How to read the trendless sequences: the "universal" set of quantitative parameters

R R Nigmatullin¹, A S Vorobev¹

¹Kazan National Research Technical University named after A.N. Tupolev–KAI, K. Marx str. 10, Kazan, Russia, 420111

e-mail: vatems14@gmail.com

Abstract. In this paper it is demonstrated a new set of “universal” parameters that help to read quantitatively any trendless sequence (TLS). This method is applied for solving the problem of selection of the “pattern” noise from the tested one and for calibration of random fluctuations expressing some qualitative inputs in terms of these quantitative parameters. This set of quantitative parameters allows to compare the TLS(s) of different nature (acoustic, mechanical, electrochemical, vibrational and etc.) with each other. Using the proposed algorithm, we analysed the differences between software simulated and experimentally generated white noises. We do suppose that the proposed “universal” scheme free from uncontrollable errors can find a wide application in solution of many practical problems.

1. Introduction
The problem of extraction of information from trendless fluctuations (always accompanied with the registered responses from open systems) became very important from the middle of the last century. Before, these fluctuations were not considered, or generally were served as a "marker" of poor-quality measurements. As an example of this approach, one can serve as a noise reduction system, given in the most known book [1]. Nowadays, the improvement of measurement and processing systems, as well as the work of many researchers [2-7], made possible to look at this problem from another angle: "Noise is a source of information." However, the analysis of the modern methods [7-16] shows that now a “universal” approach for analysis of the trendless fluctuations is absent. In many cases, the authors use the “old” methods (Fourier-transform, Wavelet-decompositions and other methods) or introduce additional processing algorithms that are based presumably on conventional methods. These methods work for solution of specific tasks and, from our point of view, they are not universal. For example, the Fourier method carries in itself unjustified supposition about “a priori” known periodicity of a random signal [7,17-19]. In addition, the F-transform creates some set of the calculated frequencies that do not belong to the system considered. Wavelet method does not contain the general criterion for selection of an optimal set of wavelets that is optimal to consideration of the chosen TLS [7, 19-21], but contains uncontrollable errors especially related to application of the specific types of wavelets to the chosen random sequence.

Many methods contain the unjustified suppositions and uncontrollable errors [2, 4], however, the analysis of fluctuations as a small part of an accompanying signal requires an accurate and specific approach, at least in its preliminary analysis. We are not able to give all literature related to this
subject (many papers are listed in book [2] and review [3] because each researcher dealing with specific noise tries to develop its own method. This preliminary analysis put forward a problem that can be formulated as follows: is it possible to suggest some “universal” set of quantitative parameters that will be useful for treatment of large massive of data? These proposed parameters should have: (a) clear interpretation, (b) free from the treatment errors and (c) can be applicable for treatment of different data, containing large number of data points. In addition, these parameters should be based on some simple principle.

In this paper, taking into account the previous attempts we want to define some simple set of quantitative parameters that can be applicable to consideration of a wide set of trendless sequences. We avoid deliberately the application of some fitting functions that are turned to be helpful for description of the SRAs [22, 23], however, the fitting error should be under a researcher control. All fitting functions with their parameters can be considered as the quantitative parameters of the second order. The basic idea (no coma!) that allows introducing this universal set is based on the consideration of trendless fluctuations as a specific “struggle” between positive and negative tendencies (amplitudes). This principle/idea helps to add some new parameters to the conventional parameters as the mean value and the standard deviation. Usually a researcher-practitioner solves the following general problem: If some input predominant parameter is changed monotonically then is it possible “to notice” and express quantitatively this monotone change from the registered output fluctuations? The proposed algorithm (confirmed on the mimic and real data) allows to find the positive answer.

2. Description the proposed algorithm

As it had been mentioned in section 1, the idea was based on reasonable supposition that a “noise” is not a “disturbing factor”; it is used as a source of additional information. It was formulated thanks to papers [2, 4, 6], where scientists tried to extract information from random fluctuations. However, from the results obtained by them, it is impossible to put forward a universal idea for working with ”noises” having different nature. We want also to find a positive answer on the following question:

Is it possible to find a set of simple “universal” parameters that can characterize the behavior of any trendless sequence irrespective to the main characteristic as the probability distribution function, which in the most cases is not known?

