DOI: 10.26411/83-1734-2015-4-44-5-19

Analysis of the Aviation Safety Management System by Fractal and Statistical Tools

Dmytro Bugayko, Oleg Leshchynskyi, Nataliya Sokolova
National Aviation University Kiev, Ukraine

Volodymyr Isaienko
National Aviation University Kiev, Ukraine
The Higher School Academy of Sciences of Ukraine

Zenon Zamiar
The International University of Logistics and Transport in Wroclaw, Poland

World civil aviation is an open-source system that is affected by a large number of related and non-related factors. Aviation safety is one of the prioritized directions in the industry. Its managerial decision-making process is primarily based on a versatile analysis of security data in which the choice of the appropriate mathematical apparatus is fundamental. This article suggests applying fractal-statistical analysis to evaluate the aviation safety management system in terms of determining the random distribution of quantitative dynamics of aircraft crashes with lethal consequences in the period from 1946 to 2017. This allows us to verify the adequacy of probabilistic approaches appliance in analysing the dynamics of aviation disasters. The results of research carried out on the basis of the Hurst exponent have allowed us to conclude that the dynamics of aviation disasters is characterized by the effect of "spatial memory". In other words, these are "hidden laws", for which further investigation can become an effective tool for the development of proactive methods in managing aviation safety.

Keywords: aviation safety, data analysis, fractal-statistical analysis, the Hurst exponent.

1. INTRODUCTION

Safety always remains the priority in the development of the aviation industry. From a statistical standpoint, world civil aviation is now the most secure means of transport. The global safety level of the industry lands at a figure of one disaster to ten million flights. However, according to the estimates conducted by the leading world civil aviation organizations, every 15-20 years the number of flights doubles. Such an increase in flights, thus, allows to refute the assertion that there will be no disasters and incidents in aviation.

According to the definition suggested by the International Civil Aviation Organization (ICAO), aviation safety means the state of an aviation system or organization in which risks associated with aviation activities, related to or in direct support of the operation of aircraft, are reduced and controlled to an acceptable level. [1].

Safety data analysis is central to the risk management system. Analysts usually resort to a variety of methods because of the complexity and multiplicity associated with safety factors. Among their number, probabilistic approaches are deemed to be the most useful ones.

The topicality of the article is to find the degree of uncertainty along with the existing patterns of distribution in the quantitative dynamics of aircraft crashes with lethal consequences in civil aviation. The discovery of the available "spatial memory" effects, that of the "hidden patterns" or the absence of the two will be fundamental for assessing the adequacy of the mathematical apparatus in safety data analysis. In case of their availability, the trend-stability of investigative processes may become observative, which in turn can be used to identify trend markers. It is during such periods that the world aviation safety system has the ability to implement enhanced precautionary measures. The obtained results can be applicable to the proactive methods of forecasting future safety indicators. In the proactive approach, the emphasis is placed on the prevention of aviation accidents by identifying threats and taking countermeasures to them before an actual danger takes place and thus constitutes a negative impact on the state of aviation safety.

The practical value for the industry lies in the introduction of preventive measures to aviation disasters determined as a result of proactive methods of forecasting and essentially strengthened, first of all, during the periods of revealing an increasing character of a trend change marker.

The obtained results can be applicable to the proactive methods of forecasting future safety indicators in order to implement a set of anticipatory actions against aviation disasters.

2. LITERATURE ANALYSIS AND THESIS STATEMENT

The ICAO emphasized the necessity to change the global approach to aviation safety. A new Annex 19 to the Convention on the International Civil Aviation Organization, "Safety Management", was proposed [2, 3]. A consistent approach to safety management is systematically implemented at the global, regional and state levels. It includes necessary organizational structures, spheres of responsibilities, policies and procedures. At the same time, the development of aviation safety data management remains an important tool for increasing the level of world civil aviation safety. Data management presupposes structuring, controlling and decision making on the processes and procedures sustainable for the industry organization. Also, it guarantees that data management systems achieve aviation safety objectives through the promotion of integrity, availability, usability and data protectability [1]. The result of safety analysis is to provide people in charge with the opportunity to make the most effective managerial decisions.

One of the major problems in safety analysis is the choice of the appropriate mathematical apparatus. For the world, civil aviation is an open-source system that is influenced by a large number of both related and unrelated factors. Keeping it in mind, the search for new methods of assessing acceptable levels of civil aviation safety seems relevant and important for future efficiency, the development of the industry and safety at large.

As it has been mentioned above, probabilistic methods are effective for solving a number of aviation safety tasks. For example, a sequential probability ratio test (SPRT) based on Wald's test can be highly effective in solving the problems of "aviation safety space" accurate defining [4]. Determining the integral probability indicator is decisive in assessing aviation safety risks [5]. The probabilistic evaluation is important in assessing the safety of aviation operations according to their types, for example, the evaluation of flights of unmanned aerial systems [6]. In the air traffic service system, the integrated safety management system is based on the probable forecasting of risks [7]. The likelihood of determining the frequency of aircraft crashes is fundamental in the multiplicative approach of calculating the matrix decision on the admissibility of risks [8].

At the same time, the research issue of verification remains unresolved. Specifically, it is that of the application adequacy of probabilistic approaches in the statistical data analysis of integral safety indicators (such as the number of aviation disasters) for large time series. It is a well-known fact that there is a sufficiently large class of random processes for which a tool of normal distribution is not suitable. These are the processes of market profits, the processes which describe dissipative systems, etc. To them, the law of large numbers is either not applicable or does not necessarily lead to obtaining adequate results. Among these processes, one can single out those with so-called "thick tails". To the latter, researchers have succeeded to apply fractal and statistical tools for the sake of a quasi-cyclic prior foreseen analysis of world oil prices [9] and fractal-statistical analysis of annual water fluctuations [10]. The scientific novelty of this work, however, lies in applying this very analysis in the aviation safety industry, namely for the study of statistical data of integral safety indicators (such as the number of aviation disasters) for large time series. The authors have thus chosen the theory of fractal statistics as the scientific approach to the estimation of uncertainty, along with the existing patterns of distribution in the quantitative dynamics of aircraft crashes with lethal consequences.

3. GOALS AND RESEARCH TASKS

With the help of fractal-statistical analysis, the research aims to evaluate the aviation safety management system in terms of determining the random distribution of quantitative dynamics of aircraft crashes with lethal consequences.

To achieve the goal, the following tasks have been set:

- to study the expediency of applying the theory of fractal statistics to the problems of the aviation safety management system;
- to analyze the quantitative dynamics of aircraft crashes with lethal consequences in the period from 1946 to 2017 by means of fractal statistics and evaluation of the corresponding Hurst exponent;
- to describe the time series in question and single out its main properties based on the calculated Hurst exponent.

