Russian Journal of Earth Sciences

On the Use of a Complex Indicator of the Stability of Permutation Entropy of Time Series Fragments When Analyzing Infrasound Monitoring Signals of the Altai Republic

N. G. Kudryavtsev*[,1], I. N. Frolov[1], and V. Yu. Safonova[1]

[1] Gorno-Altaisk State University, Gorno-Altaisk, Russia
* **Correspondence to:** Nikolay Kudryavtsev, ngkudr@mail.ru.

Abstract: This paper discusses one of the approaches that allows us to assess the degree of complexity or randomness of fragments of a time series in order to detect infrasound or geomagnetic signals in the results of observations of the dynamics of the natural or man-made processes under study. In our case, we are talking about monitoring the infrasound background on the territory of the Altai Republic. To solve the problem of estimating the required characteristics of a time series with minimal computational costs and in real time, a complex indicator of the stability of permutation entropy is introduced, since estimating the value of classical permutation entropy for $n = 3$ (the most commonly used version of permutation entropy) does not allow solving the problem with sufficient accuracy.

Keywords: infrasound monitoring, time series, permutation entropy, complexity assessment, stratospheric waveguide, turning points.

Citation: Kudryavtsev, N. G., I. N. Frolov, and V. Yu. Safonova (2023), On the Use of a Complex Indicator of the Stability of Permutation Entropy of Time Series Fragments When Analyzing Infrasound Monitoring Signals of the Altai Republic, *Russian Journal of Earth Sciences*, 23, ES6010, EDN: XXZEHM, https://doi.org/10.2205/2023es000887

1. Introduction

When studying the life activity of complex systems, measurement experiments are usually conducted, the result of which are time series of manifestations of certain parameters of the observed phenomena recorded at certain intervals. The results of such observations require significant effort for their interpretation, and the task of classifying fragments of the obtained data is one of the intermediate steps that allows one to come to an understanding of the processes being studied. Often, a "classifier" (classification algorithm) must determine, at a minimum, the presence or absence of a signal of a certain type, or simply a signal that differs from noise, in the fragment of a time series under study. Sometimes the task is complicated by the fact that the classification must be performed in real time.

So, for example, in the signal shown in Figure 1, there is no visually detectable signal other than noise. Although, sometimes a lot depends on the degree of "approximation" or detail of the signal being studied.

At the same time, in the fragment of the time series in Figure 2, a periodic signal is shown, although somewhat noisy, having almost the same amplitude and a slight shift to the positive area.

And the following figure (Figure 3) shows signals with different frequencies, amplitudes and varying degrees of noise.

There may also be time series where, among the predominant noise component, short-term rather implicitly determined oscillatory values of the amplitude of the fixed value are observed, corresponding, perhaps, to some fragments of the natural phenomena being examined (Figure 4). And these fragments must be found and isolated among the noise.

Research Article

Received: 6 December 2023
Accepted: 29 December 2023
Published: 31 December 2023

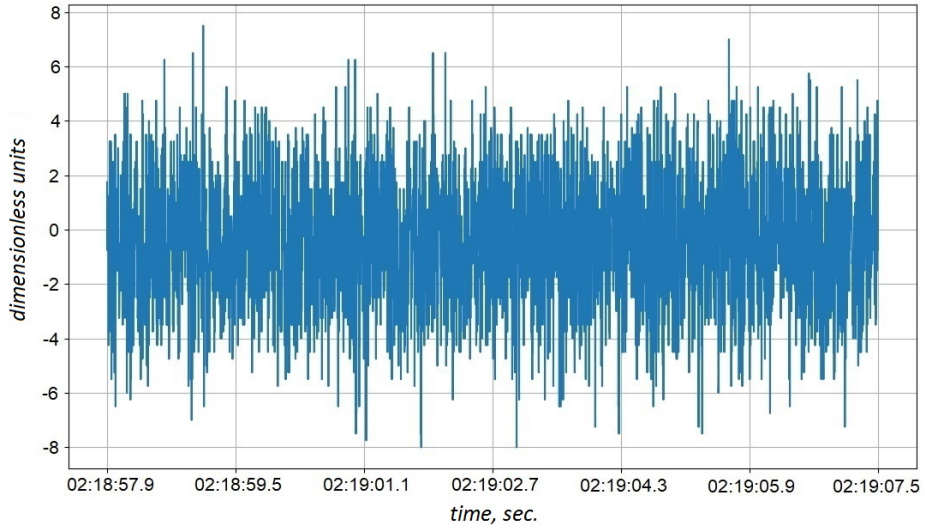

Figure 1. Noise acoustic signal. Gorno-Altaisk (03.09.2023. Ten second interval local time 02:18:57–02:19:07).

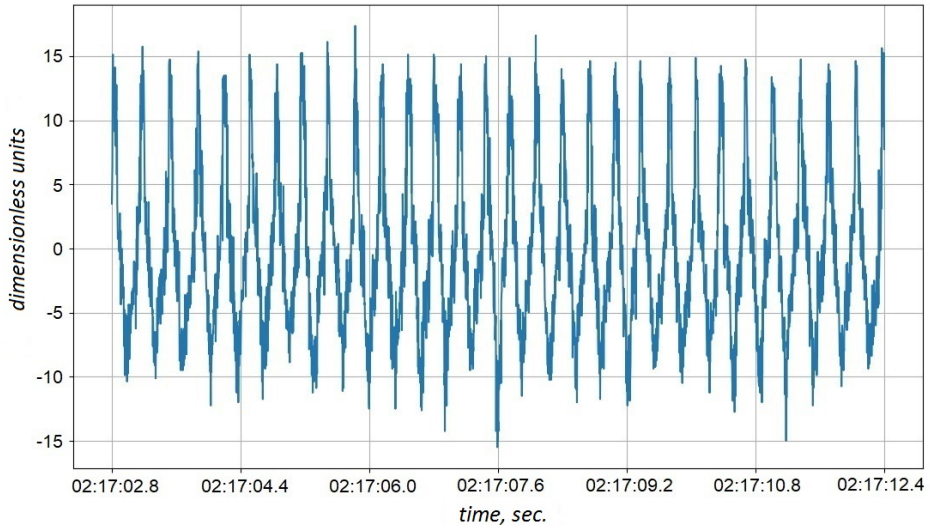

Figure 2. Periodic acoustic signal with a frequency of approximately 3 Hz. Gorno-Altaisk (03.09.2023. Ten second interval local time 02:17:02–02:17:10).

Time series that are obtained as a result of such observations can be divided into regular, stochastic and mixed; they can also be either stationary or non-stationary. For example, the study of mechanical motions that require measurements of the distance between the Moon and the Earth generates a regular time series. In this case, the results of measuring current velocities in a turbulent flow are stochastic quantities. In time series of different nature generated by non-stationary processes, there are both elements of regular data and stochastic components.

Continuing a brief excursion into the technology for processing observational results, it should be noted that when implementing long-term measurements, parameterization is used to simplify analysis procedures and reduce the amount of processed information. Parameterization is often understood as the extraction from observational data of a minimum set of the most essential parameters characterizing the system or process being studied [*Chumak*, 2012]. In fact, parameterization is a reduction of data, which, under certain conditions, allows you to restore the original data with a given accuracy based on the resulting set of parameter values and predetermined rules. The most commonly used

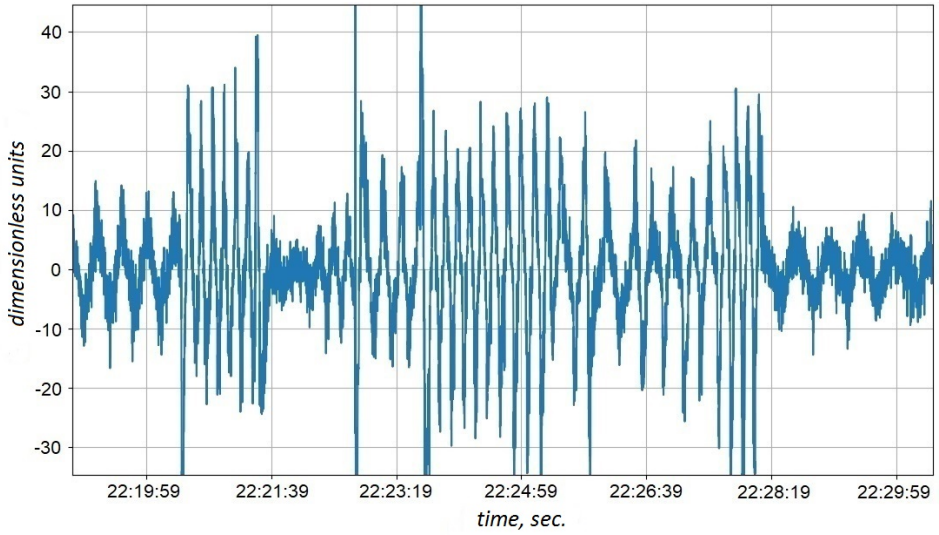

Figure 3. Mixed acoustic signal. Gorno-Altaisk (06.10.2023. Ten second interval local time 22:19:59–22:29:59).