We must also bear in mind that the number of these parameters should be minimal and they should be rather “universal”. These parameters should not contain uncontrollable treatment errors and accurately reflect a nature of the considered fluctuations. Before, we should give some definitions for better understanding the algorithm proposed below.

Under the TLS (\(Dy = y_j - \langle y \rangle\)) we understand a set of fluctuations that oscillates relatively the horizontal OX axis. If the initial sequence has a trend then one can find the smoothed trend with the help of the POLS [24] and after its subtraction one can obtain the desired TLS again.

A single fluctuation is defined as a random deviation of an amplitude relative OX axis and, therefore, can be positive or negative. From this point of view, the given TLS can be considered as a sum of amplitudes having two opposite tendencies in positive and in negative directions, respectively. Therefore, one can put forward a simple principle reflecting a “specific struggle” between positive and negative amplitudes and based on this idea one can define the following parameter:

\[ p_1 = \text{Rg}(Dy) = \max(Dy) - \min(Dy). \]  

The value defines the range of the given sequence (\(Dy: Dy_j, j=1, 2, ..., N\)). Parameter \(p_1\) is always positive and corresponds to the maximal intensity of the given TLS.

\[ p_2 = \text{Rg}(\{|Dy|\}) = \max(|Dy_j|) - |\min(Dy_j)|. \]  

Parameter \(p_2\) defines the relative contribution of amplitudes that are located in the opposite sides of the TLS. If \(\text{Rg}(\{|Dy|\}) = 0\), it corresponds to an “ideal” balance between positive and negative amplitudes. In the opposite cases, when \(\text{Rg}(\{|Dy|\}) > 0 (<0)\) we observe a specific “spike” of positive (negative) amplitudes in the given TLS relatively each other.

\[ p_3 = DN = N(x_+) - N(x_-.) \]
Parameter $p_3$ determines the number of amplitudes located in the opposite sides of the trendless sequence. If $DN > (\leq 0)$ then the number of positive amplitudes exceeds the number of negative amplitudes and (vice versa).

$$p_s=Nv = \frac{2N}{N(\text{roots})}, 2 < Nv < \frac{N}{3}.$$  (4)

The value $N(\text{roots})$ determines the number of points (roots) that can cross the OX axis. This important parameter $p_3$ determines the number of data points corresponding to one oscillation. If $Nv$ close to 2 then the oscillations of the given sequence have clearly expressed the HF character, while $Nv$ becomes close to $N/3$ we receive finally a single LF oscillation. If $Nv > N/3$ then the desired oscillations are absent. One (two) crossing points are not sufficient for creation of a single oscillation. The parameter $N(\text{roots})$ allows to evaluate approximately the period $T \approx 2\pi/\omega T$ and frequency $\omega=2\pi/T$ of the mean oscillation.

$$p_s=\frac{\sum N_{+,-}y_{+,-}}{}.$$  (5)

This parameter determines the total/cumulative contribution of all amplitudes corresponding to positive/negative amplitudes, respectively.

For better understanding of the meaning of these parameters that follows from the “struggle” principle, we listed the combination of their signs in Table 1.

| $p_1=\text{Rg}(Dy)$ and $p_2=\text{Rg}(\mid Dy\mid)$ | $p_3=DN$ | $p_4=Nv$ | $p_5=DS$ | Comments |
|---|---|---|---|---|
| “ideal” sequence | $DN>0$ | $2 < Nv < 20$ | $DS>0$ | Excess of positive amplitudes is predominant |
| $\text{Rg}(Dy)<Dy_c$ | $DN<0$ | $20 < Nv < 50$ | $DS>0$ | Excess of positive amplitudes is remained predominant |
| $\text{Rg}(Dy) > Dy_c$ | $DN<0$ | $Nv < N/10$ | $DS<0$ | Excess of negative amplitudes is predominant |
| Critical behavior | $DN<0$ | $N/10 < Nv < N/3$ | $DS>0$ | Excess of negative amplitudes is remained predominant |
| Spike of negative (positive) amplitudes | $DN>0$ | $LF$ fluctuations | $DS>0$ | |

