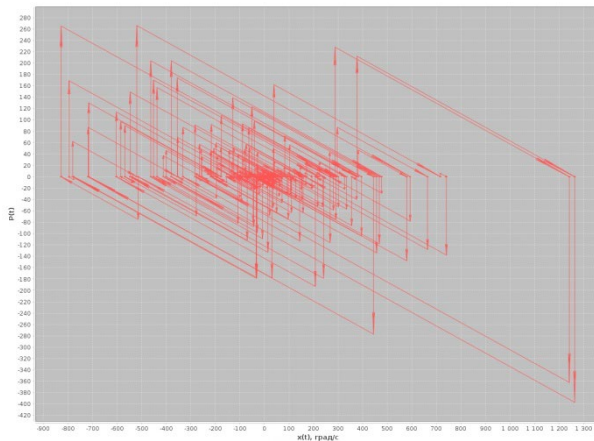Figure 8 shows a geometric illustration of the phase portrait of the gyroscopedata.

Figure 8 Gyroscope data phase portrait

The phase portrait shown in figure 8 is quite clearly and fully illustrates stream multifractal structure in phase space which presents fractal percolation and aggregation region. The space of percolation processes dominates in the region of small values of gyroscope readings (in the phase portrait – a "clot" of points in the center).

The geometry and topology of this phase portrait region reflects a regular structure on the icm–20608 quadcopter gyroscope values set. Such structure is typical for a stable and regular operating mode. Fractal

percolation covers the area of large gyroscope readings.

On the one hand, this region of the phase space reflects the singular processes in the icm–20608 gyroscope. The dynamics of a quadcopter in this case occurs along a complex trajectory, which fits into the regular mode of its operation. In Figure 8, this feature is reflected in the form of sharp changes in the values of the gyroscope readings and is indicated by arrows.

On the other hand, it allows to reflect and to describe the spatial and temporal scales of the non-standard or non-stationary gyroscope operation mode. The phase portrait shown in Fig. 8 reflects gyroscope stationary operation mode with singular phase transitions and no disturbances. Full-scale experiments on modeling the non-stationary operating mode of the gyroscope were not carried out due to objective reasons for the authors. Laboratory studies of this mode of operation and analysis of the results allow us to draw preliminary conclusions. In this case, the intermittence of the processes of fractal aggregation on fractal percolation trajectories with a positive and negative gradient will be observed.

As we can see spatial and temporal scales of the fractal percolation and aggregation processes are sufficiently fully and substantially illustrated by the phase portrait.

3 Conclusions and some generalizations

The results of processing and analysis of parametrically related data streams, indicated above, allow us to make a number of conclusions and generalizations.

First, fractal methods in large parametrically linked data streams processing and analysis is based on logical schemes of their phase portraits cognitive analysis, decoding of information hidden in them are promising and unique paradigm in the information technologies and smart information–measuring systems development.

Second, the streams of parametrically related data can be processed using various processes and methods of fractal theory and genetic data for both the collection and population of sample data and their analysis. These methods and processes reflect and define the features of the resulting estimates of fractal measures and dimensions, as well as the scope of the conclusions that can be drawn from these data. In this case, stream phase portrait are parametrically associated to the data stream. Phase portrait of the data stream determines and describes the regular and irregular properties of its structure relative to the information scale of measurements.

In a wide aspect of fundamental research in the field of intelligent information–measuring systems and intelligent information technologies, the results of this work for the first time allowed us to show how and in what the synergy of such entities as facts, laws and reality is manifested. Is it possible to draw such analogies in the framework of traditional models, algorithms, schemes, etc.? If yes, then show the results of the identified analogies and formulate trends of their theoretical development and practical continuation.

Applied aspects of the results are closely related to

the solution of a wide range of problems in the field of physical experiment, development and implementation of information technologies for control, diagnosis and control of nuclear power plants, and many others. On the one hand, the methods of fractal theory of solving complex nonlinear problems of processing, analysis and interpretation of the results of physical, biological and medical experiments are proposed. On the other hand, a new IT – technologies was developed and implemented in the trend of DAMDID processing, analysis and classification of parametrically related data.

For software implementation of IT – technologies were used data from the icm–20608 quadcopter gyroscope.

References

[1] Dvoryatkina, S.N.: Integration of fractal and neural network technologies in pedagogical monitoring andassessment of knowledge of trainees. RUDN Journal of Psychology and Pedagogics, 14 (4), pp. 451—465 (2017), DOI: 10.22363/2313-1683-2017-14-4-451-465

[2] Lymar, T. Yu., Matrova, T.S.,Staroverova, N.Yu.: Fractal search algorithm in relational databases.Bulletin of the South Ural State University Series Computational Mathematics and Software Engineering", . 3 (4), pp. 61–74 (2014), DOI: 10.14529/cmse140404

[3] Zmeskal, O., Vesely, M., Nezadal, M., B.: Fractal analysis of image structures. Harmonic and Fractal Image Analysis. 4, pp. 3–5 (2001)

[4] Mandelbrot, B: How Long Is the Coast of Britain? Statistical Self-Similarity and Fractional Dimension.Science, New Series, 156 (3775), pp. 636—638 (1967), DOI:10.1126/science.156.3775.636

[5] Feder, J.: Fractals. Plenum Press, New York, 1988

[6] Eganova, I.A.: The nature of space–time. PublishingHouse of SB RAS, "Geo" Brach, Novosibirsk, Russia (2005)

[7] Myshev, A.V.: Metrological theory of the dynamics interacting objects in the information field of a neural network and a neuron. Information Technology, 4, pp. 52–63 (2012) (in Russia)

[8] Myshev, A.V., Dunin, A.V.: Fractal Methods in Information Technologies for Processing, Analyzing and Classifying Large Flows of Astronomical Data.CEUR Workshop Proc. 1613. Selected Papers of XIX Inf. Conf. on Data Analytics and Managementin Data Intensive Domains (DAMDID/RCDL 2017).Moscow, Russia, pp. 172–176. http:// ceur–ws.org/Vol–2022/