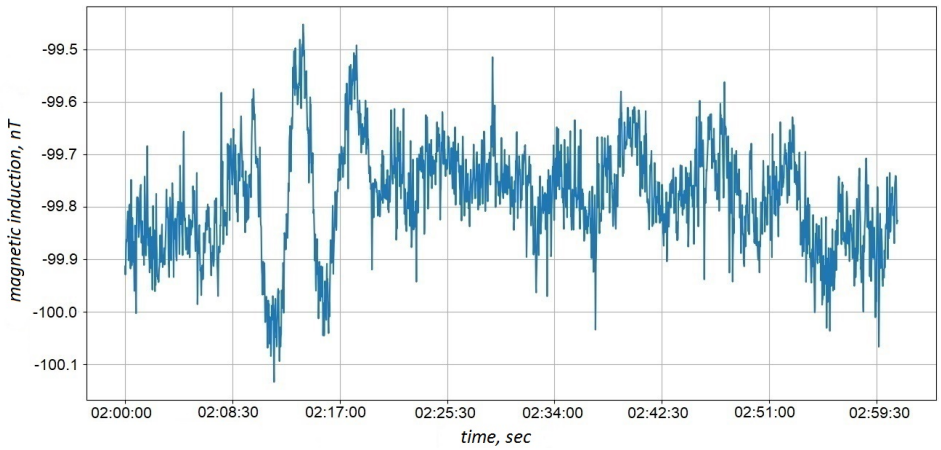

Figure 4. Difference signal between the vector sum of projections in the Euclidean metric and the magnitude of the Earth's magnetic field induction vector. Intermagnet network, SPG magnetic observatory (11.02.2024 hour interval from 2 am to 3 am GMT).

parameterization methods are integral data transformations using complete systems of orthonormal trigonometric functions (Fourier transforms) or specially selected localized functions (wavelets). In general, time series parameters are the results of applying procedures for selecting such a basic set of indicators that make it possible to quantitatively characterize the dynamics of changes in the system and understand the processes occurring in it.

Until about the middle of the last century, a naive idea of time series was used. It was believed that any series contains only a general trend, regular and random components, the numerical characteristics of which do not change over time. In this case, the trend could be considered as a fragment of change with a very long period. Time series analysis methods were reduced to extracting regular components from a data set in such a way that "white noise" remained in residuo.

The purpose of this work is to describe one of the approaches that allows us, using parameterization of measurement results, to detect a useful infrasound signal in conditions of significant interference, mainly in real time.

Generally speaking, different versions of entropies are used to act as a parameter that gives an assessment of the complexity of a certain fragment or the entire series of

observations as a whole. Permutation entropies of different orders are a useful analysis tool for a large number of time series of very different nature. They make it possible to obtain important information about dynamic processes in complex systems producing non-stationary series with a significant stochastic component.

An analysis of works describing the use of permutation entropy when processing data obtained during observations of the dynamics of various natural phenomena showed that this time series parameter has been used for quite a long time and very effectively. Thus, in [*Zunino et al.*, 2012], permutation entropy was successfully used in the study of the North Atlantic Oscillation, which has a significant impact on winter weather over Western and Central Europe. In [*Liang et al.*, 2020], using permutation entropy, an indicator of the complexity of the time series of concentration of suspended solid microparticles in moisture droplets in the air was determined, on the basis of which a model was obtained and an algorithm was formulated for predicting the trend of changes in the concentration under study over time. In [*Zhu et al.*, 2016], permutation entropy was used to study the influence of climate on the dynamics of the spread of dengue fever in Southeast Asia in the period from 2004 to 2015. In [*Fu et al.*, 2019; *Lu et al.*, 2022; *Roushangar et al.*, 2021; *Silva et al.*, 2021; *Zhang et al.*, 2019] permutation entropy has been used to model and study various aspects of climate, from studying the dynamics of drought periods in northwestern Iran and northeastern Brazil, to seasonal variations in wind speed in the Ningxia and Jilin regions, and temperature throughout mainland China.

Materials and methods

Let us consider a sample $\{x_k\}$, obtained from a time series $\{x\}$. The permutation entropy of a sample of a time series of order n is usually called the Shannon entropy n! permutations defined by the relations of n consecutive sample values. Further in this work we will talk about entropy for $n = 3$. All six possible permutations for three consecutive values of the series are shown in Figure 5, published in [*Traversaro et al.*, 2018].

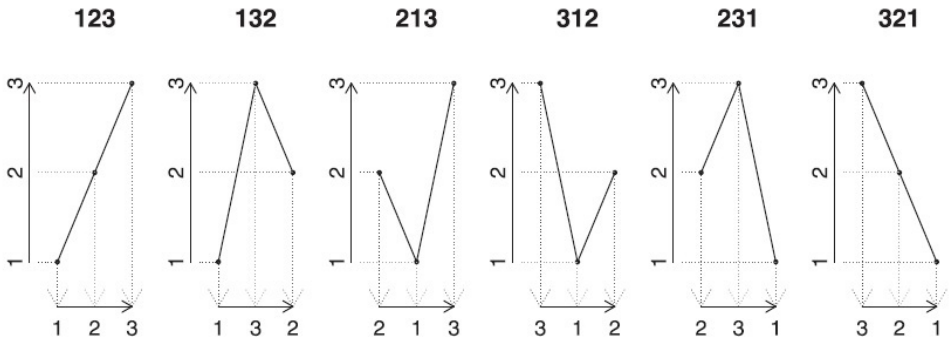

Figure 5. Permutations for $n = 3$.

The entropy itself is calculated using the classical formula (1).

$$\sum_{j=1}^{3!} m_j = N,\ p_j = m_j/N \quad and \quad H_3 = -\sum_{j=1}^{3!} p_j \log_2 p_j \tag{1}$$

Let us note that in formula (1) the letter m_j denotes the frequency of occurrence of each of the six permutations in the studied fragment of the time series. It should also be said that entropy is expressed in bits. The maximum possible value of permutation entropy will be observed when each of the permutations is uniformly distributed. Thus $\max H_3 = \log_2 3!$, which gives the result 2.584962500721156. Entropy takes its minimum value when there is only one permutation, for example (1, 2, 3). In this case, the entropy will be equal to 0.

In addition to considering the theoretical foundations of permutation entropy, we would like to dwell on the criterion proposed by *Kandal* [1981] for assessing the randomness of the sample data under study.

The method (or criterion) of turning points is one of the simplest criteria for the randomness of a fragment of an observed time series. Its essence is to count the "peaks" and "troughs" that can be formed by three neighboring points. In this case, a peak is a point that is larger than two neighboring ones, and a trough is a point that is smaller than two neighboring ones. Since out of three points, as mentioned above in the section on permutation entropy, only six permutations 123, 132, 213, 231, 312, 321 can be written, then four of them correspond to the criterion of turning points. Two permutations (132 and 231) are peaks, two more (213 and 312) are troughs. Therefore, given a random uniform distribution of time series values, the probability of encountering one of the four turning points is 4/6 or 2/3. If we do not take into account the two extreme points of the series fragment, then the mathematical expectation of the presence of a turning point at the current location is equal to $2/3(n-2)$. The value of the standard deviation for turning points is somewhat more difficult to derive. Ultimately, a sign of randomness of a time series was obtained and justified if the number of turning points in the studied fragment of the time series falls within the interval from $2\cdot(N-2)/3-((16\cdot N-29)/90)^{0.5}\cdot 1.96$ to $2\cdot(N-2)/3+((16\cdot N-29)/90)^{0.5}\cdot 1.96$.