Besides these parameters, one can add some parameters that follow from the fitting of the bell-like curve (BLC) by beta-distribution function [23]:

$$Y \equiv Bd(x; A, B, \alpha, \beta) = A(x - x_0)^\alpha (x_N - x)^\beta + B.$$  (6)

This function fits BLC $Y(x)$ that, in turn, is obtained after the integration of the SRA (the sequence of the ranged amplitudes) with preliminary elimination of its mean value. We want to stress here that the BLC $(>0)$ describes a value of maximal fluctuation located in the given interval $[x_0, x_N]$. The meaning of these fitting parameters $(A, B, \alpha, \beta)$ and their calculations are explained in papers [22,23]. In addition to these four parameters, one can add the maximal value of the BLC: $p_6=Y_{mx}$ (this characteristic point separates the positive set of amplitudes from the negative one), then the similar measures of asymmetry:

$$p_8=DX = \frac{1}{2}(x_0 + x_N) - x_{mx}, p_9 = \frac{\alpha}{\beta} \equiv r = \frac{x_{mx}-x_0}{x_N-x_{mx}}.$$  (7)

in vertical direction and a small parameter $B$ that indicates a possible asymmetry in the horizontal direction. Therefore, summarizing the parameters determined above one can propose at least 10 quantitative parameters (including also the mean value of $y$, $p_0=$mean($y$)) pretending on the “universal” description of the given TLS. We want to stress here that parameter $p_6$ determines the specific “equilibrium” line separating the set of positive amplitudes from the negative ones. Parameters $A, p_6=Y_{mx}$ describe the intensity of fluctuations. $Y_{mx}$ of the BLC separates the positive set of amplitudes from the negative ones. $B$ is asymmetry in horizontal direction with respect to OX axis. $p_7=x_{mx}$ and $p_8=DX$,
where $Dx = \frac{x_n + x_N}{2} - x_{mx}$ have two cases. If $Dx$ close to zero – BLC is symmetrical. If $Dx > 0$ ($< 0$) excess of negative amplitudes, (excess of positive amplitudes). $p_\varphi = r$, where $\frac{a}{b} \equiv r = \frac{x_{mx} - x_0}{x_N - x_{mx}}$ also have two cases. If $r \approx 1$ – BLC is symmetrical. If $r < 1$ ($> 1$) excess of negative amplitudes (excess of positive amplitudes). $S(a, \beta) = \alpha + \beta$, $0 < S(a, \beta) < 1$ is the most of amplitudes is located near the mean value, where $l < S(a, \beta) < 2$ the desired TLS exhibits the fractal properties. In this research, we will use only the parameters $p_\varphi p_\varphi$. As it follows from examples given below these parameters are sufficient for the solution of the problems formulated above.

As it has been mentioned in Introduction, we deliberately avoid the application of the FFT (the fast Fourier transform). This transformation is based on a priori (and, in the most cases, invalid) supposition that the given TLS is pure periodical and, in addition, it contains the excess of frequencies that do not exist in the given TLS reflecting the output of the studied system. That is why in the most cases the FT of the given TLS is not used as a fitting function [17,18] and is considered as an “independent” source of information. The same situation is related to application of the wavelets [19-21, 25, 26]. At present time, we do not have the justified criterion for selection of an optimal wavelet that fully corresponds to the analysis of the given TLS. Besides, the selected wavelet family contains some uncontrollable errors and application of two and more types of wavelets taken from other temperature families can lead to different/contradictory results.

