ISSN 0001-4338, Izvestiya, Atmospheric and Oceanic Physics, 2017, Vol. 53, No. 8, pp. 847–858. © Pleiades Publishing, Ltd., 2017.
Original Russian Text © A.V. Desherevskii, V.I. Zhuravlev, A.N. Nikolsky, A.Ya. Sidorin, 2017, published in Geofizicheskie Protsessy i Biosfera, 2017, Vol. 16, No. 2, pp. 55–73.

Analysis of Rhythms in Experimental Signals

A. V. Desherevskii*, V. I. Zhuravlev, A. N. Nikolsky***, and A. Ya. Sidorin******

Schmidt Institute of Physics of the Earth, Russian Academy of Sciences, Moscow, Russia
**e-mail: adeshere@ifz.ru*
***e-mail: vzhtvertsa@gmail.com*
****e-mail: anickol@yahoo.com*
*****e-mail: al_sidorin@hotmail.com*

Abstract—We compare algorithms designed to extract quasiperiodic components of a signal and estimate the amplitude, phase, stability, and other characteristics of a rhythm in a sliding window in the presence of data gaps. Each algorithm relies on its own rhythm model; therefore, it is necessary to use different algorithms depending on the research objectives. The described set of algorithms and methods is implemented in the WinABD software package, which includes a time-series database management system, a powerful research complex, and an interactive data-visualization environment.

Keywords: rhythm, hidden periodicity, parameter estimation, algorithm, WinABD

DOI: 10.1134/S0001433817080035

INTRODUCTION

For many years, the authors of this work have been conducting regime observations at various geophysical areas. We found that daily, seasonal, and other rhythms were present in most of the observed processes. For the analysis of their characteristics, various algorithms were created, including those that are tolerant of missed observations and other data defects. All these methods can be applied not only in geophysical applications, but also in medicine, biology, economics, etc.

The aim of this work is to present some methods of rhythm analysis which are suitable for data series with gaps. We also aimed to demonstrate the features of these methods using an example of experimental signals of the motor activity of bioindicators, which were obtained at the Gharm Testing Area within the work on searching for earthquake precursors (Sidorin, 1990).

APPLIED DATA

To illustrate the features of the developed methods of signal processing, two series of signals of a biological nature with sharply differing characteristics of the daily rhythm are selected from the database (Deshcherevskii, Sidorin, 2002). The first series shows the number of electrical pulses of a Nile elephant fish (*Gnathonemus leopoldianus*), the SLON2M series; a mildly expressed, relatively smooth, circadian rhythm is typical for this series, which is almost invisible against the background of other variations. The second series—the SOM4—shows the number of mechanical impulses (surfacing) for a catfish (*Hoplosternum thoracatum*); the diurnal rhythm in this series is a brief and very sharp burst of activity, the amplitude of which varies greatly from day to day, and the time shifts by 1–2 h. The possible spectral range of such a signal can be quite wide, but due to the specificity of measurements with strong antialiasing and averaging (the number of pulses per hour was counted in observations) we consider only the daily cycle of activity.

Both bioobjects were observed for several years. All calculations were performed for the full series, but for greater clarity only small fragments of the series are shown on the graphs, which allow us to analyze and compare the selected signals in detail. The data from 15 days (from November 13 to November 27, 1983) are shown for the SLON2M series (see Fig. 1a, curve *1*) and from 18 days for the SOM4 series (from February 13 to March 1, 1992; see Fig. 1b, curve *1*). Since the SLON2M series contained a trend, low-frequency variations were filtered from it prior to analysis with periods of more than 168 h.

ALGORITHMS FOR RHYTHMS EXTRACTION

The Need to "Refine" the Rhythm before Evaluating Its Parameters

In the analysis of rhythms, their phase and amplitude characteristics are usually studied (Komarov, 1989). Also, the shape of the rhythm, its stability, and other parameters can be considered (Deshcherevskii and Sidorin, 2003). These and other characteristics of rhythm can be estimated directly from the original