The authors of this work decided to use permutation entropy to analyze the dynamics of time series representing the results of monitoring the infrasound background at an experimental site in the area of the main building of Gorno-Altaisk State University (Gorno-Altaisk, Altai Republic). To monitor the infrasound background, at the first stage of research, an acoustic sensor was used, developed on the basis of an INMP 441 MEMS microphone, the description of which is given in [*Microsin.net*, 2020]. As an undoubted advantage of the presented design, we can note the presence of a built-in 24-bit analog-to-digital converter and a special I2S bus, supported by many modern controllers for high-speed transmission of digitized audio data, synchronization of several microphones, etc. Despite the presence of a built-in bandpass filter, evaluation calibration tests have shown that with sound pressure fluctuations of the order of 0.5 Pa, an acceptable signal/noise ratio is observed starting at frequencies of the order of 2.5–3 Hz. When choosing this microphone as an acoustic sensor capturing part of the infrasound range, we proceeded from satisfactory price-quality ratios, the availability of this equipment on the market and the sufficiency of the observed part of the infrasound range to solve the assigned tasks [*Schwardt et al.*, 2022].

To increase sensitivity at low frequencies and provide protection from rain, snow, and gusts of wind, the microphone was placed in a plastic tube with a diameter of 110 mm, hermetically sealed at both ends, playing the role of a Helmholtz resonator (neck dimensions 5 mm × 20 mm). The dimensions of the pipe (620 mm) were selected taking into account a short distance from the control circuit, attachment to the mast and natural protection from the wind.

The resonant frequency for a Helmholtz resonator can be calculated using the formula (2).

$$f_H = \frac{v}{2\pi}\sqrt{\frac{S}{V_0 L}}, \tag{2}$$

where f_H – frequency, Hz; v – speed sound of a medium; S – neck section, m^2; L – neck length, m; V_0 – resonator volume, m^3. Estimated calculations showed a resonant frequency of the order of 7 Hz.

The digitization (sampling) frequency when working with a datalogger is set by the microcontroller and depends on the data transfer rate on the I2S bus. In our case, the frequency is set to about 330 Hz. At this frequency, a relatively small amount of daily data

is accumulated, and at the same time it remains possible to work quite well with signals up to 100–150 Hz.

Figure 1–3 show examples of signals obtained through this acoustic sensor. The authors of this work were tasked with developing a "classifier" algorithm that allows identifying intervals of a time series containing a "noise" signal (containing only noise), a mixed signal (a periodic highly noisy signal) and a "good signal" (a periodic signal with minimal amount of noise). In this case, the duration of the interval was assumed to be equal to 3000 samples of the time series (approximately 10 seconds). In addition, the classification algorithm in the future was planned to be executed on a microcontroller in real time, which imposed requirements on a certain ease of implementation on the developed "classifier" algorithm.

The best approach, according to our preliminary estimates, for dividing fragments of a time series into "signal" or "noise" was to parameterize the analyzed data using permutation entropy (the degree of chaos of the fragment under study). To calculate this parameter, it was only necessary to "count" the number of three consecutive sample values located in a certain way in the time series stream. Even the hardware implementation of such "counters" would not be very difficult. It was assumed that a well-structured sinusoidal-like signal should show low entropy, and a noise signal should show correspondingly high entropy.

However, the calculation of entropy, performed at the preliminary stage of developing the algorithm in Python using the function `dis,cis = op.complexity\_entropy(x1)` from the `ordpy` library (corresponds to the formula (1) described above), for the signal presented in Figure 2, showed the value `dis` = 0.9992368090042947, i.e. high degree of chaos or uncertainty. At the same time, for the signal shown in Figure 1, the entropy turned out to be `dis` = 0.9863362446476029. Thus, the degree of uncertainty of the "noise" signal was estimated to be much lower than the same value for a signal resembling a sinusoidal one. A closer examination of the source data showed that the analyzed sinusoid was modulated in amplitude by a "weak" noise signal. Hence the high degree of estimated uncertainty.

To improve the classification procedure, it was decided to use the turning point criterion, which was described in the "materials and methods" section of this work.

If you look closely at the criterion of turning points, you will notice a certain "kinship" of this criterion with permutation entropy, since the turning points, the number of which is estimated when checking the required criterion, are nothing more than four of the six permutations for $n = 3$ used for calculations of permutation entropy.

Returning to the computational experiment we carried out, we note that after performing the appropriate computational procedure, the criterion of turning points was also not "able" to distinguish a noisy periodic signal from a "noisy" signal.

Of course, it was possible to filter the signal from noise before calculating the permutation entropy or turning point criterion. Even in real time this is not difficult to do. However, after "coarse" filtering using the moving average method, the turning point criterion began to "mark" both analyzed signals as periodic, i.e. the "noise" has "turned" into a "reasonably smooth" signal.

To find a solution to the problem in the current situation, the following hypothesis was put forward. Let us have a sample corresponding to a non-random noisy periodic signal, for example the one presented in Figure 2. If for this time series we calculate, for example, the normalized permutation entropy or the normalized complexity indicator or the normalized indicator of the number of turning points, first for the entire sample, then for the elements of this sample numbered 0, 2, 4, 6, 8, 10..., then for sample elements numbered 0, 3, 6, 9, 12... etc. to sample elements numbered 0, k, $2k$, $3k$..., where k is an empirically selected number depending on the sample size, then we get a sequence of values of the parameter we have defined, consisting of k elements. We call such a sequence the k-complex of the parameter under study. The assumption we accepted without proof was that for a periodic, albeit noisy time series, the obtained points of the k-complex themselves must represent

some kind of "conditionally non-random series". If the original time series is random or very noisy, then the points of the *k*-complex must also be a "conditionally random series". Thus, we can say that by introducing its *k*-complex instead of a time series, we reduce the number of elements of the series without changing its degree of randomness. In this case, of course, we need to decide how we will evaluate the degree of randomness of the *k*-complex itself. As an assessment of the degree of randomness or regularity of a *k*-complex, it was proposed to use the same criterion of Kandal turning points. In addition, within the framework of the hypothesis we accepted, it was agreed that if we obtain a k1-complex of the same parameter, but for a sample not from the original time series, but from the *k*-complex, then we will call such a sequence a second-order k1-complex. Thus, the increase in the orders of *k*-complexes can be continued to a "reasonable" limit.

Note that a similar technique with "sparse" samples from the original time series was used, for example, in Higuchi's article when calculating a parameter that was called "fractal dimension" [*Higuchi*, 1988].

Results

For example, let us consider a fragment of the time series presented in Figure 2, *k*-complex for $k = 64$, and as a parameter the "normalized complexity" indicator (cis-parameter) of the time series. Essentially, the cis parameter is the one's complement of the normalized permutation entropy or dis parameter from the *ordpy* library. The desired 64-cis complex is shown in Figure 6

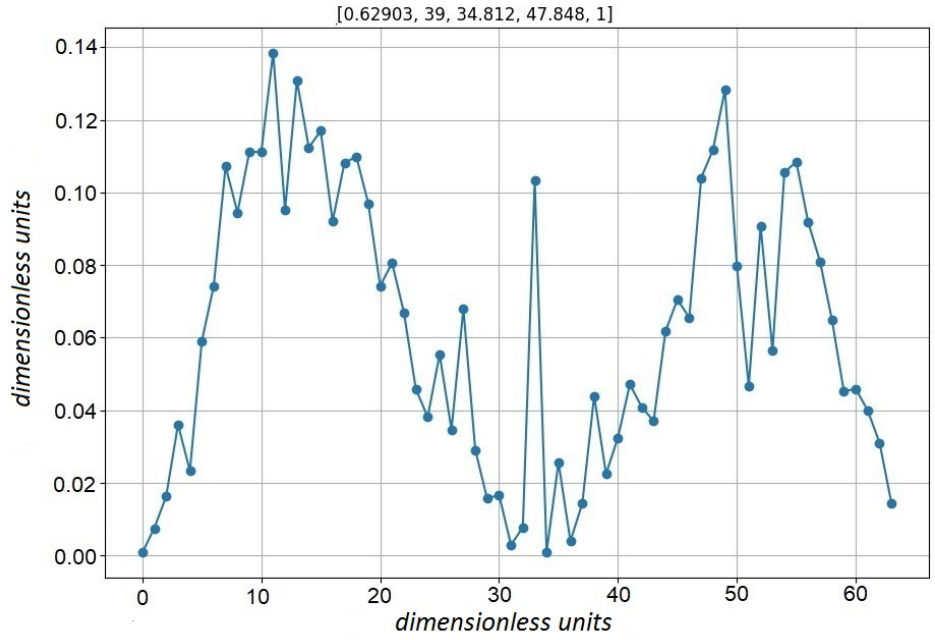

Figure 6. Sequence of values of the 64-cis complex.