4. THEORETICAL BASIS OF THE HURST APPROACH TO THE ANALYSIS OF AVIATION SAFETY MANAGEMENT SYSTEM

The authors have studied the annual data on the number of aircraft crashes with lethal consequences for the time period from 1946 to 2017. Systematic scientific analysis of such processes allows not only to make effective managerial decisions, but also obtain and gather experiences helpful in improving accuracy and reliability of analytical models, methods and algorithms for further decision-making. The methods of analysis that underlie preventive management can reveal the general prospects and trends of the processes in question, while ensuring the balance of short-term and long-term action programs. A very important stage in the analysis is to establish the whole set of factors capable of affecting the development of such processes. The current characteristics of these factors are variability, instability, conceptual inconsistency and incompleteness, which significantly influences the adoption of adequate managerial decisions in new areas of human activity, including the aviation safety management system. While choosing the methodology for study and research of the existing time series, the authors noticed the latter's short actual part (SAP) in the description of the production, economic and socioeconomic processes. Besides, lots of new processes in the social and industrial spheres are also represented by physically short time series. The reason for that is evident; namely, such processes never became a subject of statistical accounting. The main difficulties encountered in the analysis and the forecast of the dynamics of processes with a short actual part is a high degree of their unsteadiness depicted by the time series (TS). In other words, it was crucial to see if there is a need to develop new alternative approaches to the problem of analysing TS with SAP, given that no unified scientific view on the volume of statistics and the classification of time series has been developed yet. As a rule, $n \leq 50$ is considered a short hour series, while $n \geq 50$ - a long one [11].

According to the above mentioned classification, the investigated time-series, the authors refer neither to a short time series, nor to a time series with a short actual part. Having previously analysed various time series, they have concluded that actual time series describing the technical, social and socio-economic systems with possible external similarities may sometimes have quite different interpretations and perceptions. This is explained, in particular, by the fact that for technical systems one or a finite number of sources that generate the corresponding "signal" of the system's behaviour can be defined, which is not the case for the analysis of socio-economic phenomena to which the authors also refer the aviation safety.

Consider a partial case of a discrete random process, the time series $X(t),\ t \in Z$ *, where* $X(t)$ can be interpreted as the number of annual air crashes in at year (Fig. 1) [12].

Initially, the hypothesis about the normal distribution of the general population (time series X(t)) has been tested. For its verification, the $\chi 2$-criterion of Pearson's consent has been used. For this, the plurality of data on the number of fatal accidents has been divided into 7 intervals (Tab.1).

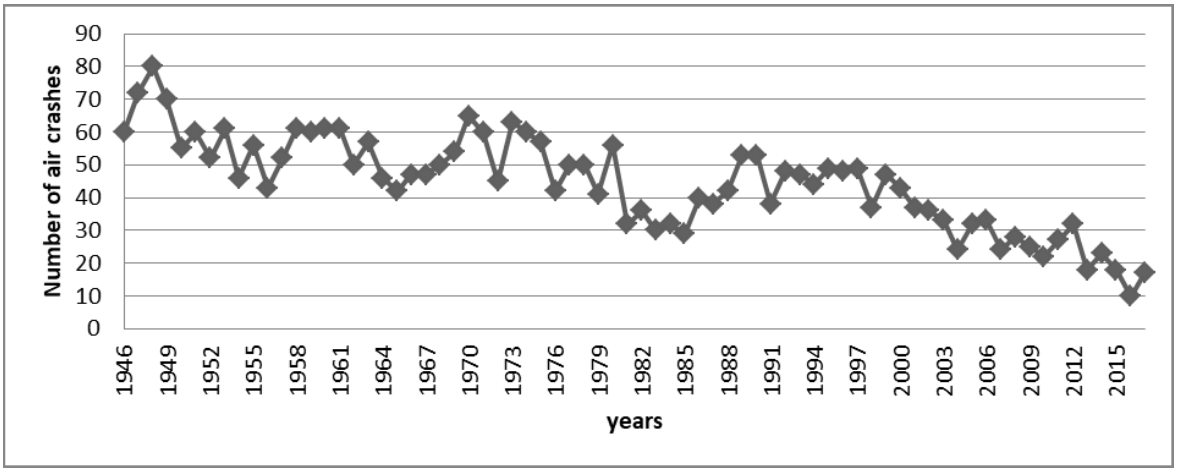

Fig. 1. The quantitative dynamics of aircraft crashes with lethal consequences in the period from 1946 to 2017.

Table 1. Calculations of Q^2 statistics

Quartiles	ni	ni/n	φ(x)	Pi	mi	(ni-mi)	(ni-mi)2	((ni-mi)2)/mi
10-20	5	0.069444	0.425	0.425	30.6	-25.6	655.36	21.41699346
21-30	9	0.125	0.456	0.031	2.232	6.768	45.80582	20.52232258
31-40	13	0.180556	0.488	0.032	2.304	10.696	114.4044	49.65469444
41-50	22	0.305556	0.52	0.032	2.304	19.696	387.9324	168.3734444
51-60	15	0.208333	0.552	0.032	2.304	12.696	161.1884	69.96025
61-70	6	0.083333	0.583	0.031	2.232	3.768	14.19782	6.361032258
71-80	2	0.027778	0.614	0.031	2.232	-0.232	0.053824	0.024114695

The H_0 hypothesis presupposed the normal distribution of the studied general variety. The following statistics have been used to test the hypothesis H_0:

$$Q^2 = \sum_{i=1}^{k} \frac{(n_i - m_i)^2}{m_i}, \tag{1}$$

where:
mi - theoretical frequencies, mi = ni * pi
ni - empirical frequencies

The random variable Q^2 has χ^2 - the distribution with the number of degrees of freedom *(k-r-1)*

where:
k - the number of intervals (in our case, it is 7)
r - the number of parameters of the theoretical distribution (in our case, *r*=2): Xsr = 43,82 the dispersion σ2 = 218.2.

By setting the level of significance α = 0.01, we got the next boundary of the critical region:
$X_\alpha = 7.779$

It is known that the more Q^2, the worse agreement of the theoretical (normal) distribution with the empirical one. With a sufficiently large Q^2, the H_0 hypothesis is rejected, so only the right-sided critical region is used. Having received Q^2 at 336.313, the authors have came to the conclusion that the found value belongs to the critical region and thus refutes the hypothesis about the normal distribution of the general variety.

Having studied the structure of this time series, an autocorrelation analysis of its levels has been carried out using the classical formula of the autocorrelation coefficient:

$$\rho_k = \frac{\sum_{t=1}^{n-k}(Y_t - \overline{Y})(Y_{t+k} - \overline{Y})}{\sum_{t=1}^{n}(Y_t - \overline{Y})^2}, \qquad (2)$$

where:
k - the number of periods according to which the autocorrelation coefficients have been calculated.

The autocorrelation coefficient has two main properties:
- firstly, it is determined by analogy with the linear correlation coefficient and, therefore, characterizes the density of an exclusively linear relationship between the current and the previous levels of a series. Briefly, the autocorrelation coefficient can detect only the presence of a linear (or close to linear) tendency;
- secondly, the autocorrelation coefficient is not sufficient to conclude whether the trend is increasing or decreasing in the levels of time series.