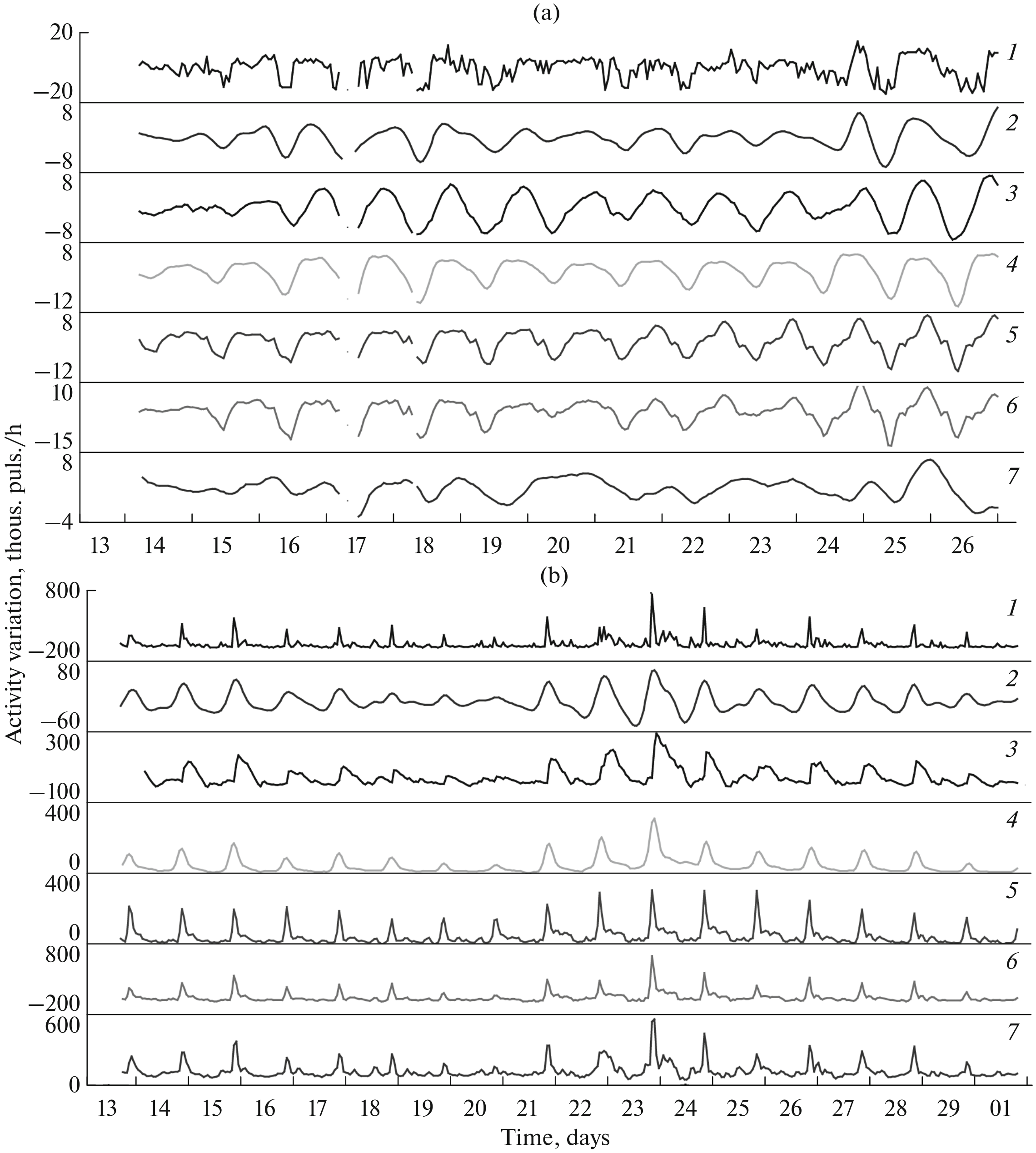

Fig. 1. Extraction of diurnal rhythm by different methods: (a) SLON2M series (November 13–27, 1983); (b) SOM4 series (February 13 to March 3, 1992). Original signal (*1*) and refined rhythm according to the models M1–M6 respectively (*2*–*7*).

S series; however, the actual experimental signals always contain various types of noise, which is why the rhythm parameters are evaluated with a very large error. For this reason, before evaluating the rhythm parameters, it is advisable to extract it in a refined form, i.e., to filter it from noise.

Filtering the signal to increase the signal-to-noise ratio is a typical signal-processing technique. Frequency filtering is mostly used, which is based on differences in the spectral properties of the noise and the signal being selected (Kanasevich, 1981; Hemming, 1987). This approach is effective in the study of quasi-harmonic rhythms, the dispersion of which is concentrated in a limited frequency band. However, rhythms are far from always being sinusoidal (Terebizh, 1992). The deviations from the harmonic function are often not just significant, but constitute the most essential feature of rhythm; for example, the asymmetry of the seasonal course, biorhythms, etc. To describe such variations and study their characteristics, the models that consider rhythm as a periodic function R of a complex form are more adequate.

The effectiveness of such models was shown by us when analyzing seasonal variations of electrical parameters at the Gharm Testing Area (Sidorin, 1990); therefore, the name "average seasonal function" was used for the rhythm function **R** (Deshcheverevskii et al., 1996; Deshcherevskii and Sidorin, 1999). This approach made it possible to detect the flicker-noise character of nonseasonal components of variations of the parameters under consideration (Deshcherevskii et al., 1997b) and also to obtain interesting results, in particular, in analyzing the seasonal variations of apparent resistance (Deshcherevskii et al., 1997a; Deshcherevskii and Sidorin, 2004).

In some cases (for example, in astrophysics (Terebizh, 1992)), the form of the rhythm function **R** can be specified from a priori considerations. In biological, geophysical, and other applications, the form of variation should, as a rule, be estimated empirically. The most natural way of such an assessment is based on the method of epoch superposition. This method assumes that we only know the value of the period of rhythm originally. In addition to this—for some types of rhythms—one can assume a certain degree of smoothness of the periodic function or impose other restrictions on it, which makes it possible to improve the estimate (Deshcherevskii and Sidorin, 1999).

Knowing the model rhythm function **R**, you can use it to extract a purified signal $\hat{\mathbf{S}}$. As a rule, the signal **S** is approximated by the function **R** in some window. In the case of a linear model, the approximating function $\hat{\mathbf{S}}$ is calculated as

$$\hat{\mathbf{S}} = A \cdot \mathbf{R} + B, \tag{1a}$$

where A and B are the coefficients of the model, which are chosen to minimize the remainder **Z**:

$$\mathbf{S} = \hat{\mathbf{S}} + \mathbf{Z} = A \cdot \mathbf{R} + B + \mathbf{Z}. \tag{1b}$$

Bold type in Eq. (1) designates vector values that correspond to the time series within the window (Deshcherevskii et al., 2016a).

An estimation of coefficients A and B and a calculation of the values of approximating function $\hat{\mathbf{S}}$ are performed anew for each position of the time window. The wider it is, the stronger noise is suppressed and more stable the parameters of rhythm are. However, to improve the temporal resolution or in the case of strong variability of rhythm composition, it is advisable to choose a window with a small width. To ensure that the duty cycle of the filtered signal corresponds to the duty cycle of the source series, the evaluation window is shifted to one point at each step (Deshcherevskii et al., 2016b). Among other things, this avoids sharp jumps in parameter values at the boundaries of the window and ensures that function $\hat{\mathbf{S}}$ is smooth and continuous.

Depending on the problem being solved, the calculation of $\hat{\mathbf{S}}$ can be performed in the middle of the window (which increases stability and improves allocation of rhythm) or on its right-hand boundary, which gives the possibility to extrapolate signal $\hat{\mathbf{S}}$ forward into the future (beyond the window) when solving forecast problems. Having the model or cleaned rhythm $\hat{\mathbf{S}}$ formed, it is possible then to evaluate its amplitude, phase, stability, and other characteristics. In addition, signal $\hat{\mathbf{S}}$ can be used to fill data gaps in the original signal, which substantially increases the stability of estimation of any parameters of rhythm and can also be useful in other calculations (Deshcherevskii et al., 2016b).

Methods of Rhythm Extraction

Let us consider the methods or algorithms that we used to extract the refined rhythm. For brevity, they are indicated as M1–M6 below.