The heading of Figure 6 (and other images of *k*-complexes) dispalys a vector in square brackets containing as the first element the normalized number of turning points of the *k*-complex. The second element of the vector is the absolute actual value of the number of turning points, the third and fourth elements of the vector are the left and right boundaries of the range, if the actual number of turning points fell within the range, the series would be assessed as random. A single value of the last element of the vector shows that in our case a series of 64 points (64-cis-complex) is classified as random according to the criterion of turning points, although subjectively, Figure 6 shows a fairly regular sequence.

A more reliable result, presented in Figure 7, was obtained for the 64-pnp complex, where the normalized number of turning points was used as a parameter. According to the criterion of turning points, this *k*-complex was assessed as regular, i.e. non-random (as

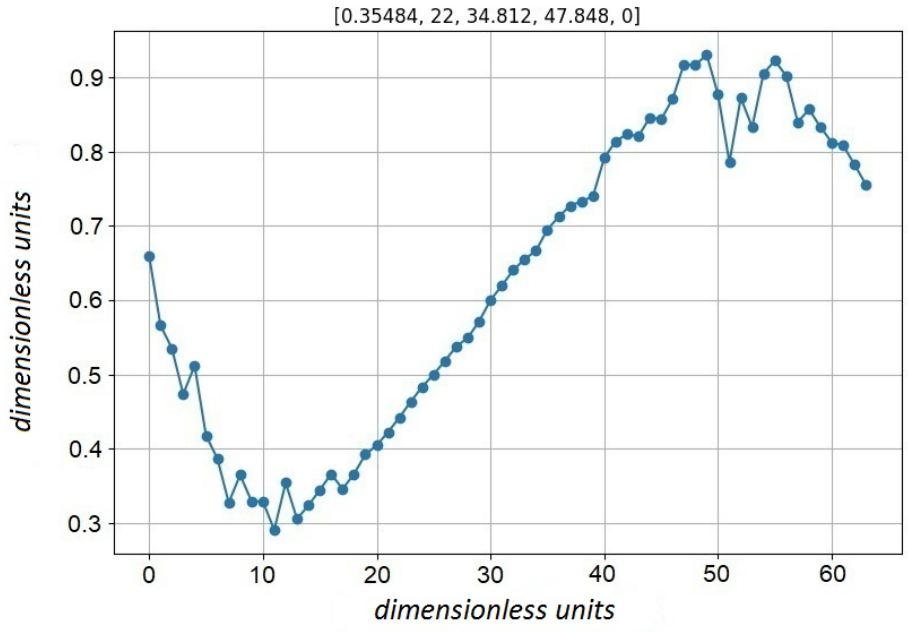

Figure 7. Sequence of values of the 64-pnp complex for the periodic signal shown in Figure 2.

expected in accordance with our accepted hypothesis). For the noise signal in Figure 1, the 64-pnp complex is shown in Figure 8. This sequence, as expected, was determined to be random by the turning point criterion.

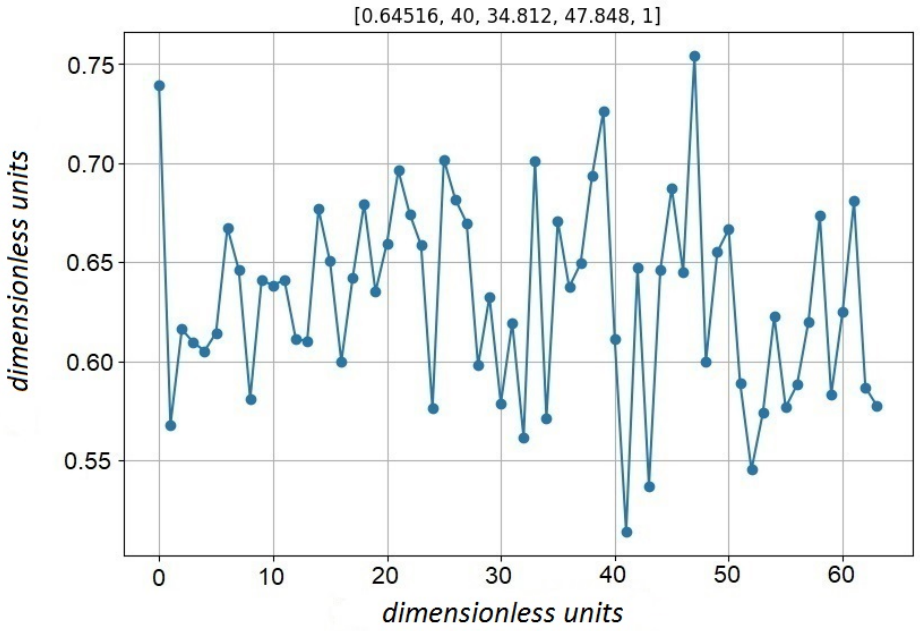

Figure 8. 64-pnp complex for the "noise" signal shown in Figure 1.

Thus, the performance of the discriminating procedure based on the *k*-complex was empirically tested. We could have stopped there if there had not been a need for a more complex classification of the signals under study. In fragments of the time series, it was necessary to detect a "good signal" and classify it as the "gs" class (an example of a signal is shown in Figure 2), the "noise signal" – the "nn" class (Figure 1) and the "bad signal" – the "bs" class. (Figure 9)

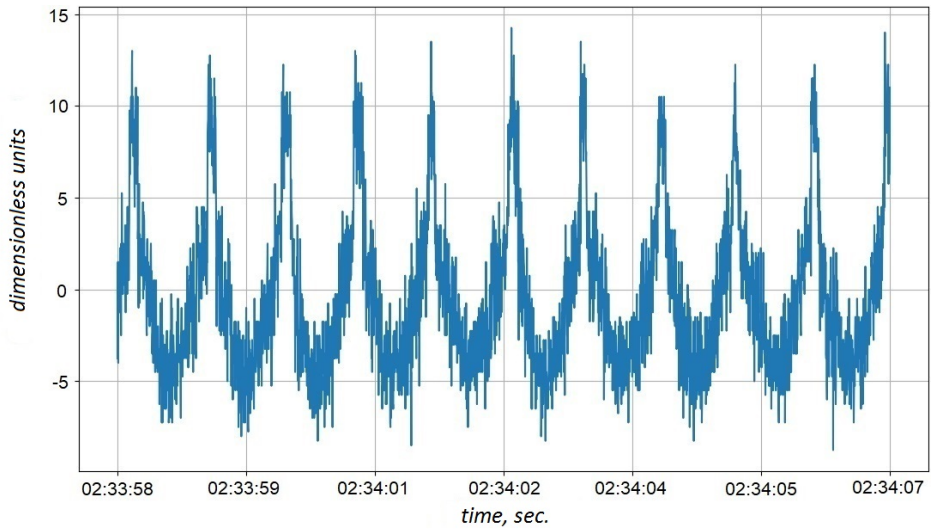

Figure 9. Example of a "bad" signal.

The 64-pnp complex for the signal presented in Figure 9 (shown in Figure 10) was assessed by the turning point criterion as random, although visually the graph (Figure 10) shows a certain regularity of the trend.