The correlogram of the time series in Figure 2 shows that the autocorrelation coefficient of the first order has proved to be the highest. Such a result allows to accept the hypothesis that the studied time series is tendentious and does not contain explicit cyclic oscillations.

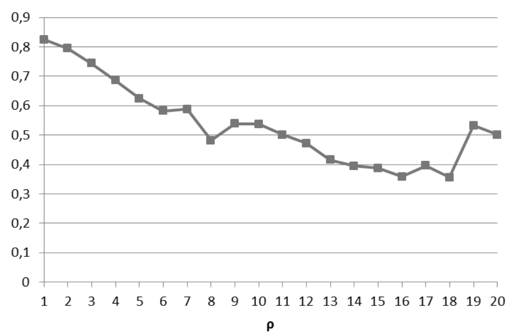

Fig. 2. Correlogram of the studied time series of crashes with lethal consequences in the period from 1946 to 2017

Table 2. Autocorrelation coefficients of the time series of crashes with lethal consequences in the period from 1946 to 2017

i	ρi
1	0.824316
2	0.794177
3	0.744269
4	0.686522
5	0.624615
6	0.582267
7	0.587575
8	0.481821
9	0.538899
10	0.537706
11	0.501316
12	0.472212
13	0.415393
14	0.394818
15	0.387381
16	0.358142
17	0.395507

18	0.355772
19	0.532619
20	0.501497

Based on the received correlogram, the authors have also adopted the hypothesis of the possible existence of a quasi cycle with a period of 19 years (19= 0.532619).

The analysis of the time series in terms of the number of aircraft crashes with lethal consequences in the period from 1946 to 2017 can be carried out in accordance with the two currently existing paradigms, namely a linear and nonlinear one. The ratio of these paradigms can be represented in the following table 3:

The authors of the article are firm in delineation of the study fields, such as fractal analysis, fractal geometry and fractal statistics; in particular, fractal analysis refers to a field of study that solves fractal functions. Specifically, there are continuous non-differentiated functions and their properties, Cantor's projectors, their constructions and properties, singular functions, etc. [17].

Considering everything above mentioned, the authors have decided to attempt at applying nonlinear approaches and in accordance with them, the elements of fractal statistics to the study of the time series of the number of air crashes with lethal consequences in the period from 1946 to 2017.

A number of Ukrainian scientists have worked on the issue of applying fractal analysis to the study of time series. In particular, Antonova and Chikina investigated time series that characterize the spread of various skin diseases in Ukraine. Based on a specialized fractal procedure for the analysis of time series, Skalozub and Klymenko studied the railway processes of Ukraine. Kramarenko, Nechai and Skalozub researched the possibilities of applying the methods of chaotic dynamics to the problems of analysing and forecasting economic and technological properties of carriages. Kyrychenko and Radyvylova investigated the long-term dependence of network traffic with the help of a R/S analysis. Kyrychenko, Radyvylova and Synelnykova studied the method of calculating the Hurst exponent for a modeled self-similar network traffic along with the method of estimating the Hurst parameter of self-similar processes. Kyrychenko and Chala constructed a comprehensive approach to the analysis of fractal properties of self-similar random processes by time series of small length. Among the foreign authors, Liubyshyna studied fractal analysis of time series and its application in geological exploration activities. Kuzenkov and Lohinov used the method of scaling in the analysis of linguistic pathologies of neurological genesis. Finally, Shelukhin, Teniakshev and Osyn investigated fractal processes in telecommunications [13-15].

Table 3. Linear and nonlinear paradigms

Linear paradigm	Nonlinear paradigm
Each influence on the initial conditions of the process causes a proportional reaction of the result obtained. The value distribution of all-time series describing the processes is subject to a normal or almost normal law.	Insignificant disturbances of the initial conditions cause bifurcations, that is, exponential super reactions of the process results. The value distribution of many time series describing the actual processes does not obey to the normal or almost normal law.

Table 4. In accordance with these paradigms, one can compare classical statistics and a separate study field of fractal statistics in the following way

The main tasks of classical and fractal statistics	
Classic statistics	Fractal statistics
Investigation of the distribution centre of the statistical aggregate (mean, mode, median, etc.), the study of variation variables (variance, deviation), density and communication directions studies (correlation ratio, correlation coefficient and autocorrelation, covariance, determination, etc.).	Trend-stability research, numerical estimates of memory depth (research on persistence to anti-persistence), fractal dimension of statistical aggregate, cyclicity characteristics, etc.).

Table 5. In its turn, the origin of linear and nonlinear paradigms can be presented as follows

Origin of linear and nonlinear paradigms	
Linear paradigm	Nonlinear paradigm
$R \approx T^{0.5}$ In the process of Brownian motion, the random particle passes the distance R which increases proportionally to the square root of the time *T* observing this particle. (Einstein, 1908)	the built-in bracket template is necessary to use *C* - a constant *n* - the number of observations $(R / S)_n$- normalized swing *H* - the Hurst exponent It is used in time series which describe the actual evolutionary processes and phenomena.

Figure 3 illustrates the analysis of the time series of data on the number of aircraft crashes with lethal consequences in the period from 1946 to 2017 fragmentally and as a whole

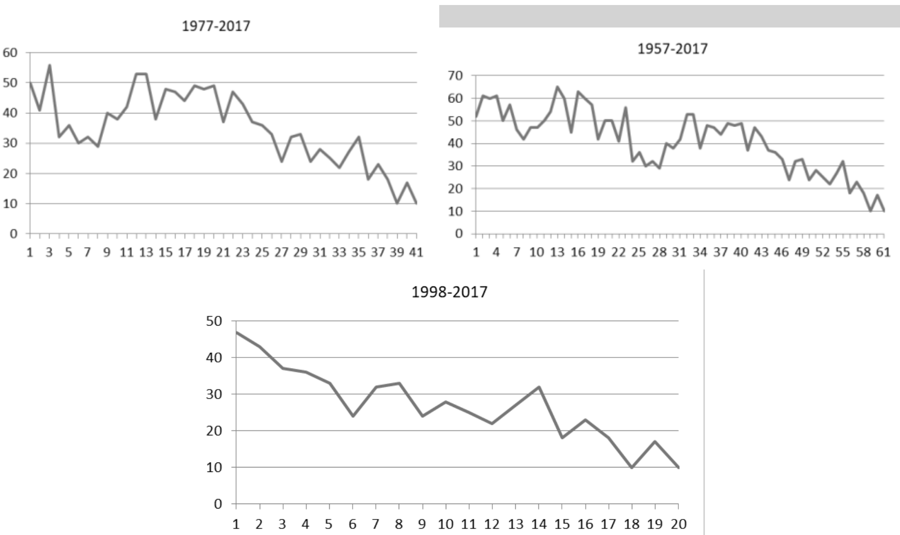

Figure 3. Fragments of the time series of crashes with lethal consequences in the period from 1946 to 2017

The authors emphasize the possibility of its self-similarity.