Algorithm M1. Bandpass frequency filtering. The standard algorithm of frequency filtration is based on the calculation of the filter coefficients in accordance with the given parameters of the gain-frequency characteristic (usually the limits of the suppression/transmission band and the slope of the filter cut are specified) (Kanasevich, 1981; Hemming, 1987). This algorithm generates alternating filtering sequences that require a uniform step between samples and, therefore, do not allow data gaps. For series with an uneven time step, the calculation of special filtering sequences is possible. However, a recalculation of the filter coefficients at each step, taking into account the actual distribution of gaps in the window, would require a disproportionate computing time. Therefore, we used the method of sliding kernel smoothing (with the Gaussian kernel) for frequency filtration (Hardle, 1993; Lagutin, 2009).

Due to positively defined weighting function of the kernel, this algorithm is tolerant to data gaps. Using the so-called "collapse of the window" allows us to perform filtering without decreasing the length of the filtered signal, not only at the beginning and end of the series, but also at the boundaries of the data-gap intervals (Deshcherevskii et al., 2016a–2016d). To extract the daily rhythm from the signal, we excluded (filtered) the variations with periods of less than 17 and more than 35 h.

With frequency filtering, function $\hat{\mathbf{S}}$ is constructed without using the model (1).

Algorithm M2. Approximation by a sinusoid. As a model rhythm **R** in the model (1), a sinusoid fragment with a period P is used (Deshcherevskii et al., 1996). The degree of adaptability of this algorithm can be adjusted by varying the width of the approximation window, and it can be either less or more than the period P, depending on the priorities: maximum noise suppression or improvement of time resolution of the method. Given the properties of the signals, a 48-h window was used when processing the SLON2M series and a 12-h window when processing the SOM4 series.

Algorithm M3. Approximation by an average rhythm with adjustment of amplitude. The average oscillation estimated by epoch superposition is used as a model rhythm **R** (Deshcherevskii and Sidorin, 1999):

$$R(t) = (1/N)\sum(S(t + iP)), \tag{2}$$

where t varies from 1 to p, summation is over all integers i for which the point of time $t + iP$ lies within the series, and N is the number of summable values (non-gaps). Thus, the series is as if cut along the length of the period to a certain number of parts (epochs) of duration P, and then all these epochs are superimposed and averaged. Then, the constructed image of the rhythm is further smoothed out and coherently (i.e., with the conservation of phase) multiplied (duplicated) on the entire time axis:

$$R(t + iP) \equiv R(t). \tag{3}$$

In fact, this means that the current rhythm is compared with an ideally stable reference rhythm that has a fixed shape and phase throughout the time axis. Thus, the M3 algorithm does not suggest the possibility of changing the phase of rhythm and does not allow us to evaluate these changes, even if they actually exist. On the other hand, excluding one degree of freedom (phase variability) from the model increases the stability of amplitude estimation in the case when the rhythm phase does not really change.

For highly noisy rhythms, the error in estimating the coefficients of regression model (1) can be reduced by increasing the width of the sliding window. If the rhythm is sufficiently clear, its value of sliding window can be reduced, which increases the temporal resolution. In any case, the window width should be a multiple of period P. Taking these considerations into account, we used a sliding window with a width of 72 h when processing the SLON2M series and 24 h for the SOM4 series.

Algorithms M4 and M5. Approximation by the current average rhythm. Both these algorithms are based on a calculation of the average rhythm $R(t)$ by Eq. (2). However, unlike the M3 algorithm, the evaluation is performed not over the entire row, but in a sliding window whose width is chosen to be several periods P. This leads to the fact that the shape of the average rhythm $\mathbf{R}$ (and, consequently, its phase characteristics) can vary with a window displacement.

It is assumed in algorithm M4 that the resulting function R is a refined rhythm: $\hat{\mathbf{S}} \equiv \mathbf{R}$, i.e., approximation by model (1) is not used. The M5 algorithm differs from algorithm M4 in such a way that, after evaluating shape of the rhythm, the amplitude of the rhythm is additionally adjusted according to model (1). It is usually assumed that the shape of the rhythm varies more slowly than its amplitude. Therefore, function $\mathbf{R}$ is evaluated in a relatively wide window, while parameters of the model (1) are evaluated in narrower window. The M5 algorithm tracks only slow enough changes in the rhythm phase.

For the SLON2M series, we estimated the shape of rhythm using Eq. (2) in a 7-day-wide window, and the amplitude adjustment (Eqs. (1), (2)) was performed in a 2-day window. For the SOM4 series, windows with a width of 5 and 1 day were used.

Algorithm M6. Approximation by the epoch superposition method. The algorithm of epoch superposition is also based on the use of model (1). However, the average rhythm $\mathbf{R}$ calculated according to Eq. (2) in this case moves along with the window; i.e., unlike the algorithm M3, only the $\mathbf{S}$ series is fixed now. It is as if the fragment of the $\mathbf{R}$ series that hit the sliding window in its initial position "freezes" into this window and then moves along with it. Thus, the phase of the average rhythm $\mathbf{R}$ is fixed not relative to the beginning of the signal, but relative to the beginning of the window. If the reference variation fits well with the actual vibration at some position of the sliding window, the correlation between them approaches unity and the regression coefficient shows the exact amplitude of the rhythm.

Thus, this algorithm allows us to take into account variations, amplitudes, and phases of the rhythm simultaneously without restriction on the magnitude of the phase shift. The width of the sliding window in this algorithm is usually chosen to be equal to the period of oscillation, which was applied in this present work.

Comparison of Filtered Rhythm

The results of the diurnal rhythm extraction by different methods are shown in Fig. 1. As can be seen in Fig. 1a, for the SLON2M series (where the amplitude of the rhythm is small) the differences between filtered signals $\hat{\mathbf{S}}$ are very significant. In fact, each model defines the concept of "rhythm" in its own way, which leads to such a significant difference in the results. For the SOM4 series (see Fig. 1b), there are also significant differences between the models. However, the differences in this case are due to the fact that some models are clearly not optimal; i.e., the results of the filtration do not correspond to the intuitive notions on the structure of the rhythm. Let us consider in more detail the features of the described algorithms.