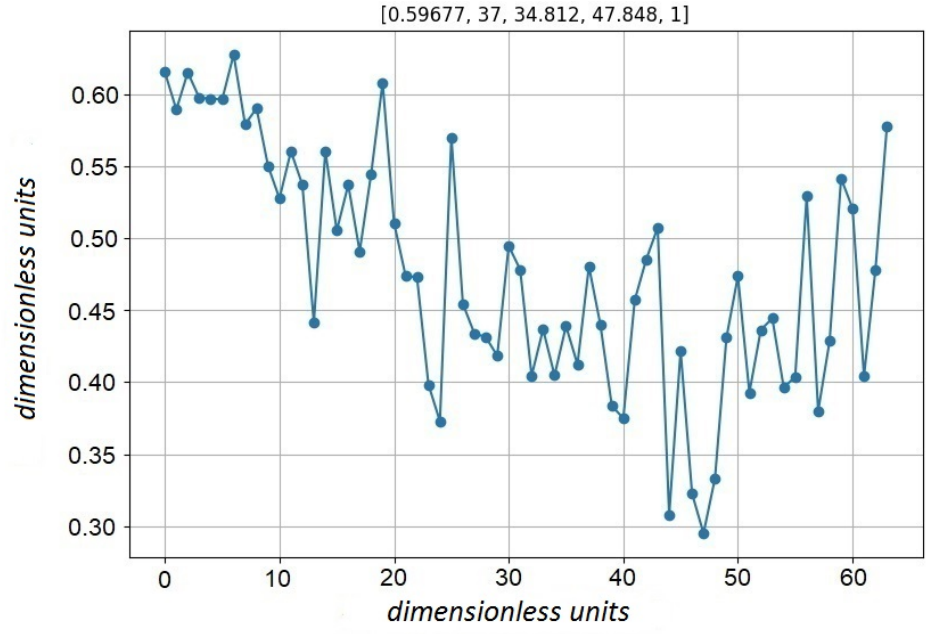

Figure 10. 64-pnp complex for the signal shown in Figure 9.

To evaluate complex signals, such as those presented in Figure 9 or 12, we decided to use a second-order k-complex. At the same time, for greater reliability, it was decided to use a 128-pnp complex as a first-order complex, and a 32-pnp complex for a second-order complex. In total we have a 128-32-pnp complex. By the way, already during the transition of the first-order pnp complex from $k = 64$ to $k = 128$ for the signal shown in Figure 9, the 128-pnp complex was assessed as regular (Figure 11).

It can be reasonably assumed that adjusting the orders of pnp complexes and selecting numbers k for each of them will allow us, if necessary, to implement quite complex classifications of signals..

Now let us give an example of more "complex" bad signals. One such signal is shown in Figure 12.

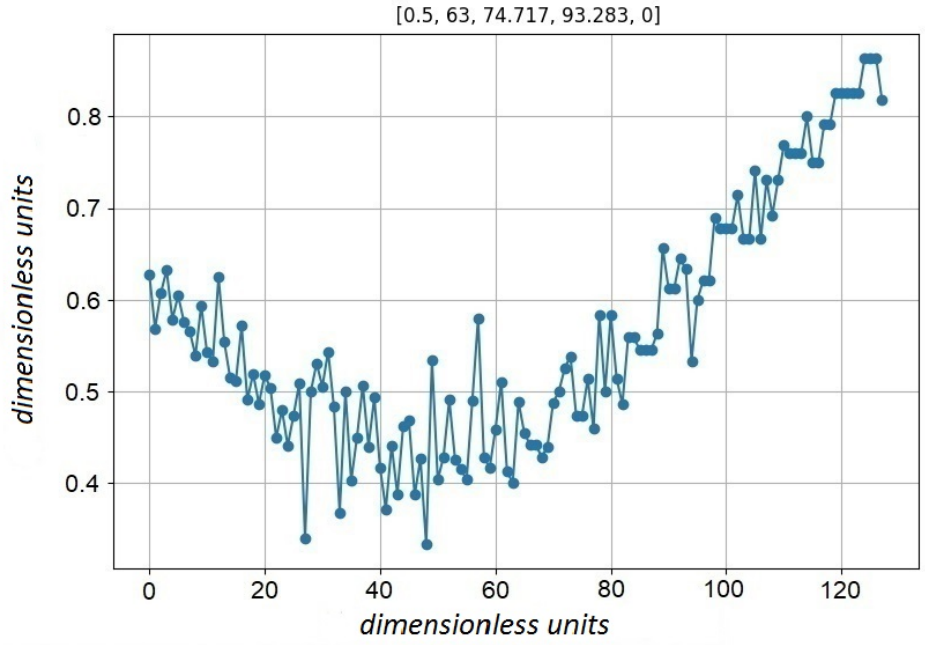

Figure 11. 128-pnp-complex for the signal (Figure 9).

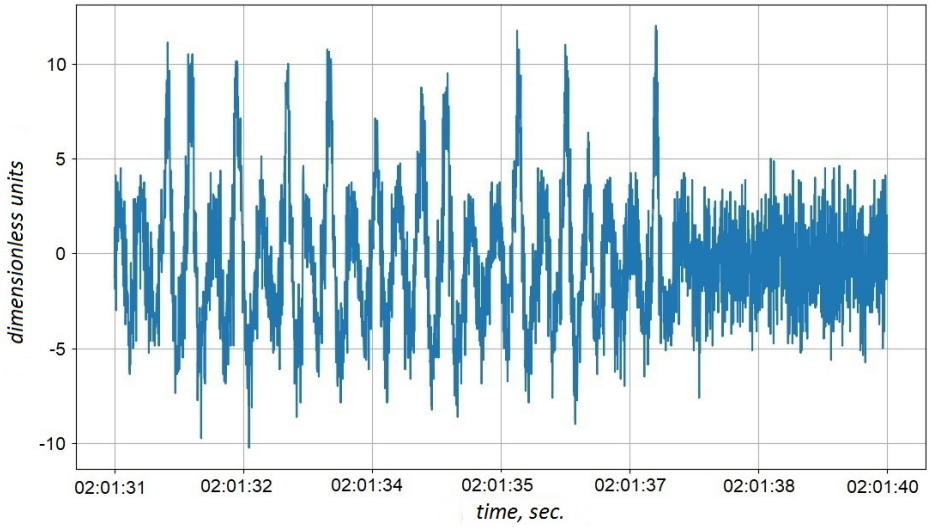

Figure 12. Example of a "bad" signal from the "bs" class.

Figure 13 shows a 128-bpn complex for this signal, which was estimated to be random.

Figure 14 shows a 128-32-pnp second-order complex. This complex was assessed as regular. Thus the signal was recognized as belonging to the "bs" (bad signal) class.

We called the above-described indicator (even to some extent a classifier) of the degree of regularity/chaoticity of the studied fragments of time series based on k-complexes of different orders a complex indicator of the stability of permutation entropy.

Now let us consider how, as a first approximation, we can organize a classifier of the studied fragments of time series, using in our case a 128-pnp complex of the first order and a 128-32-pnp complex of the second order. As already mentioned above and noted in the headings of the figures in which the above complexes are presented, each complex has a one-bit regularity indicator. If the complex is chaotic (irregular), then the bit value is equal to 1. If the complex is regular, then the bit is equal to 0. So, we have two complexes, and therefore two bits. The input for the classifier will be a two-bit binary vector, where the first bit (the left bit of the input vector) corresponds to the

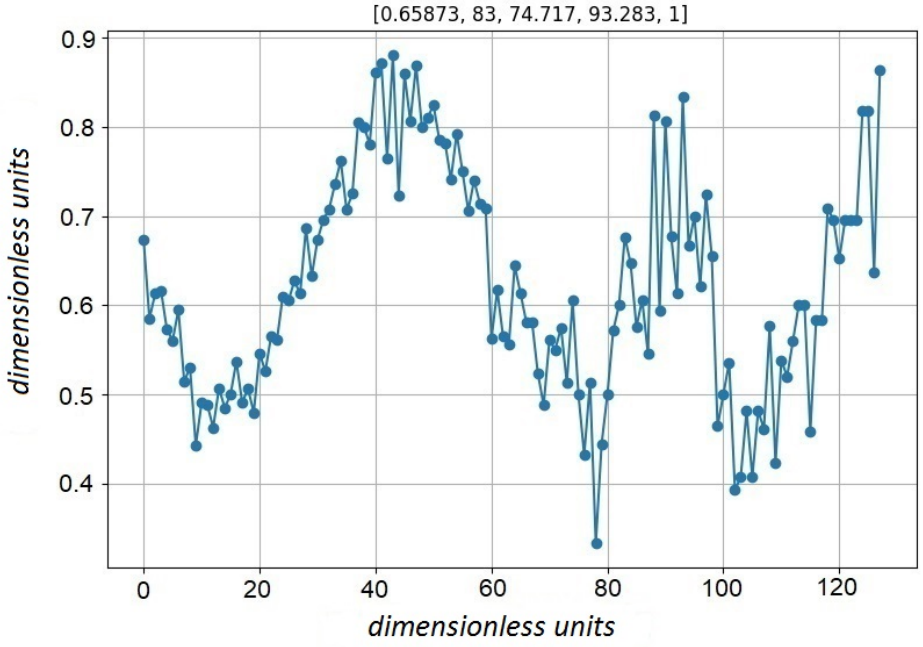

Figure 13. 128-pnp complex for the signal (Figure 12).