It is known that for finite-dimensional distributions, a real-valued process

$\{X(t), t \in R\}$

has stationary increments if

$$\{X(t+\Delta t) - X(\Delta t), t \in R\} = \{X(t) - X(0), t \in R\} \quad \text{is for all } \Delta t \in R. \tag{3}$$

Denote the sequence of increments for

$\{X(t), t \in R\}$

at a discrete time

$Y_k = X(k+1) - X(k), k \in Z.$

The process *X(t)* is broadly called stationary if the covariance function

$R(t_1, t_2) = M\left[(X(t_1) - m)(X(t_2) - m)\right]$

is invariant to the displacement, presupposing

$R(t_1, t_2) = R(t_1 + k, t_2 + k)$

for any

$t_1, t_2, k \in Z$.
Assume that the first two moments
$m_1 = M\left[X(t)\right], m_2 = \sigma^2 = M\left[X(t) - m_1{}^2\right]$
exist and are finite for arbitrary
$t \in Z$.
M can be taken as the mean; m1 as the first moment (mathematical expectation);
m2= σ^2
as the second moment (dispersion of the random process X(t)). For convenience, assume that m1=0. Then in case of stationarity
$R(t_1, t_2) = R(t_2 - t_1, 0)$,
denote covariance as *R(k)* and a correlation coefficient as
$r(k) = R(k) / R(0) = R(k) / \sigma^2$.
The real-valued random process
$\left\{X(t), t \in R\right\}$
possesses the property of self-similarity [7] with an exponent
H>0
(*H-Self-Similar* hereinafter called $H - ses$) if for arbitrary real numbers
$a \rangle 0$
finite-dimensional distributions for
$\left\{X(at), t \in R\right\}$
are identical to finite-dimensional distributions
$\left\{a^H X(t), t \in R\right\}$
simply, if for arbitrary
$k \geq 1, t_1, t_2, \ldots t_k \in R$
and
$a \rangle 0$,
the equation is the following:

$$(X(at_1), X(at_2), \ldots, X(at_k) = (a^H X(t_1), a^H X(t_2), \ldots, a^H X(t_k)) \quad (4)$$

$$\{X(at), t \in R\} = \{a^H X(t), t \in R\} \quad (5)$$

Based on the formula (5), we can conclude that changes in the time scale are equivalent to those in the spatial scale. Therefore, typical realizations of a self-similar process are visually similar regardless of the time scale at which they are studied. At the same time, it does not necessarily presuppose that the random process is exactly repeated, but rather has a similarity of statistical properties due to the fact that the statistical characteristics remain stable on a scale [16]. Parameter H is called the Hurst exponent and is used in the theory of self-similar processes owing to its perceptibility as a self-similar indicator of a random process, characterizing the property of long-term dependence.

From the theory of self-similar random processes, it is known that a non-degenerate self-similar H-ses process does not possess the stationary property. There is rather a significant connection between self-similar and stationary processes, which establishes the following theorem:

Theorem 1. [16] If $\{X(t), 0 < t < \infty\}$ equals *H-ses,* then

$$Y(t) = e^{-tH}X(e^t), -\infty \quad (6)$$

is stationary, and if $Y(t), -\infty Y(t), -\infty < t < \infty$ is a stationary process indeed, the process

$$(t) = t^H Y(lnt), 0 \quad (7)$$

is then self-similar to *H-ses*.

Theorem No. 1 nearly confirms the existence of various self-similar processes. Practically, however, the processes with stationary increments draw more attention as they lead to stationary sequences with special properties.

The processes *H-ses* with stationary increments [12] are marked as *H-sssi* (self-similar process with self-similarity parameter H and stationary increments).

Definition. The process $\{X(t), t \in R\}\{X(t), 0 < t < \infty\}$ is called *H-sssi* if it is self-similar with the parameter *H* and has stationary increments.

The following statement is well-known[6]; namely, if to suppose that$\{X(t), t \in R\}$ $\{X(t), t \in R\}$is a non-degenerate process of *H-sssi* with infinite dispersion, then $0 < H \leq 1, X(0) = 0$and covariance satisfies the equation:

$$R(t_1, t_2) = \frac{1}{2}\{[|t_1|^{2H} + |t_2|^{2H} - |t_1 - t_2|^{2H}]\sigma_X^2\} \quad (8)$$

Applying the theory of fractal statistics, the most commonly used is the range of 0.5 <H <10.5 <H <1 since the *H-sssi* process of *X(t)* with H <0H <0 is immaterial and represent a pathological case. In the case of H> 1H> 1, autocorrelation of the increment process does not exist. In the range of 0 <H <0.5, the increments process is a short range dependent process (SRD). Well-known SRD processes include that of Poisson, Markov and autoregressive processes.

In the range of 0.5 <H <10.5 <H <1, the normalized correlation function (correlation coefficient) for the increments process of X (t)

$$Y_k = X(k) - X(k-1), k \in Z \quad (9)$$

is the following:

$$r(k) = \frac{1}{2} \quad (10)$$

The theory of fractal statistics is also applicable to aggregated random processes. Let *Y = Yi, iZY = Yi, iZ* be a stationary process with the correlation function *R(k)*. The m-aggregated time series $Y^{(m)}Y^{(m)}$ is obtained by the mean of a given time series in non-overlapping time intervals (blocks of the m-length parameters). If to replace each parameters' block of the time series with their mean, that is,

$$Y_i^{(m)} = \frac{1}{m}\left(Y_{(I-m+1)} + Y_{(I-m+2)} + \cdots . + Y_I\right),$$
$$m = 1,2, ..$$

or shortly present as

$$Y_k^{(m)} = \frac{1}{m^H} \sum_{i=(k-12)m+1}^{km} Y_i, k \in Z, 0 < H < 1 \quad (11)$$

and denote the corresponding correlation function as $R^{(m)}(k)$, the study of the m-aggregated time series $Y^{(m)}Y^{(m)}$ may then be more constructive by reducing the amount of data.

Definition. The discrete random process $\{Y_k, k \in Z\}$ is exact second-order self-similar [16] in the broad sense with the self-similarity parameter H (12 <H <1) H if

$$R(k) = \frac{\sigma^2}{2}, \quad (12)$$

for arbitrary $k \geq 1$ $k \geq 1$.*X(t)* is second-order asymptotical self-similar –H-sssa in the broad sense if

$$lim_{m\to\infty} R^{(m)}(k) = \frac{\sigma^2}{2}, \quad (13)$$

where $R(k) = R^{(m)}(k)R(k) = R^{(m)}(k)$ for arbitrary $m \geq 1$. Therefore, self-similarity in the broad sense presupposes the covariance structure perseverance in aggregation of the time series.

The equation

$$R(k) = R(k) = [(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}]\frac{\sigma^2}{2}$$

means the availability of an additional structure, namely long range dependence.