SLON2M series. As can be seen in Fig. 1a, frequency filtering (algorithm M1, curve *2*) identifies a very unstable diurnal rhythm. An approximation of the rhythm with a sinusoid in a 48-h window (algorithm M2, curve *3*) yields close results, but there are significant differences in details (shape, amplitude, and phase of the rhythm). The visual comparison of the rhythms selected (curves *2*, *3*) with the original series (curve *1*) shows that the models that have a harmonic basis are apparently not quite adequate to the characteristics of this series. When approximating by an average rhythm with amplitude adjustment (algorithm M3, curve *4*), the form of the variation is fixed, and only its amplitude varies. At the same time, a comparison of the refined rhythm with the original signal shows that the correspondence between them is rather coarse.

When the average rhythm is approximated in a sliding window (algorithm M4, curve *5*), the shape of the rhythm varies noticeably within the time interval under consideration. It can be seen that the method is quite successful in simulating the daily rhythm during the period from November 15 to 18: the results of filtration are rather close to the a priori (subjective) concept about the rhythm selected. However, this algorithm does not track rapid changes in the nature of the rhythm. Therefore, on the night of November 23–24, there is no usual night outburst of activity in the original signal, which is clearly seen from the original series (curve *1*); however, a similar effect is not noticeable in the filtered rhythm (curve *5*). The change in the shape of the rhythm on November 25 is also not reflected.

The M5 algorithm provides a much more accurate adjustment to the amplitude of the rhythm is (curve *6*), although it also does not cope with the change in the rhythm shape on November 25. However, it should not be forgotten that this algorithm has the largest number of adjustable parameters (degrees of freedom). This means that, upon an unfavorable combination of circumstances, this algorithm can be adjusted not only to the rhythm, but also to random noises. For the signal under consideration, this algorithm seems to correspond most closely to our a priori concepts about the nature of rhythm.

The rhythm extracted by the M6 algorithm seems to correspond least of all to our intuitive understanding of the biological rhythm (Fig. 1a, curve *7*). It can be seen that an algorithm that allows arbitrary phase variations has a much worse selectivity than the algorithms considered above. However, it is possible that it is our expectations that are wrong rather than the model. When analyzing such a complex signal as a series of motor activity, it is impossible to determine the optimal rhythm pattern a priori, since we do not have a clear understanding of the causes and patterns of changes in the motor activity of a fish. One can only assume that the daily rhythm of activity exists, as this is confirmed by both an analysis of the signal spectra and a priori representations.

In this situation, the choice of the optimal model of rhythm can be made only through expert evaluation. Each of the refined rhythms shown in Fig. 1a implements one of the possible models of rhythm. Given the complex structure of the original series and the low signal-to-noise ratio, it is not surprising that the models differ significantly. At the same time, most algorithms demonstrate a similar dynamics of the change in the amplitude of the rhythm over time, which indicates the reality of these changes.

SOM4 series. This signal is the exact opposite of the previous case. The diurnal rhythm is distinguished very clearly, as a rule, in the form of a burst of activity of very large amplitude (see Fig. 1b, curve *1*). However, its phase is unstable, and the amplitude falls almost to zero on some days. Quite often, along with the main burst of activity, there are side maxima of smaller amplitude. In this case, the task of a formalized description of such a rhythm is not obvious. Let us see how suitable the algorithms are for solving it.

With frequency filtering (algorithm M1), the filtered rhythm (see Fig. 1b, curve *2*) gives a certain idea on the relative magnitude of the bursts; however, the amplitude of the rhythm decreased by an order of magnitude in absolute terms. The shape of the rhythm was distorted beyond recognition, and the phase undergoes obvious biases each time when there are additional bursts of activity in the original signal.

When the sinusoid is used for approximation (algorithm M2), the window width of the regression model (1) was chosen to be 12 h in order to enhance the adaptive properties of the method (see Fig. 1b, curve *3*). Reducing the window size degrades the signal-to-noise ratio of the filtered signal, but improves the temporal resolution. Indeed, the amplitude of the rhythm is traced slightly better in this case than with frequency filtering (compare curves *2* and *3* in Fig. 1b). However, the filtered series is very noisy, and the shape of the extremum still "blurs" and poorly conveys the sharp features inherent in the original curve; distortions of the phase are also noticeable. A fairly obvious fact is confirmed that the methods M1 and M2, based on the harmonic model, are of little use for the analysis of rhythms with sharp bursts.

When using M3 algorithm (curve *4*), the averaged rhythm **R** must be further smoothed to ensure that bursts that are shifted in time relative to the average position are tracked. This is due to the fact that this algorithm does not allow phase variation. Therefore, the width of the local burst in signal **R** must be sufficient to allow the actual burst of activity to coincide with the burst of activity in **R** even for a small phase shift. If this alignment is ensured (the phase is fixed), the M3 algorithm tracks the amplitude of the rhythm quite well.

When approximation by an average rhythm in a sliding window (algorithm M4, curve *5*) us used, rapid changes in the amplitude of the rhythm (the duration of which is less than the window of the estimation of the function **R**) are not tracked. For the SOM4 series, we used a window with a width of 5 days, which provides good noise suppression. However, the amplitude of the rhythm in this case corresponds to its average amplitude over five periods. In fact, the amplitude varies significantly faster (with a characteristic time of 1–2 days). Therefore, the M4 model is not very suitable for the SOM4 rhythm. The inclusion of amplitude tuning (algorithm M5, curve *6*) leads to a more realistic picture. However, the resolution improvement is achieved at the cost of less effective suppression of random noise.

Quite ambiguous results were obtained using M6 algorithm (see Fig. 1b, curve *7*). On the one hand, the filtered rhythm reflects all the important features of the original signal well enough. On the other hand, it also contains quite a few small (random) features, which can be considered a drawback of the algorithm. However, without a clear idea of the criteria for optimality, it is possible to compare different models only through expert assessments.