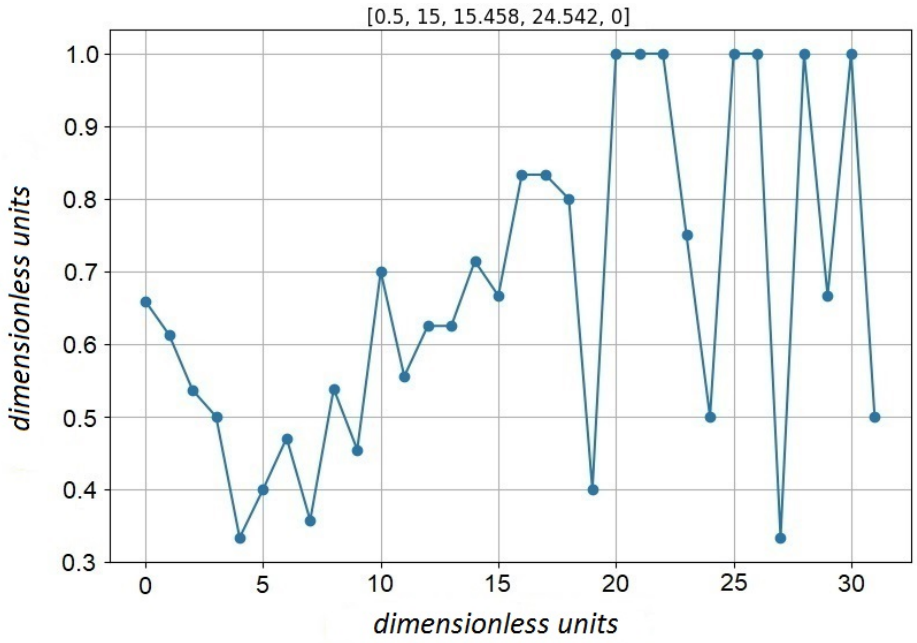

Figure 14. 128-32-pnp second order complex for the signal (Figure 12).

chaoticity index of the first-order complex, and the second bit (the right bit of the input vector) serves as an index of the chaoticity of the second-order complex. Thus we have:

(00) – class (gs) good signal,
(10) or (01) – class (bs) bad signal,
(11) – class (nn) noise signal.

Let us consider an example of the practical application of the classifier we developed. We tried to estimate the time distribution of the degree of "activity" of the observed infrasound background. At the same time, it was agreed to evaluate "activity" in proportion to the number of ten-second fragments of the time series assessed by the classifier as "gs" divided by the total number of ten-second fragments in the hourly interval. To analyze the "activity", about 390 calculations were performed for each hourly fragment of the time

series. Thus, for each approximately 10-second fragment of a time series containing 3000 samples, a 128-pnp first-order complex was calculated along with an assessment of its "chaoticity" and a 128-32-pnp second-order complex, also along with an assessment of "chaoticity". Then a classification procedure was launched, assigning the current 10-second fragment to one of the three classes described above. Thus, by adding one (if the current fragment belongs to a given class) or zero, three lists of 10-second fragments were formed – members of the given classes. At the end of processing the hourly fragment of the time series, the single values of the lists were summed up, and the resulting sum was divided by the total number of 10-second fragments in the analyzed hourly interval. As a result, for each hourly fragment of the series, three numbers were obtained, the sum of which gave one. For our task in the future, we used only the "gs" list, conditionally containing information about the "activity" of the observed time series.

In a similar way, 120 hours of time series were analyzed and an array of 120 points was obtained, the value of which should, according to our hypothesis, characterize the "activity" of the infrasound background recorded at our infrasound monitoring point. The result of the analysis is shown in Figure 15 and presents information for five days from the first to the fifth of September 2023. The graph shown in Figure 15 clearly shows five peaks of infrasonic "activity" occurring during the night hours of the analyzed time series. Moreover, it should be noted that the peaks relating to the second, third and fourth of September show a significant increase in the "activity" under study.

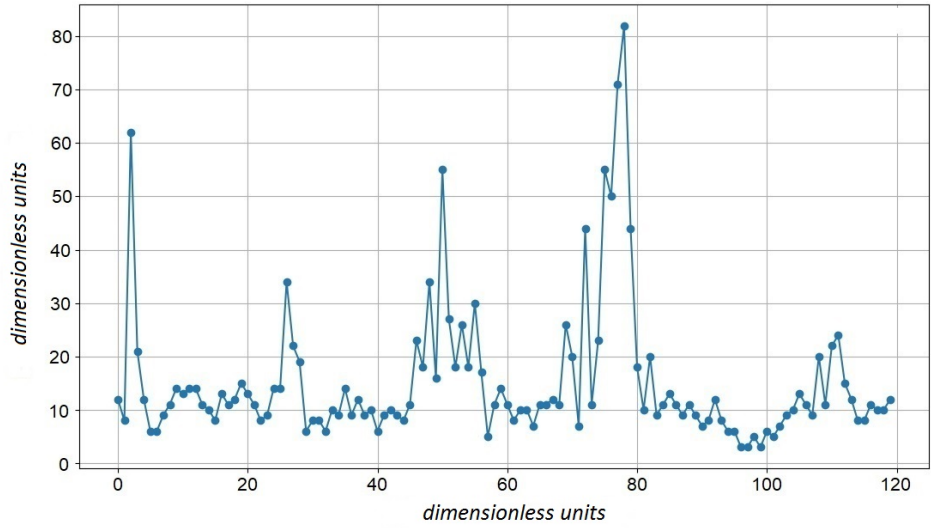

Figure 15. Infrasound activity for the period from September 1 to September 5, 2023. Gorno-Altaisk.

Further research showed that for samples containing a relatively small number of points, the k-complex randomness measure based on turning points is unstable. An alternative criterion has been developed especially for such cases.

In order to verify with sufficient reliability the adequacy of the algorithm, we take verified geomagnetic data containing noise and a useful signal, isolated using indices of geomagnetic activity or simultaneous recording of morphology at several spatially distributed observatories. Figure 4 shows a time series representing the difference signal between the vector sum of projections in the Euclidean metric and the magnitude of the Earth's magnetic field induction vector. This value is a generally accepted indicator of the quality of measurements obtained using observatory instruments, and is called Delta F (component G) [*St-Louis*, 2020]:

$$G = F(v) - F(s), \tag{3}$$

where $F(v)$ – modulus of the magnetic field strength vector, calculated from vector magnetometer data, $F(s)$ – direct measurements of the modulus on a scalar magnetometer.

The measurements were carried out at the SPG magnetic observatory [*Sidorov et al.*, 2017], part of the INTERMAGNET network, on 02/11/2024. Figure 4 itself shows the hour interval (3600 counts) from 2 am to 3 am GMT. The X axis shows time, the Y axis shows the value of magnetic induction in nT. Conventionally, the task of the "classifier" was to detect and isolate from the background noise the signal shown on the graph in the time interval approximately between 02:08:30. and 02:18:30. At the same time, the "classifier" should not have responded to any other signals.