Finishing this section, we should show a possible way of application of these 10 parameters in analysis of a large amount of data. As it was mentioned above, the most part of researches solve the following problem: there is one predominant and external factor $F$ (for example, pressure $P$, temperature $T$, humidity $H$, intensity of radiation $I$, substance concentration $C$ and etc.) that is changed monotonically in the certain range $F_k = a k + b$ ($k = 0,1,\ldots, K$). For each fixed value of $F_k$ we have some rectangle matrix of measurements $N \times M$ ($N \geq M$), where ($j = 1,2,\ldots,N$) determines the number of data points and $M$ ($m = 1,2,\ldots,M$) determines the total number of repeatable (or statistically similar) measurements. We imply that this initial matrix contains only fluctuations (serving as an additional source of information) expressed in the form of the TLS(s). Having these huge massive of data is it possible to reduce them to the minimal number of quantitative parameters reflecting some common features of the phenomenon studied? The solution can be divided on some steps:

S1. We transform each TLS to the reduced sequence containing only $n$ (in our case $n = 10$) quantitative parameters. Therefore, we obtain the reduced matrix $n \times M$. We transpose this matrix ($n < M$) and obtain the matrix $M \times n = (n \times M)^T$, where number of measurements (rows) exceeds the number of the reduced parameters $n$ (columns). Each column ($p_{m,l}$: $m = 1,2,\ldots,M$; $l = 1,2,\ldots,n$) for the fixed $l$ has a different statistical meaning and in some cases it has a sense to make them statistically similar to each other with the help of transformation:

$$
N_{rm_l} = -1 < \frac{p_{l} - \text{mean}(p_l)}{\max(p_l) - \text{mean}(p_l)} < 1, l = 1,2,\ldots,n.
$$

(8)

Here $p_l$ ($l = 1,2,\ldots,n$) determines the vector of one of the reduced parameters belonging to the reduced set $n$.

S2. Then one can apply the reduction procedure to the $N_{rm_l}$ vertical vector having $m = 1,2,\ldots,M$ data “points”. After that, we obtain the square matrix $n \times n$ containing only parameters taken over all measurements.

S3. Finally, for the remaining reduced matrix $Pr_{n,n}$ one can apply the singular valued decomposition (SVD) operation [27] (widely used in the PCA [28]) and find the eigenvalues of this matrix. These eigenvalues (associated with the principal components) located in the descending order ($Ev_1 > Ev_2 > \ldots > Ev_n$) can characterize the initial rectangle matrix $N \times M$ forming $n$-possible functions $Ev_k(k)$ depending on the external factor $F_k$. From these functions one can select the most sensitive and monotone function that reflects the monotone behaviour of the predominant factor $F_k$. We should stress that this approach is rather “universal” and can be applied to a wide set of TLS(s).

3. Quantitative “reading” of real sequences

For analysis, we chose a "white" noise. For this purpose, a fluctuation sequence called white noise was generated (in the Matlab application software package), and a white noise source (R&S FS-SNS 26
Smart Noise Source) was used to generate real data. The power source needed for the generation was supplied to the noise source input, its output was connected to an oscilloscope (Agilent oscilloscope DSOX3014A). The received signals were processed and evaluated in the computer algebra system Matcad.

Each measurement for both cases was reproduced 100 times. The total number of the data points in each recording was \( N = 1000 \). For treatment of the data the following algorithm was proposed, it includes in itself the following steps:

**S1.** For definition we have the rectangle matrix of the size \( N \times M \) \((N=1000, M=100)\). With the help of the proposed algorithm we reduce the initial matrix to the reduced matrix \( n \times 3 \) \((n=10)\) and it includes the set of parameters \((p_0-p_9)\) defined in section 2). From each measurement we took only three basic parameters \((s=1 \text{ (mean}(m)), s=2 \text{ (stdev}(m))) \text{ and } s=3 \text{ (range}(m)=\text{max}(m)-\text{min}(m)), \) where \(m=1,2,\ldots,M \) \((M=100)\).

**S2.** In order to see the possible differences between the real and model noises one can consider the ratio:

\[
R_{ts}(p) = \left( \frac{M_{d}(s)(p)}{M_{n}(s)(p)} \right) - 1, s = 1,2,3; p = 0,1, \ldots, 9.
\]

In the result of realization of these two steps, we obtain three functions depending on the 10 parameters \((p_0-p_9)\). The subsequent analysis of these functions is simple: if function (9) exceeds the value of one, then this relationship will have a distinctive feature that shows the difference between the data being analysed. Figure 5 shows the fixed sampling of the initial set of data associated with the modelled white noise created by embedded program. Other data look similar and therefore are omitted.