EVALUATION OF THE AMPLITUDE OF RHYTHM

Methods for Estimating the Amplitude of Rhythm

One of the most important parameters of the rhythm is its amplitude. When monitoring the rhythm parameters, the amplitude can be measured in different ways. In the simplest case, amplitude A is estimated as the difference between the maximum and minimum value of the series over the period. Let us to designate an amplitude as $A_{\max-\min}$. To estimate $A_{\max-\min}(t)$ at the time t, the filtered signal $\hat{S}(t)$ is considered in the time interval from $t - P/2$ to $t + P/2$, where P is the period. The maximum $\mathrm{Max}(\hat{S})$ and minimum $\mathrm{Min}(\hat{S})$ values of the signal in the specified interval are sought after. Amplitude is estimated by formula

$$A_{\text{max-min}}(t) = \mathrm{Max}(\hat{S}) - \mathrm{Min}(\hat{S}). \quad (5)$$

Then the sliding window of amplitude estimation is shifted by one point. This allows us to calculate variations $A_{\max-\min}(t)$ with the same resolution as the original signal (and not with the step P, as is done sometimes). Such a method of constructing the $A_{\max-\min}(t)$ series is useful not so much for a more detailed representation of the time scanning (the use of window estimates in any case leads to a certain averaging–smoothing), but rather for convenience in the subsequent joint processing of the initial and parametric (calculated) series.

Another method for estimating the amplitude is to measure a variance of the filtered signal in the same time window $(t - P/2, t + P/2)$. For practical purposes, it is more convenient to use a standard deviation rather than variance, since it has the same dimension as the signal. We denote the amplitude measured in this way as A_σ. It should be noted that both estimates $A_{\max-\min}$ and A_σ are not applicable for a very noisy signal, since in this case the amplitude or standard deviation of the noise rather than the signal is measured. However, for the filtered rhythm, both methods are fully functional.

When regression model (1) is used to estimate the amplitude of the rhythm, one can analyze coefficient A of this model. Let us designate the amplitude of the rhythm measured in such a way as A_{regr}. This estimate is convenient, because it is much less subject to the influence of various random factors, such as separate outbursts or data gaps. Unfortunately, it is not always possible to evaluate the amplitude in this way. Therefore, with frequency filtering (algorithm M1) or when the signal is approximated by the average rhythm in the sliding window (M4 algorithm), the regression model cannot be constructed and the A_{regr} coefficient cannot be calculated. Although the regression coefficient can be estimated in the case of the M5 model, to measure the amplitude of the rhythm it still has to be multiplied by the amplitude of the reference R, which changes when the window is shifted. Unfortunately, this usually leads to an additional noise of the resulting curve, since the random features of the two functions are multiplied. Therefore, in practice, this method of estimating the amplitude of rhythm is hardly advisable.

Although the M6 algorithm uses a regression model, the A coefficient in this case has a different meaning and does not imply the amplitude of rhythm. The point is that frame R is "frozen" in the sliding window in the M6 algorithm. Therefore, as the window moves, the phase difference between the frame R and the rhythm under study cycles through all values from 0 to 2π and so on. Therefore, the time course of the regression coefficient A looks like a quasiperiodic function in this case, which reaches its maximum when the phase of the reference and actual rhythms coincide and the minimum when they are in antiphase. In particular, if both the studied and the model rhythm are two identical sinusoids, the graph of the change of the coefficient A will also be a sinusoid whose values are $+1$ when the phases coincide and -1 when the phases are shifted by π. In fact, the amplitude of the rhythm in this case is determined by the maximum value of the coefficient A over the period. This allows us to estimate the regression amplitude only with a resolution (time step) P, and not with resolution of the original signal, as in the case of other algorithms.

Thus, the amplitude of the rhythm can be estimated in different ways. The most universal methods are those that estimate the amplitude from the filtered signal. An alternative approach is based on the use of coefficients of the regression model of the filter. The analysis shows that the regression method for estimating the amplitude of rhythm is much more resistant to data gaps, and the edge effects are much weaker than when estimating the amplitude of the rhythm by the span or standard deviation. However, the regression estimate is available only for particular rhythm models, which significantly limits its applicability.

The following question arises: how consistent are the estimates of $A_{\max-\min}$, A_σ, and A_{regr}? Figure 2 shows the daily rhythm of the SLON2M series filtered by the M4 algorithm, as well as its amplitude calculated in three ways. As can be seen from the figure, the largest differences between the three estimates of the amplitude of the rhythm are most noticeable at boundaries of the intervals with missed observations. This, in general, is not surprising, since the way gaps are processed when calculating these parameters is different. It is worth noting that the calculations were carried out at $\eta = 49\%$, that is, only those data sets were excluded where there were more than half misses (Deshcherevskii et al., 2016d). If you allow calculations only for those positions of the window where the proportion of

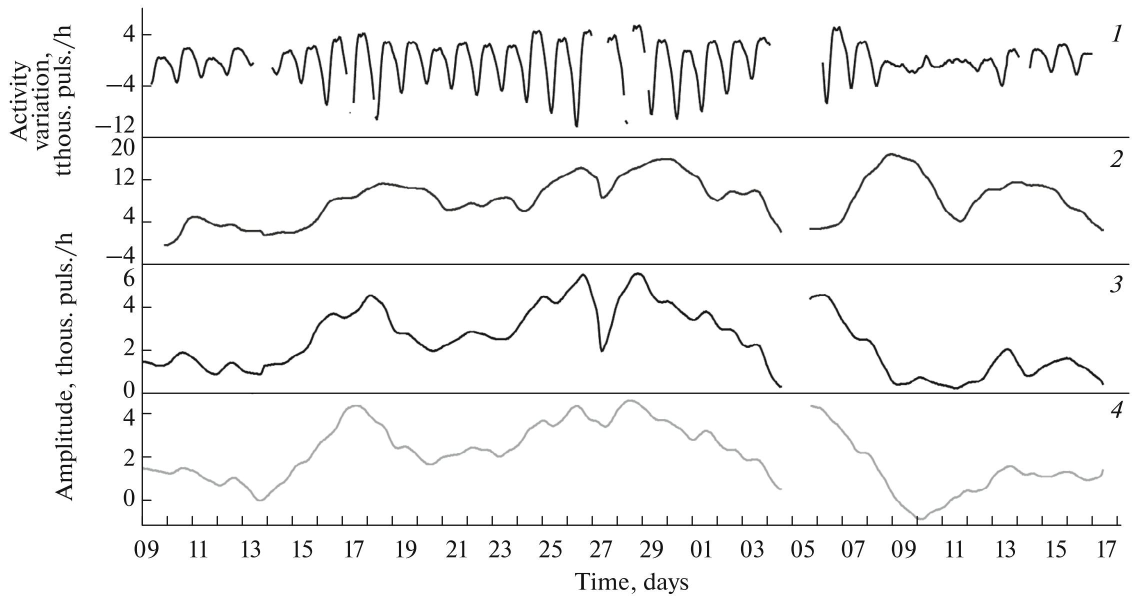

Fig. 2. Amplitude of the diurnal rhythm filtered by M4 algorithm (SLON2M series) estimated in three ways (November 9 to December 18, 1983): filtered rhythm (*1*), amplitude estimation by span (*2*), amplitude estimation by standard deviation (*3*), and amplitude estimation by the regression coefficient of the model (*4*).

misses does not exceed 25%, the differences between the curves are significantly reduced.