We decided to divide the presented hourly fragment of the time series into ten-minute intervals containing 600 samples and to build for each interval a 108-pnp complex of the first order. For the desired fragment of the time series, the 108-pnp complex of the first order is presented in Figure 16. The red color in the figure shows a fourth-degree polynomial (`np.polyfit()`), which approximates the points of this complex. Below in green is a graph of the absolute values of the element-wise difference between the points of our 108-pnp complex and the approximating polynomial.

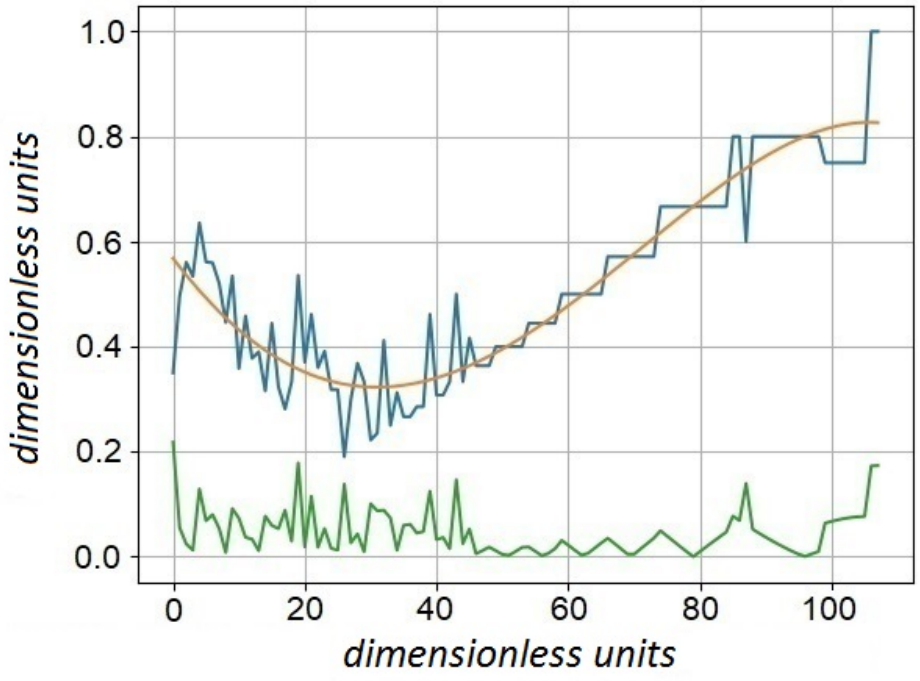

Figure 16. 108-pnp-complex of the first order.

It was hypothesized that the smaller is the standard deviation for the vector of absolute values of the element-wise difference between the points of our 108-pnp-complex and the approximating polynomial, the less random time series corresponds to the k-complex under consideration. Thus, the value of the desired standard deviation can serve as a threshold criterion for identifying more or less noisy signals.

To isolate a ten-minute fragment of the signal shown in Figure 17, the criterion STD < 0.06 was set. This criterion worked only once, because only for the k-complex presented in Figure 16, the STD value turned out to be 0.044.

As the analysis of the initial data shows (Figure 18), this interval is a natural signal and is observed at other observatories and stations of the network at the same time (Figure 19). It should be noted that isolating a useful signal using delta F analysis of one observatory is generally not an easy task that arises in the process of processing geomagnetic data [*Soloviev et al.*, 2018]. The proposed algorithm can be used to automate the selection of useful intervals when processing 1-second geomagnetic data.

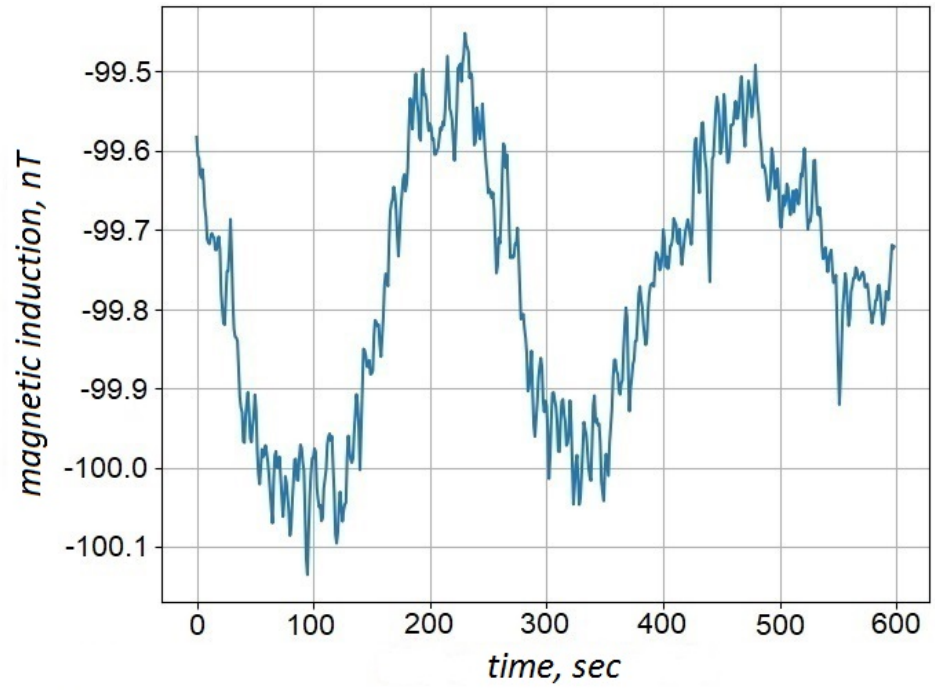

Figure 17. 108-pnp-complex of the first order.

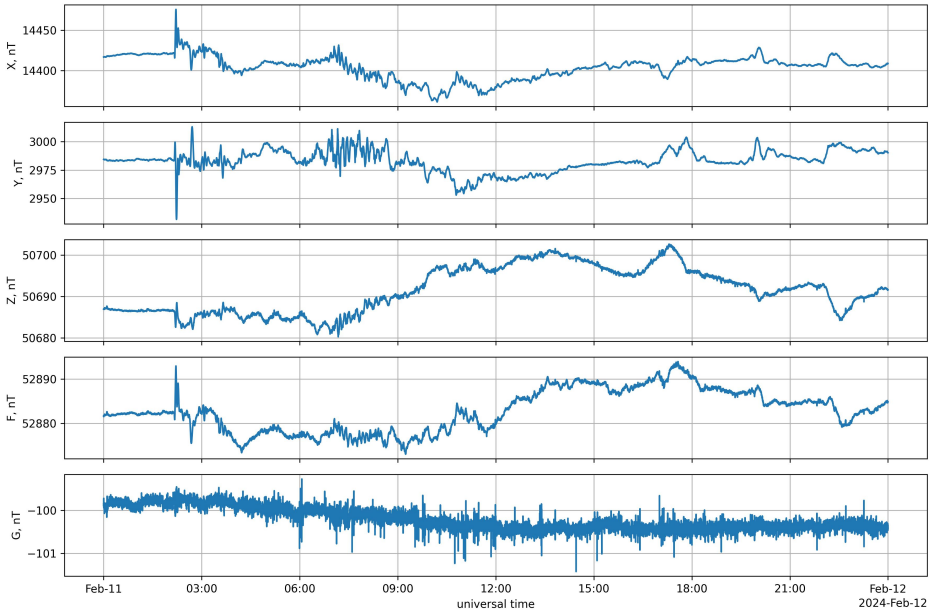

Figure 18. Fragment of data from magnetometers of the SPG observatory around 2024-02–11 02:00 UT.