The connection between the exact second-order self-similar process in the broad sense and the self-similar one in the narrow sense can be defined as follows; the process X is called strict self-similar in the narrow sense with the parameter

$R = 1 - \frac{\beta}{2}, 0 < \beta < 1, H = 1 - \frac{\beta}{2}, 0 < \beta < 1$ if $m^{1-H}X^{(m)} =, m \in N,$

where the sign «=» means the equality of the finite-dimensional distributions;

$X^{(m)} = \left(X_1^{(m)}, X_2^{(m)}, \dots\right) X^{(m)} = \left(X_1^{(m)}, X_2^{(m)}, \dots\right)$

- is the process X averaged by the blocks of the m-length, the $X^{(m)}$components of which are determined by the equation below:

$$X_k^{(m)} = \frac{1}{m}\left(X_{(km-m+1)} + \dots + X_{km}\right), m, k \in N, \quad (14)$$

The connection between the exact second-order self-similar process in the broad sense and that of strict self-similarity in the narrow sense can be perceived by analogy with the connection between stationary processes in the broad and narrow sense.

In addition to the usage of statistical similarity analysis while scaling self-similar processes, the latter can be detected by the following equivalent features.

1. In an availability case of hyperbolically decaying covariance function of the following type:

$$R(k) \approx k^{(2H-2)}L(t) \text{ при } k \to \infty, \quad (15)$$

where:

L(t) - the function of slow variation at infinity, namely for all

$$x > 0 \frac{\lim_{t\to\infty} L(tx)}{L(t)} = 1. x > 0 \frac{L(tx)}{L(t)} = 1$$

From the above mentioned, it follows that the covariance function in this very case is not summable, and the series formed by the successive values of the covariance function runs out:

$$\sum_k R(k) = \infty \quad (16)$$

The last infinite sum is another definition of long range dependence (LRD). Therefore, almost all self-similar processes are long range dependent. The consequences of such a phenomenon are significant enough, as the cumulative effect in a wide range of delays may differ significantly from that observed in the short range dependent processes such as the Poisson, Markov and autoregressive processes.

2. The sample variance of aggregated self-similar processes decays more slowly as compared to the magnitude inverse to the sample size. If to introduce a new time sequence

   $\left\{X_i^{(m)}; i = 1,2,\ldots\right\}, \left\{X_i^{(m)}; i = 1,2,\ldots\right\}$,

   being obtained by the mean of the original sequence for non-overlapping successive blocks of the m-length, then a slower decay of dispersion turns out to be characteristic for the self-similar processes by law.

3.

$$\sigma^2\left(X^{(m)}\right) \approx m^{2H-2} \text{ при } m \to \sum\infty \quad (17)$$

At that time, for common (non-self-similar) stationary random processes
$\sigma^2\left(\left\{X_i^{(m)}; i = 1,2,\ldots\right\}\right) = \sigma^2 m^{-1}, \sigma^2\left(\left\{X_i^{(m)}; i = 1,2,\ldots\right\}\right) = \sigma^2 m^{-1}$
that is, declays inversely proportional to the length of the sample. This indicates that the statistical characteristics of the sample, in particular the mean and the variance, will coincide very slowly, especially at $H \to 1H \to 1$. This property is displayed on all measures of self-similar processes.

4. If to consider self-similar processes in the frequency domain, the phenomenon of long range dependence then leads to a power-law nature of the spectral density near zero:

$$(\omega) \approx \omega^{-\gamma} L_2(\omega) S(\omega) \omega^{-\gamma} L_2(\omega), \quad \text{при } \omega \to 0\omega \to 0 \quad (18)$$

where:
$o < \gamma > 1,\ 0 < \gamma < 1; L_2$ - the function that is slowly changing at a point 0
$S(\omega) = \sum_k R(k) e^{ik\omega}$ - the spectral density

From the position of spectral analysis, the long range dependence thus implies that
$$S(0) = \sum_k R(k) = \infty$$
the spectral density tends to $\infty\infty$when the frequency $\omega\omega$approaches 0 (a similar phenomenon is called *1/f*-noise). Short range dependent processes are characterized by spectral density which has a positive and finite value for $\omega = 0\omega = 0$.

Recent ratios associated with the *H* parameter are called the Hurst exponent. The Hurst exponent of a self-similar process is in the range of 0.5 to 1. At approximation of *H* to unity, the time series becomes "more visibly self-similar", revealing itself in a slower vanishing covariance.

Long Range and Short Range Dependencies

Self-similarity as a certain property of the time series affects not only the stationaryness of the second order, but also the value and property of the *H* exponent, in particular, the boundary properties.

Among the set of LRD **processes in probability theory as well as in the simulation of time series, self-similar processes are interesting and important in connection to the subordination of their boundary theorems at a rather simple structure.**

Definition [16] $\{\boldsymbol{v_i}, \boldsymbol{i} \in \boldsymbol{Z}\}\{\boldsymbol{v_i}, \boldsymbol{i} \in \boldsymbol{Z}\}$ is called a stationary process with LRD if there is a constant $c_r > 0$ $c_r > 0$and such a real number $\alpha \in (0;1), \alpha = 2 - 2H\alpha \in (0;1), \alpha = 2 - 2H$that

$$\lim_{k\to\infty} \frac{r(k)}{c_r k^{-\alpha}} = 1 \tag{19}$$

The process $\{Y_i, i \in Z\}\{Y_i, i \in Z\}$is called stationary with SRD if there is such a constant $0 < c_0 < 1$that $\frac{lim_{k\to\infty} r(k)}{c_0^k} = 1.$

The given definition of long range dependence has an asymptotic interpretation and indicates only some limiting behaviour of the correlation coefficients and approaching of the delay to infinity. Here, only the degree of convergence is determined and not the absolute value. Defining the latter, the c_r constant, in fact, only complicates the identification of LRD [16].

The asymptotic behaviour of the coefficient *r(k)* can be studied using a Taylor series:

$$r(k) = H(2H-1)k^{2H-2} + k^{2H-2}\text{при } k \to \infty \tag{20}$$

According to the last definition, the process $\{Y_i, i \in Z\}$with $0.5 < H < 1$s thus LRD with the parameter $\alpha = 2 - 2H\alpha = 2 - 2H$. This also means that the correlations are not summed yet:

$$\sum_{k=-\infty} r(k) = \infty \tag{21}$$

In the case of the cord *r(k)* decaying hyperbolically, the corresponding process$\{Y_i, i \in Z\}$ is thus LRD.

Several partial cases for the values of *H* and its influence on *r(k)* can be drawn. If *H = 1/2, r(k) = 0*, the time series *X(t)* is a process with SRD which is explained by its complete non-correlation.

If $0 < H < \frac{1}{2}, \sum_{k=-\infty}^{\infty} r(k) = 0,$

$0 < H < \frac{1}{2}, \sum_{k=-\infty}^{\infty} r(k) = 0,$

it is theoretically possible, but now hardly encountered in real world applications. In fractal statistics, such a series is called anti-persistent. The *H=1* case leads to a degenerate situation of *r(k)=1* for arbitrary *k1*. The value *H>1* is forbidden by the stationary condition which is superimposed on the investigated process $\{Y_i, i \in Z\}\{Y_i, i \in Z\}$.