For a numerical measurement of the degree of similarity between the estimates, we calculated the correlation coefficient between $A_{\text{max-min}}$ and A_σ for six variants of the filtered rhythm (algorithms M1–M6). It turned out that, for the SLON2M series as a whole and for the fragments shown in Figs. 1–2, it varies in the range of 0.96–0.98 and almost never drops below 0.94. For the SOM4 series, the situation is similar.

The correlation between A_{regr} and $A_{\text{max-min}}$ or A_σ can only be estimated using M2, M3, and M5 algorithms; in the latter case, the A_{regr} evaluation is performed using a more complex algorithm (the regression coefficient is multiplied by the amplitude of the reference rhythm **R**, which must also be estimated in some way). Calculations show that the differences between A_{regr} on the one hand and $A_{\max-\min}$ or A_σ, on the other, are slightly larger than the differences between $A_{\max-\min}$ and A_σ. The correlation coefficient in the pairs (A_{regr}, $A_{\max-\min}$) and (A_{regr}, A_σ) is slightly lower than in the pair ($A_{\max-\min}$, A_σ), but in this case it almost never drops below 0.9. And if you reduce the value of η to 25%, the correlation increases even more and reaches 0.98–0.99.

Thus, for the filtered SLON2M and SOM4 rhythms, all methods for estimating the amplitude give almost identical results.

Dependence of the Amplitude on the Way of Rhythm Selection

Let us now consider the question of how much the amplitude estimate depends on the way of rhythm filtration (extraction). Figure 3 shows the time dependence of $A_{\max-\min}$ for the daily rhythm of SOM4 filtered in different ways (algorithms M1–M6).

As the analysis shows, the differences between these curves are quite noticeable, which is related to the specifics of the rhythm patterns. For example, a decrease in the amplitude of the rhythm on January 16–28 is considered by the M4 algorithm as almost monotonic, while the other algorithms show more detail. There is a rather strong difference between algorithms M1 and M2 at the very beginning of the interval, and on the whole for the first fragment of the signal there are differences between algorithms M1 and M2, on the one hand, and M3, M5, and M6 on the other hand. It is interesting that, despite the very significant difference in the rhythm patterns of M3, M5, and M6, the dynamics of the amplitude of the rhythm is displayed by them in practically the same way. Apparently, this is a consequence of the fact that the rhythm in this case is very clear.

For the second fragment of the signal (February 13–28), dynamics of the amplitude of rhythm is estimated by all methods almost equally. The only exception is the M4 method, which shows a smoother dynamics; this is quite expected based on the rhythm model used by this method. Recall that the average rhythm for the SOM4 series was estimated in a sliding window with a width of 120 h, which does not permit responding to rapid changes in the rhythm amplitude lasting less than 2–3 days. However, all the main features of the amplitude dynamics have been fully worked out in this case.

When analyzing the $A_{\max-\min}$ graphs for the SLON2M series (not shown in the figures), the results are approximately the same, although there is a speci-

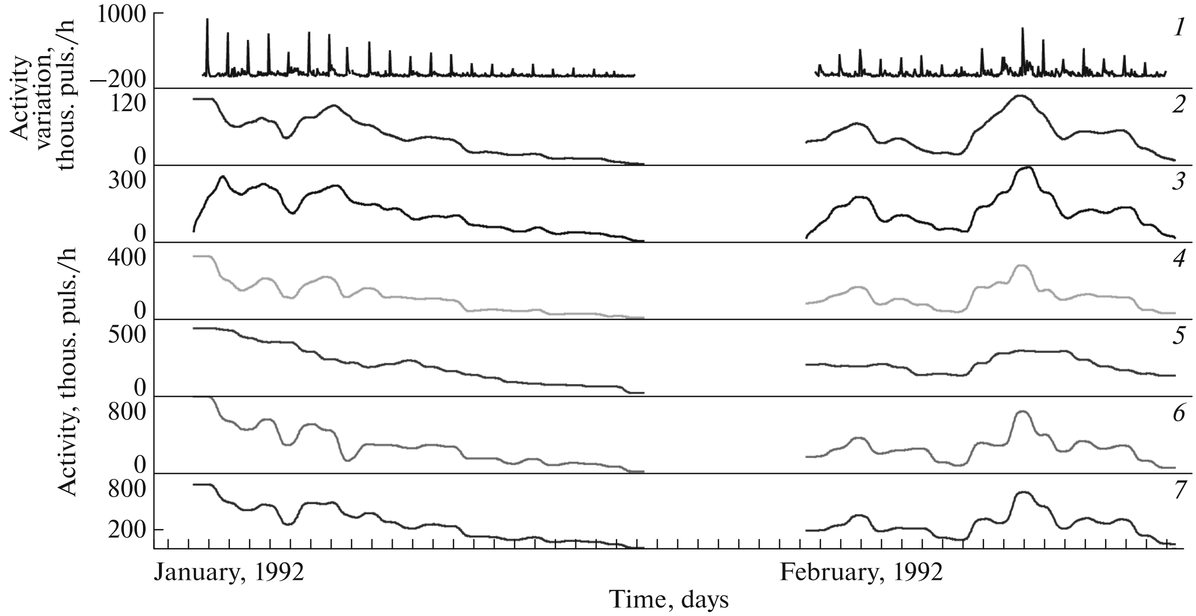

Fig. 3. Amplitude of the diurnal rhythm estimated by the swing method for the SOM4 series (January 12 to March 2, 1992): (*1*–*7*) are the same as in Fig. 1.

ficity associated with a lower signal-to-noise ratio for the diurnal SNR2M dependence. Because of this, the differences between rhythms filtered using different patterns of rhythm are more pronounced.