Conclusion

As a result of the research presented in this work, an algorithm was proposed for detecting fragments of a time series that meet the specified "non-randomness" criteria. The proposed "classifying" algorithm is called a complex indicator of the stability of the permutation entropy of fragments of a time series. Its properties are based on reducing the number of studied points of a fragment of a time series by calculating the k-complex of some easily calculated parametric indicator of the studied fragment and subsequent assessment of the degree of randomness of the resulting k-complex. In our case, the normalized indicator of the number of turning points calculated for k sparse samples from

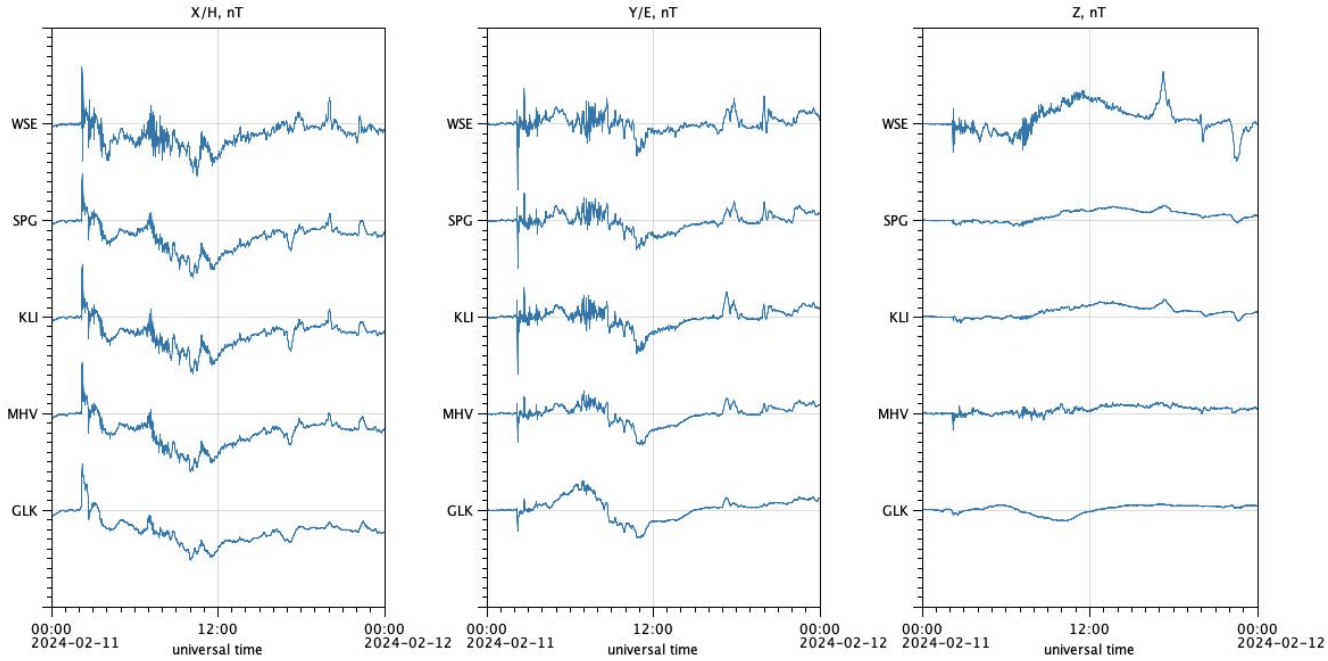

Figure 19. Comparison of vector components around 2024-02-11 UT at the magnetic observatories St. Petersburg (SPG), Klimovskaya (KLI), Mikhnevo (MHV), White Sea (WSE), Gyulagarak (GLK).

the desired fragment of the time series was used as such an indicator. For sufficiently large fragments of the time series (the first example considered was 3000 points), the degree of randomness was assessed quite well by the criterion of turning points applied to k-complexes of the first and second order. For relatively small samples (for example, 600 points), the degree of randomness was estimated by the minimum value of the standard deviation for the vector of absolute values of the element-wise difference between the points of our k-pnp-complex and the approximating polynomial of the fourth degree.

Acknowledgments. The facilities of the GC RAS Common Use Center "Analytical Center of Geomagnetic Data" (http://ckp.gcras.ru/) were used for conducting the research. The research was carried out with funds from the Russian Science Foundation (RSF) and the Ministry of Education and Science of the Altai Republic No. 23-21-10087.

References

Chumak, O. V. (2012), *Entropies and fractals in data analysis*, R&C Dynamics, https://doi.org/10.13140/2.1.4739.6800 (in Russian).

Fu, S., Y. Huang, T. Feng, D. Nian, and Z. Fu (2019), Regional contrasting DTR's predictability over China, *Physica A: Statistical Mechanics and its Applications*, *521*, 282–292, https://doi.org/10.1016/j.physa.2019.01.077.

Higuchi, T. (1988), Approach to an irregular time series on the basis of the fractal theory, *Physica D: Nonlinear Phenomena*, *31*(2), 277–283, https://doi.org/10.1016/0167-2789(88)90081-4.

Kandal, M. (1981), *Time series*, Finance and Statistics, Moscow (in Russian).

Liang, T., G. Xie, D. Mi, W. Jiang, and G. Xu (2020), PM2.5 Concentration Forecasting Based on Data Preprocessing Strategy and LSTM Neural Network, *International Journal of Machine Learning and Computing*, *10*(6), 729–734, https://doi.org/10.18178/ijmlc.2020.10.6.997.

Lu, P., L. Ye, M. Pei, Y. Zhao, B. Dai, and Z. Li (2022), Short-term wind power forecasting based on meteorological feature extraction and optimization strategy, *Renewable Energy*, *184*, 642–661, https://doi.org/10.1016/j.renene.2021.11.072.

Microsin.net (2020), INMP441: digital microphone with interface I2S, https://microsin.net/adminstuff/hardware/inmp441-i2s-omnidirectional-digital-microphone.html (in Russian), (visited on 18.02.2024).

Roushangar, K., R. Ghasempour, V. S. O. Kirca, and M. C. Demirel (2021), Hybrid point and interval prediction approaches for drought modeling using ground-based and remote sensing data, *Hydrology Research*, *52*(6), 1469–1489, https://doi.org/10.2166/nh.2021.028.

Schwardt, M., C. Pilger, P. Gaebler, P. Hupe, and L. Ceranna (2022), Natural and Anthropogenic Sources of Seismic, Hydroacoustic, and Infrasonic Waves: Waveforms and Spectral Characteristics (and Their Applicability for Sensor Calibration), *Surveys in Geophysics*, *43*(5), 1265–1361, https://doi.org/10.1007/s10712-022-09713-4.

Sidorov, R., A. Soloviev, R. Krasnoperov, D. Kudin, A. Grudnev, Y. Kopytenko, A. Kotikov, and P. Sergushin (2017), Saint Petersburg magnetic observatory: from Voeikovo subdivision to INTERMAGNET certification, *Geoscientific Instrumentation, Methods and Data Systems*, *6*(2), 473–485, https://doi.org/10.5194/gi-6-473-2017.

Silva, A. S. A., R. S. C. Menezes, O. A. Rosso, B. Stosic, and T. Stosic (2021), Complexity entropy-analysis of monthly rainfall time series in northeastern Brazil, *Chaos, Solitons & Fractals*, *143*, https://doi.org/10.1016/j.chaos.2020.110623.

Soloviev, A., V. Lesur, and D. Kudin (2018), On the feasibility of routine baseline improvement in processing of geomagnetic observatory data, *Earth, Planets and Space*, *70*(1), https://doi.org/10.1186/s40623-018-0786-8.

St-Louis, B. (Ed.) (2020), *INTERMAGNET Technical Reference Manual*, *Version 5.0.0*, INTERMAGNET Operations Committee and Executive Council, https://doi.org/10.48440/INTERMAGNET.2020.001.

Traversaro, F., F. O. Redelico, M. R. Risk, A. C. Frery, and O. A. Rosso (2018), Bandt-Pompe symbolization dynamics for time series with tied values: A data-driven approach, *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *28*(7), https://doi.org/10.1063/1.5022021.

Zhang, T., C. Cheng, and P. Gao (2019), Permutation Entropy-Based Analysis of Temperature Complexity Spatial-Temporal Variation and Its Driving Factors in China, *Entropy*, *21*(10), 1001, https://doi.org/10.3390/e21101001.

Zhu, G., J. Hunter, and Y. Jiang (2016), Improved Prediction of Dengue Outbreak Using the Delay Permutation Entropy, in *2016 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, IEEE, https://doi.org/10.1109/iThings-GreenCom-CPSCom-SmartData.2016.172.

Zunino, L., M. C. Soriano, and O. A. Rosso (2012), Distinguishing chaotic and stochastic dynamics from time series by using a multiscale symbolic approach, *Physical Review E*, *86*(4), https://doi.org/10.1103/PhysRevE.86.046210.