![Figure 1.](image)

**Figure 1.** (a). Typical noise obtained by software embedded in Matlab. It corresponds to the first sampling. Solid red line shows the desired SRA that divides the positive set of amplitudes from the negative ones. If we divide this line relatively mean value and then integrate it, we obtain the curve showing the summarized contributions of these amplitudes. These curves are shown in the next figure; (b). These two curves determine the total contribution of the positive (blue line) and negative (red line) amplitudes. They are slightly differed from each other. On the right-hand side we show the total set of the reduced parameters corresponding to the chosen sampling. The meaning of these parameters is explained in Tables 1 and in the text.

Figures 2(a,b) shows the difference between the two types of white noise (a) in the form of 10 parameters (b). We constructed them in accordance with expression (8) in order to see the desired differences.

As one can see from analysis of this figure, the proposed parameters \((p_0-p_9)\) are sufficient to detect the TLS difference.

Based on this method, we also analysed the acoustic noise recorded from the frictionless bearings (FB) in a normal state and noise from the FBs with artificially created defects [34]. The proposed algorithm allows detecting the desired defects that initially had a qualitative description only.
In this work, some universal “platform” for processing noises data was demonstrated. In our opinion, it can attract the attention of many researchers working in various branches of engineering and the natural sciences. Let us summarize and focus on their distinctive features:

1. The proposed “platform” is based on a simple “struggle” principle between positive and negative amplitudes. It is free from the uncontrollable errors and does not use any unjustified supposition as the FFT or wavelet analysis;

2. The proposed platform reminds a “violin cleats” which allow a researcher to "tune" and customize flexibly this tool. If necessary, the sensitivity can be increased by adding of number of the previously omitted parameters. This peculiarity makes possible in increasing the detection of the predominant external factor;

3. The instructive example based on real data shows that the proposed platform enables to detect the difference between modelled and experimental data associated with "white" noise.

Before, any researcher knew presumably only two basic “universal parameters” as the mean value (it can be interpreted as a specific equilibrium line separating the positive amplitudes from the negative ones) and the standard deviation (it evaluates approximately the deviations from this line). The proposed scheme can add another set of universal parameters based on specific tendencies of a “struggle” between positive and negative amplitudes.