ESTIMATION OF ACROPHASE AND BATHYPHASE

When analyzing rhythms, characteristics such as acrophase and bathyphase are often used (Komarov, 1989). To calculate the acrophase t_{max}, we are looking for a time t when the monitored index S reaches a maximum value S_{max} for the period. The time t_{min} to reach the minimum value S_{mix} for the period is called the bathyphase. When working with a daily rhythm, it is most convenient to represented results in hours. It should be remembered that the difference between 23 h and 0 h is only 1 h, although this change may look like a phase jump on the graph. For the convenience of visualization in such cases, it is advisable to introduce corrections into the phase, subtracting or adding an integer number of periods. In practice, this means that, instead of a 23-h phase, the graph depicts a phase of −1 h, which is much more convenient for analysis, although it may be somewhat unusual.

Figure 4 shows t_{max} statistics reflecting the time of maximum activity of the SOM4. The specificity of the SOM4 activity (Fig. 4, curve *1*) is that the acrophase is determined very clearly even from the original signal, not refined from noise (curve *2*). At the same time, the jumps of the acrophase on January 20–21, and especially on February 21–22, look random. Estimation of the acrophase in the filtered signals (curves *3*–*6*) shows that not all algorithms are equally useful. For example, refining the rhythm by M1 and M2 algorithms not only does not stabilize t_{max}, but, on the contrary, leads to a "bounce" of the acrophase, within both the first and second fragments of the signal (see Fig. 4, curves *3*, *4*). Obviously, this is due to the fact that the rhythm patterns embedded in M1 and M2 are based on a sinusoid, which is completely inadequate to the real form of the variation. This leads to a phase shift when the amplitude of the rhythm changes. Algorithm M5 (curve *5*), on the contrary, stabilizes the phase of the filtered rhythm. In this case, it may be optimal. It is interesting that an evaluation of the phase by the M6 algorithm (curve *6*) is very stable. This is a consequence of the specific form of variation. For rhythms with a more complex form of periodicity, the results are not always as unambiguous as an analysis of similar graphs for the SLON2M series shows (not shown in this paper).

The batiphase is calculated similarly, but it is analyzed much less frequently. In biological systems, the rhythm quite often has a sharp splash (as in the SOM4), and, much less often, a sharp dip. For example, when analyzing the rhythm of the SOM4, the acrophase will be evaluated steadily (since the maximum is clearly visible), while the bathyphase is not (see Fig. 1b). For the SLON2M rhythm, extracted by the superposition of epochs in the sliding window (see Fig. 1a, curves *5* and *6*), the acrophase estimation is unstable during the period from November 15 to 19.

We noted above that, when evaluating the amplitude characteristics, as can be seen from Fig. 3, the type of model that approximates the rhythm does not play an essential role. When analyzing the phase

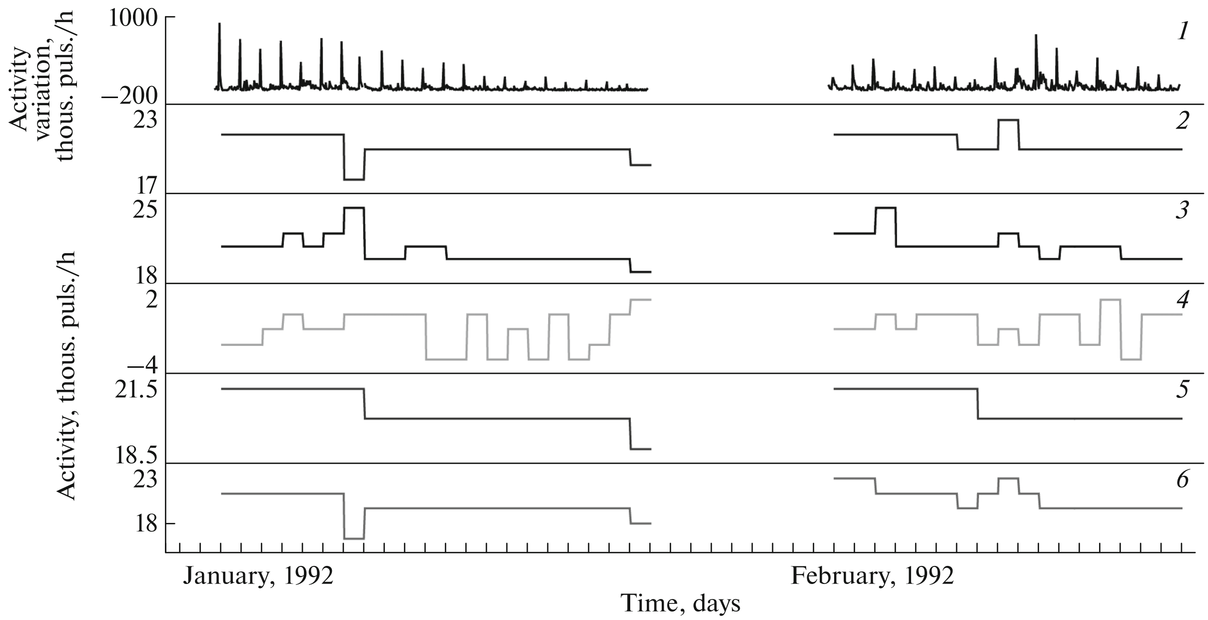

Fig. 4. Acrophase activity for the SOM4 series estimated using different methods of rhythm extraction (January 12 to March 2, 1992): the original series (*1*), from the original series (*2*), from the filtered rhythm (*3*–*6*): models M1 and M2 (*3*, *4*), model M4 (*5*), and model M6 (*6*).

parameters, the correct choice of the rhythm model is much more important. If the model does not correspond to the actual variation, the phase is evaluated very unstably, with large errors.