The process $\{Y_i, i \in Z\}\{Y_i, i \in Z\}$ is thus SRD if the normalized correlation function can be represented as a finite sum In the frequency domain, there is an equivalent definition of long range dependence where the necessary condition is to meet the spectral density of the process $S(\omega) =$ to the next definition.

Definition $\{Y_i, i \in Z\}$ is called a stationary process with LRD [16] if there is a real number $\beta \in (0;1)$and such a constant $c_f > 0$ that

$$\frac{\lim_{\lambda\to 0} s(\omega)}{c_f|\omega|^{-\beta}} = 1 \tag{22}$$

In such a way, the process $\{Y_i, i \in Z\}$with $0.5 < H < 10{,}5 < H < 1$ is long range dependent with the parameter $\beta = 2H - 1. \beta = 2H - 1$

Behaviour $S(\omega)$in the neighborhood of the origin of coordinates is well-described by the behavior of the function at zero.

$$S(\omega) = c_f|\omega|^{1-2H} + |\omega|^{\min(3-2H)} \tag{23}$$

$c_f = \frac{1}{2\pi} sin(\pi H) L(2H+1)\sigma^2; L(z) = \int_0^{+\infty} x^{z-1} e^{-x} dx, z > 0.$

Approximation

$S(\omega) \approx c_f|\omega|^{-\beta}, \omega \to 0, 0 < \beta = 2H - 1 < 1$ is sufficiently acceptable even for relatively large frequencies. Thus, the last equation is handy for the estimation of H in the frequency domain.

For a self-similar process, the variance of the sample mean decreases slower than the value inverse to sample's size:

$$\sigma^2 \left[X_t^{(m)}\right] \approx m^{-\beta}, 0 < \beta = 2H - 1 < 1, \tag{24}$$

where:

m – is quite big

For SRD processes, the parameter $\beta = 1\beta = 1$.

Therefore, the properties of a slowly decaying variance are usually clarified by plotting the function $\sigma^2 \left[X_y^{(m)}\right]$ from the m on to the *log-log* coordinate system (the graph of variance variation). A straight line with a negative inclination of less than 1 in a wide range of m visually shows a slowly decaying variance.

This property is also defined by the index of dispersion for counts (IDC).

$$\sigma^2 \frac{\left(X_t^{(m)}\right)}{M} \left(X_t^{(m)}\right) \tag{25}$$

The estimation of the self-similarity parameter (in our case, that of the Hurst exponent) is possible based on one of the properties characteristic for the self-similar process, in particular the analysis in the time domain which includes the R/S analysis. The latter is a classic method of finding the H parameter proposed by Hurst himself.

For a given set of observations $X = \{X_n, n \in N\}$ with the selective mean $\underline{X} = \frac{1}{n}\sum_{j=1}^{n} X_j$, the scope is defined as the difference between the maximum and the minimum deviation:

$$R(n) = \max_{1 \le j \le n} \Delta_j - \min_{1 \le j \le n} \Delta_j, \tag{26}$$

$$\text{де } \Delta_k = \sum_{i=1}^{k} X_i - k\underline{X} \ \forall k = \underline{1, n}. \tag{27}$$

The specified definition of the scope differs from that of the time sequence of the random variable X_j which equals:

$$\max_{1 \le j \le n} X_j - \min_{1 \le j \le n} X_j \tag{28}$$

In the Hurst's definition, the scope takes into account the cumulative Δ_j which characterizes the variation of the *X* dimension relative to the mean value.

$$\frac{R(n)}{S(n)} = \frac{\max\limits_{1\le j\le n} \Delta_j - \min\limits_{1\le j\le n} \Delta_j}{\frac{1}{n-1}\sum_{j=1}^{n}\left[x_j - \underline{X}\right]^2 = \frac{max(0,\Delta_1,\Delta_2,\dots\Delta_n)}{S(n)}} \tag{29}$$

It is known that for many natural phenomena such an empirical relation is applicable:

by

$$n \to \infty M\left[\frac{R(n)}{S(n)}\right] \approx cn^n \tag{30}$$

where:
c – the positive finite constant independent of *n*.

After logarithm of both parts of the last relation, the following relation can be gained:
by

$$n \to \infty ln\left\{M\left[\frac{R(n)}{S(n)}\right]\right\} \approx Hlnn + lnc \tag{31}$$

The *H* parameter can be estimated using the graph of the function ln from *lnn lnn* by means of correlation-regression analysis.

Using the tools of R/S analysis, their approximate nature should be considered, in particular, because of their capability to estimate only the level of self-similarity in the time series. The proposed in the article method is only used to verify the self-similarity of a given time series and, in case of a positive answer to a stated question, obtain the rough estimate of H.

If observations are related to the SRD process [16], then

$$n \to \infty M\left[R(n)/S(n)\right] \approx d\sqrt{n}, \tag{32}$$

where:
d – the positive finite constant independent of *n*. This case characterizes the process devoid of self-similarity

5. RESULTS DISCUSSION: FRACTAL-STATISTICAL ANALYSIS IN THE AVIATION SAFETY MANAGEMENT SYSTEM

Studying statistical data on the quantity of aircraft crashes with lethal consequences in the period from 1946 to 2017, the equation of the trend line has been obtained by common means of correlation-regression analysis and looks as *Ni=-*
*0,5789*t+65,473*. Having filtered the trend, its graphic representation obtains the following shape:

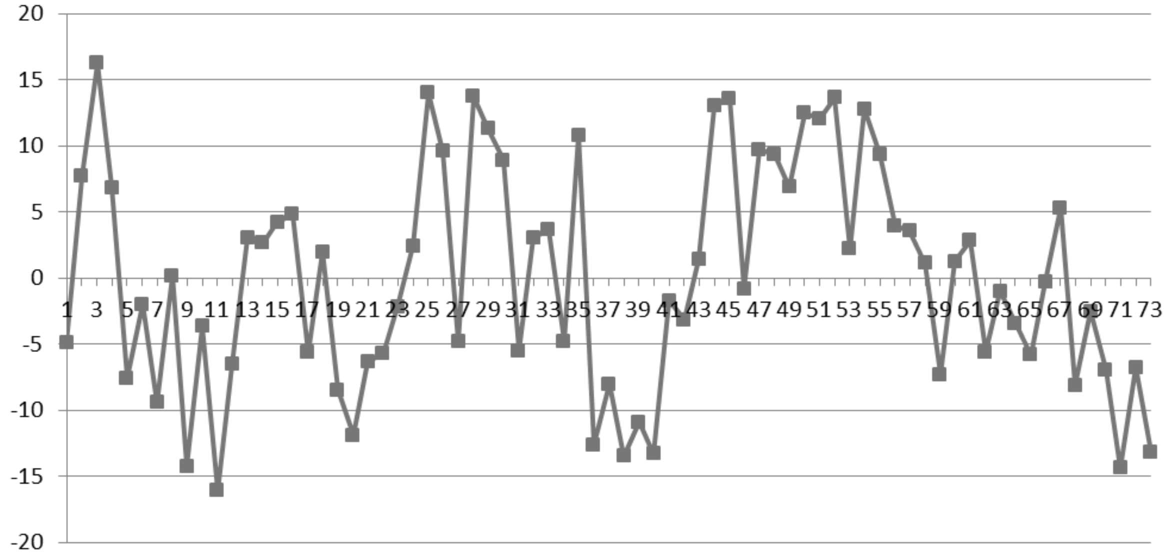

Fig. 4. Graphical representation of statistical data on the quantity of aircraft crashes with lethal consequences in the period from 1946 to 2017