5. References
[1] Kharkevich A A 1965 Struggle with Disturbances Radio and Connection
[2] Timashev S F 2007 Flicker – noise spectroscopy Phys. Math. Lit.
[3] Timashev S F and Polyakov Yu S 2007 Review of Flicker-noise spectroscopy in electrochemistry Fluctuation and Noise Letters 7(2) R15-R47
[4] Timashev S F and Polyakov Yu S 2008 Analysis of discrete signals with stochastic components with flicker noise spectroscopy International Journal of Bifurcation and Chaos 18(9) 2793-2797
[5] Yulmetyev R M 2000 Stochastic dynamics of time correlation in complex systems with discrete time Phys. Rev E 62 6178-6194
[6] Yulmetyev R M 2002 Quantification of heart rate variability by discrete nonstationarity non-Markov stochastic processes Phys. Rev. E 65 046107
[7] Wenlong L, Junwei X, Heming W and Chuan S 2016 Non-stationary component extraction in noisy multicomponent signal using polynomial chirping Fourier transform Springer Plus 51177
[8] Xiaoling C and Liangjun Y 2018 Power line interference noise elimination method based on independent component analysis in wavelet domain for magnetotelluric signal Geosystem Engineering 21(5) 251-261
[9] Zeman P M, Till B C, Livingston N J, Tanaka J W and Driessen P F 2007 Independent component analysis and clustering improve signal-to-noise ratio for statistical analysis of event-related potentials Clin Neurophysiol 118(12) 2591-2604
[10] Rabiner L R and Gold B 1975 Theory and application of digital signal processing (Englewood Cliffs NJ Prentice-Hall Inc.)
[11] Mendel J M 1995 Lessons in estimation theory for signal processing, communications, and control Pearson Education
[12] Hagan M T, Demuth H B and Beale M H 1996 Neural network design (Boston: Pws Pub.)
[13] Ifeachor E C and Jervis B W 2002 Digital signal processing: a practical approach Pearson Education
[14] Bendat J S and Piersol A G 2011 Random data: analysis and measurement procedures (John Wiley & Sons)
[15] Gelman A, Carlin J B, Stern H S, Dunson D B, Vehtari A and Rubin D B 2013 Bayesian data analysis (CRC Press)
[16] Box G E P, Jenkins G M and Reinsel G C 2013 Time series analysis: forecasting and control (John Wiley & Sons)
[17] Alt R 1978 Error propagation in Fourier transforms Mathematics and Computers in Simulation, 20(1) 37-43
[18] Becker R I and Morrison N 1996 The errors in FFT estimation of the Fourier transform IEEE Transactions on Signal Processing 44(8) 2073-2077
[19] Gao R and Yan R 2011 From Fourier Transform to Wavelet Transform: A Historical Perspective Springer Wavelets 17-32
[20] Jameson L and Waseda T 2000 Error Estimation Using Wavelet Analysis for Data Assimilation: EEWADA American Meteorological Society Journal of Atmospheric and Oceanic Technology 17 1235-1246
[21] Sheefa R G D and Sheela M I 2015 A Study on Asphyxiating the Drawbacks of Wavelet Transform by Using Curvelet Transform International Journal of Computer Science & Mobile Computing 4(9) 318-323
[22] Nigmatullin R R 2009 Quantitative Universal Label: How to use it for marking of any randomness? Allerton Press Inc Physics of Wave Phenomenon 17(2) 100-131
[23] Nigmatullin R R, Giniatullin R A and Skorinkin A I 2014 Membrane current series monitoring: essential reduction of data points to finite number of stable parameters Computational Neuroscience 8 120
[24] Ciurea M L, Lazanu S, Stavaracher I, Lepadatu A-M, Iancu V, Mitroi M R, Nigmatullin R R and Baleanu C M 2011 Stress-induced traps in multilayered structures J. of Applied Phys. 109 013717
[25] Liu L C, Chien J C, Chuang H Y and Li C C 2001 A Wavelet-Based Preprocessing for Moving Object Segmentation in Video Sequences Wavelet Analysis and Its Applications Lecture Notes in Computer Science 2251
[26] Graps A 1995 An Introduction to Wavelets IEEE Comp. Sci. Engr. 2 50-61
[27] Moonen M and De Moor B 1995 SVD and Signal Processing III Elsevier Science B.V. p v
[28] Jolliffe I T 2002 Principal Component Analysis Springer Series in Statistics p 487
[29] Nigmatullin R R and Gubaidullin I A 2017 NAFASS: Fluctuation spectroscopy and the Prony spectrum for description of multi-frequency signals in complex systems Communications in Nonlinear Science and Numerical Simulations 56 1263-1280
[30] Nigmatullin R R, Maione G, Lino P, Saponaro F and Zhang W 2017 The general theory of the quasireproducible experiments: How to describe the measured data of complex systems? Communications in Nonlinear Science and Numerical Simulation 42 324-341

[31] Nigmatullin R R 2016 Detection of Quasi-Periodic Processes in Experimental Measurements: Reduction to an “Ideal Experiment” Complex Motions and Chaos in Nonlinear Systems, Nonlinear Systems and Complexity 15 1 1-37

[32] Nigmatullin R R, Toboev V A, Lino P and Maione G 2015 Reduced fractal model for quantitative analysis of averaged micromotions in mesoscale: Characterization of blow-like signals Chaos, Solitons & Fractals 76 166-181

[33] Nigmatullin R R, Ceglie C, Maione G and Striccoli D 2014 Reduced fractional modeling of 3D video streams: the FERMA approach Nonlinear Dynamics 80(4) 1661-1664

[34] Nigmatullin R R and Vorobev A S 2019 The “Universal” Set of Quantitative Parameters for Reading of the Trendless Sequences Fluctuation and Noise Letters 18(4) 19