Using the Hurst classical theory of R/S analysis [18] for estimating the absolute values of the studied time series, the empirical relationship between the normalized R/S magnitude and the length of the τinterval is analytically represented as follows:

$$\frac{R}{S} \approx (\tau/2)^H, H = \frac{\ln(R(\tau)/S(\tau))}{\ln\tau - \ln 2} \tag{33}$$

$$\overline{X}_\tau = \frac{1}{\tau}\int_0^\tau x(t)dt \tag{34}$$

$$V(t,\tau) = \int_0^\tau (x(t') - \overline{X}_\tau \tag{35}$$

$$R(t) = V_{\max}(t,\tau) - V_{\min}(t,\tau) \tag{36}$$

$$S(t) = \sqrt{(\frac{1}{\tau}\int_0^\tau (x(t') - \overline{X}_\tau)^2 dt')} \tag{37}$$

$$\frac{R}{S} = (\frac{\tau}{2})^H \tag{38}$$

$$\lg(\frac{R}{S}) = H \cdot \log(\frac{\tau}{2}) + b \tag{39}$$

In the process of R/S analysis, which is adapted for the application to natural evolutionary time series, we calculate.

$$X_{k,q}^{t} = \sum_{i=1}^{q}(z_j^t - z^t) \tag{40}$$

$$R = R(\tau) = \max_{1\le t\le r} X_{r,t} - \min_{1\le t\le r} X_{r,t} \tag{41}$$

$$S_k^t = \sqrt{(\frac{1}{n_k}\sum_{j=1}^{n_k}(z_j^t - z^t)^2} \tag{42}$$

$$(\frac{R}{S})_k = \frac{1}{r_k}\sum_{t=1}^{r_k}(\frac{R}{S})_k^t \tag{43}$$

The Hurst exponent is equal to the tangent of the angle of inclination of the line regression in the field where each point has its coordinates and can be found by applying the linear least squares fitting technique (LSFT). It follows that the Hurst exponent is actually the coefficient of this linear regression. In practice, a certain number of points for sufficiently small values of$\tau$is rejected. To reduce the variance of the H parameter, a certain number of points at large τ is discarded. Among the advantages of this method is the stability of the estimator relative to the distribution form, including asymmetric distributions and those with "long tails." Some of the disadvantages include the absence of a reliability analysis and error tolerance. In addition, for the Gaussian models, the R/S method loses its advantages in comparison with the maximum likelihood estimation in White's approximation or in the case of non-stationary series.

For the time-series filtered from the trend (NII(t)=N(t)-NI(t)) of the quantity of aircraft crashes with lethal consequences in the period from 1946 to 2017, such results of calculating the Hurst exponent have been obtained and summarized in Table 6.:

Table 6. Results of fractal-statistical analysis of the filtered time series of the first differences in the absolute values of aircraft crashes

No.	The initial length of the time interval (number of years)	The end length of the time interval (number of years)	The value of the step change in the length of the time interval (number of years)	The number of breakdowns	The Hurst exponent
1	8	36	1	29	0.971
2	8	36	2	15	0.972
3	9	36	1	28	0.959
4	9	36	2	14	0.969
5	10	36	1	27	0.963
6	10	36	2	14	0.947
7	11	36	1	26	0.967
8	11	36	2	13	0.983
9	12	36	1	25	0.928
10	12	36	2	13	0.948

The Hurst exponent in the ten numerical experiments does not fall below the value of 0.928. Such a result allows to preserve the hypothesis of the time series being a "black noise" (0,6≤H≤1). The specified time series are called persistent. They are also trend-resistant, account for the effect of long memory as well as possess periodic cycles and quasi cycles. Scientists believe that the enumerated properties are also characteristic for sufficiently long time series. In the formula, the use of quotation marks (“”) presupposes that the dimensions of the left and right parts do not match and for the appliance of the exact equation, a special dimensional coefficient K needs to be introduced.

The statistical value of the Hurst exponent as the coefficient of the corresponding pairwise linear regression was verified using the t-criterion, at a significance level of α=0,01. The obtained Hurst exponent has proved to be significant with such trusted intervals as (0.93-0.98).

As the regression coefficient, the Hurst exponent can be characterized as follows. The the regression equation as non-bias, substantiated, effective and invariant.

Fractal dimension is another of the characteristics of the time series and it corresponds to the empirical Hurst exponent estimated by the formula: D = 2-H.

In our case, the value of D is within (1,02-1,07). The undertaken study allows to conclude that, at first glance, the allegedly unrelated data of the number of crashes with lethal consequences for the period from 1946 to 2017 has the effect of "spatial memory", presupposing "hidden regularities".estimates obtained by means of the least squares fitting technique (LSFT) define the parameters of

The fact that the investigated time series has a complicated local structure is confirmed by studies of the time series of the first differences, the geometric image of which has the following form (Figure 6):

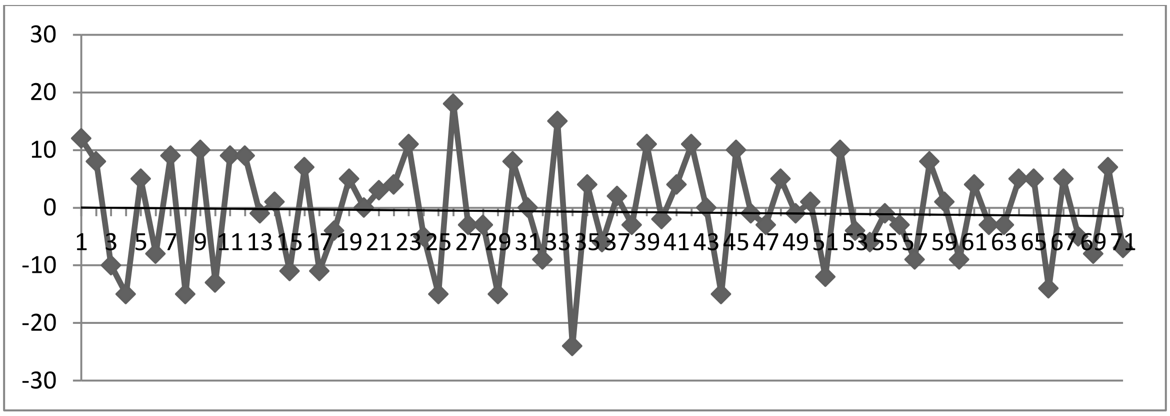

Fig. 6. Schedule of the first time-series differences in the quantity of aircraft crashes with lethal consequences in the period from 1946 to 2017

By filtering the trend of the time series of the first differences, the equation of which is ΔN = -0,0213x + 0,0624, we can obtain a new time series. The latter's results drawn on the basis of the common Hurst analysis have been summarized in Table 7.

Table 7. Results of fractal-statistical analysis of the filtered time series of the first differences in the absolute values of aircraft crashes

No.	The initial length of the time interval (number of years)	The end length of the time interval (number of years)	The value of the step change in the length of the time interval (number of years)	The number of breakdowns	The Hurst exponent
1	8	35	1	28	0.38
2	8	35	2	14	0.392
3	9	35	1	27	0.343
4	9	35	2	14	0.36
5	10	35	1	26	0.334
6	10	35	2	13	0.322
7	11	35	1	25	0.34
8	11	35	2	13	0.34
9	12	35	1	24	0.361
10	12	35	2	12	0.34

The Hurst exponent in the ten numerical experiments does not exceed the value of 0.392. Such a result indicates that this time series refers to the so-called "pink noise" $(0{,}6 \leq H \leq 1)$. The time series with the specified characteristic are called anti-persistent and are subject to a very frequent return to their mean value. The studied time series of the quantity of aircraft crashes with the lethal consequences in the period from 1946 to 2017 can be characterized by the following hypothetical property of the first increments. After a further decrease in the number of aircraft crashes, there is a certain increase with a general downward trend.

During the research, the software product *HerstCalc* (developed by Levchenko) has been used after the special preparation of the time series for its application.

This at the beginning of the research unexpected outcome inspires to further study the connections between the data, in particular, with the help of quasi cyclic and fractal-graph analysis.

6. CONCLUSIONS

1. The authors did not find any arguments that would deny the possibility of using the theory of fractal statistics, namely the Hurst analysis, for studying the dynamics of the quantity of aircraft crashes with lethal consequences in the period from 1946 to 2017. The dynamics has proven to possess a complex local structure.
2. In the process of calculating the Hurst exponent, a set of values has been considered. These are the initial and end lengths of the time interval, the value of its step change and the number of breakdowns. Together, they have helped to find the range of the Hurst exponent for the calculation method used. Ranging from 0.992 to 0.983, such a result has allowed to refer the studies time series to "black noise" $(0{,}6 \leq H \leq 1)$.
3. The minimum obtained value of the Hurst exponent confirms the referential validity of the time series of the dynamics of the quantity of aircraft crashes with lethal consequences to the class of persistent time series. Its main characteristics include trend-stability, long-term memory as well as periodic and quasi periodic cycles. The obtained results indicate that due to the impossibility of entering the studied time series to the group of common random Gaussian processes, they are to be studied by the theoretical foundations of systems with chaotic behaviour.

REFERENCES

[1] *Safety Management System*, Handbook Doc 9859, Fourth Edition, ICAO, Montreal 2018.

[2] The Convention on International Civil Aviation Doc 7300, signed in Chicago on December 7, 1944, ICAO, Montreal 2006.

[3] Annex 19 to the Convention on the International Civil Aviation Organization, Safety Management System, ICAO, Montreal 2016.

[4] Kharchenko V., Paweska M., Bugayko D., Antonova A., Grigorak M., *Theoretical Approaches for Safety Levels Measurements – Sequential Probability Ratio Test (SPRT)*, “Logistics and Transport”, 34 (2017)/2, pp. 25–31.

[5] Kharchenko V., Bugayko D., *Modern Trends of Aviation Logistics Development – Effectiveness, Safety and Security Aspects,* “Logistics and Transport”, 18 (2013)/2, pp. 17–23.

[6] Hak-Tae L., Meyn L. A., Kim S. Y., Probabilistic *Safety Assessment of Unmanned Aerial System Operations,* “Journal of Guidance, Control and Dynamics”, 36 (2013)/2, pp. 610-617.

[7] Kharchenko V., Chynchenko Y., *Integrated safety management system in air traffic services*, „Proceedings of the National Aviation University”, 58 (2014)/1, pp. 6–10.

[8] Borsuk S., Reva O., Kharchenko V., *Multiplication of Air Accidents Frequency and Hazard Desirability Coefficients for ICAO Safety Risk Tolerability Matrix Solution*, “Logistics and Transport”, 25 (2015)/1, pp. 63-70.

[9] Aref'eva O. V., Oleshko T. I., Marusich O. V., Leshchynsky O. L., Quasicyclical pre-forecast analysis of world oil prices, [in:] Scientific proceedings: a collection of scientific articles, NAU, Kiev 2011, pp. 25-31.

[10] Borisenko Y. G., Groza V. A., Leschinsky O. L., *Fractal-statistical analysis of fluctuations of Trubiz river water levels*, “Knowledge Technology”, (2014), pp. 21-25.

[11] Demidova L. A., Pyilkin A. N., Skvortsov S. V., SkvortsovaT. S., *Gibridnyie modeli prognozirovaniya korotkih vremennyih ryadov*, Goryachaya liniya, Telekom, 2015.

[12] Aviation Safety Network (ASN) - https://aviation-safety.net/statistics

[13] Kalush Yu. A., *Pokazatel Hersta i ego skryityie svoystva*, "Sib. zhurn. industr. Matem.", 5 (2002)/4, pp. 29–37.
[14] Feder E., Fraktaly, M. Mir, s. l. 1991.
[15] Kirichenko L. O., *Sravnitelnyiy analiz metodov otsenki parametra Hersta samopodobnyih protsessov,* „Sistemi obrobki Informatsyi", 48 (2005)/8, pp. 113-117.
[16] *Beran J., Statistics* for long-memory process, Chapman&Hall, New York 1994.
[17] Pratsevity M. V., A fractal approach in singular distribution studies, NPU MP Drahomanov, Kiev 1998.
[18] Hurst H. E., *Long-term storage capacity of reservoirs*, "Transactions of the American Society of Civil Engineers", 116 (1951)/5, pp. 770-808.
[19] Byistray G. P., Korshunov L. A., Nikulin N. L., Lyikov I. A., *Diagnostika i prognozirovanie sotsialno-ekonomicheskogo razvitiya regionov v ramkah nelineynoy dinamiki*, "Vestnik Tyumenskogo Gosudarstvennogo Universiteta", (2010)/4, pp. 164-170.
[20] *Nagornov O. V., Nikitaev V. G.,* Veyvlet-analiz v primerah: ucheb.posobie, M.: Izd-vo NIYaU MIFI, 2010.

Dmytro Bugayko
National Aviation University Kiev, Ukraine
bugaiko@nau.edu.ua

Volodymyr Isaienko
National Aviation University Kiev, Ukraine
volodymyr.isaienko@gmail.com

Nataliya Sokolova
National Aviation University Kiev, Ukraine
NataSokolova@bigmir.net

Oleg Leschinskij
National Aviation University Kiev, Ukraine
dimka_92@ukr.net

Zenon Zamiar
The International University of Logistics and Transport
in Wroclaw, Poland
zzamiar@msl.com.pl