DISCUSSION

Insufficiency of Approaches Based on Spectral–Temporal Analysis

To analyze the rhythmic structure of the signal, various approaches and methods are used in monitoring. The time-frequency signal analysis (TFSA) and wavelet analysis allow us not only to detect rhythms, but also to track changes in the amplitude and period of the rhythm in time. In many cases, these tools allow us to isolate a whole spectrum of harmonics and rhythms, even for a signal that at first glance looks completely random. However, in practice, it often turns out that such a model of a signal, which represents it in the form of a set of "flickering" rhythms, is very difficult to interpret in meaningful physical terms. Therefore, even when analyzing the results presented in the form of TFSA diagrams, most attention is usually paid not to chaotic "flickering" rhythms, but to more regular periodicities that are rather notable against the background and essentially determines the structure of the process.

However, the rhythms with changing periods are not so often found in nature. In most cases, rhythms tied to regular external influences (seasonal, diurnal, tidal, etc.) are of more interest. Such fluctuations, tied to certain preknown periods, usually have a complex form, determined by the nonlinearity of the system under study and multifactor effects. Such a rhythm will be "smeared" over a whole spectrum of multiple frequencies on the TFSA diagram due to the Gibbs phenomenon. The representation of a rhythm—in fact an integral object (an entity according to Occam)—in the form of a package of independent oscillations is hardly an optimal approach.

Additional problems arise when there are any sharp features in the signal: steps, outliers, activity bursts, and data gaps. On a TFSA or wavelet diagram, such features manifest themselves on many frequencies simultaneously, which greatly complicates the interpretation, especially in the case of nonstationary processes (that are just typical for monitoring).

When processing the data of regular monitoring, the very fact of the existence of a rhythm does not usually cause doubts, and the main task is to track changes in the characteristics of the rhythm in time. Thus, the task of rhythm detection loses its relevance. If, in addition, the period of the rhythm is fixed, the task of tracking the variations of the period becomes irrelevant too. It is clear that methods that try to extract this information from a signal can do this only at the cost of resolution degradation while tracking other signal characteristics (an analogue of the Heisenberg uncertainty principle). Including excessive parameters in a model, we always lose accuracy in estimating other parameters, because the amount of information in the experimental signal is limited.

If the fact of rhythm presence is known and its period is fixed, it is better to use specialized tools designed to track changes in the parameters of the rhythm.

Model of Rhythm as a Tool for Analysis

In connection with the wide application of Fourier analysis, predominantly harmonic oscillations are often understood under a periodicity. More complex periodic processes were investigated in a much smaller number of papers (Anderson, 1976; Rytov, 1976; Deshcherevskii and Lukk, 2002). Within these approaches, such periodicity is represented as a superposition of harmonics with multiple periods. Indeed, a periodic function with reasonable properties can always be approximated with desired accuracy by a set of harmonics of the Fourier expansion, and their parameters (period, amplitude, and phase) uniquely determine the characteristics of the rhythm (Anderson, 1976). However, for a number of reasons, such decomposition is convenient only for theoretical analysis, but it is very impractical in the study of experimental signals. It is enough to mention the difficulties arising when trying to extract significant and insignificant harmonics in the Fourier decomposition of a rhythm.

Criteria recommended by theory are applicable only when rather strong conditions are satisfied (the stationarity of the series, absence of data gaps, mutual independence of the values of the aperiodic (residual) component, etc.), which is not feasible in practice. But even if the signal can be filtered so that these conditions are met, the result of the analysis will be only a multiparameter description of a strictly periodic process. This is very inconvenient for monitoring, the purpose of which is to track changes in the monitored system. For example, the analysis shows that the Fourier decomposition of the rhythm consists of 15 harmonics, and the amplitude and phase of each harmonic varies in time according to its own law. How do we interpret this result in physical terms?

In fact, it is not that easy to offer a universal approach for studying arbitrary rhythms. The rhythms encountered in practice have a great variety and do not fit into the Procrustean bed of standard models. For each quasiperiodic process, its own features are specific and their study is complicated by incomplete information. How does one choose the model that best describes a particular rhythm if the experimental signal is noisy and has a limited length, data gaps, and other defects?

One way to solve this problem is to approximate the experimental rhythm using several different models, for example, those proposed in this paper. Each such model actually offers its own definition of the concept of "rhythm," concretizing and detailing the general understanding of rhythm as something repeating (reproducing) in different cycles with possible changes from cycle to cycle. After that, an informal analysis of the merits and demerits of each model is made from the point of view of the scientific task being solved, and the optimal rhythm model is chosen. In such a situation, it is impossible to propose a formal criterion that would allow choosing the preferred model without the participation of an expert—researcher. Instead, it is necessary to consider the totality of qualities and characteristics of the considered and other models of rhythm and evaluate their usefulness in processing specific signals.

Discussion of the Considered Models of Rhythm

Let us now turn to the discussion of algorithms, the work of which was demonstrated using the SLON2M and SOM4 series. All these algorithms provide effective rhythm extraction even for highly noisy signals and allow tracking changes in the characteristics of the rhythm over time. At the same time, each of these algorithms has its own peculiarities, which make it possible to outline the optimal area of its application.

Algorithm M1, which is based on frequency filtering, can be used for slightly noisy rhythms with a smooth form of variation without sharp features. In such situations, the M1 algorithm provides effective tracking of changes in both the amplitude and phase of the rhythm. Algorithm M2 works in a similar way. It is based on an approximation of rhythm by a sinusoid. In contrast to the M1 algorithm, the degree of adaptability of the filter can be varied in this case in accordance with the properties of the series. The wider the approximation window, the higher the noise suppression and the worse the resolution of the method in the time domain is. Like algorithm M1, algorithm M2 is most effective for a smooth, quasi-harmonic form of variation.

M3 algorithm, which uses average oscillation as the model rhythm **R**, provides a good approximation of the rhythms with a complex shape and makes it possible to accurately estimate A_{regr} amplitude even under conditions of strong noise. However, this method is applicable only if the phase of the rhythm is not changing, which is not always the case. For example, the time of thawing of snow or the beginning of a high water may vary greatly from year to year, which entails a shift in the phase of the seasonal rhythm of respected parameters.

The M4 algorithm also approximates the rhythmic process by an average oscillation; however, the latter it is calculated not over the entire row, but only within the sliding window. This allows tracking relatively slow changes in the amplitude and phase of the rhythm. Reducing the width of the window, one can increase the adaptive properties of the algorithm; however, this also reduces the degree of noise suppression. Therefore, the average rhythm estimated in the sliding window should be smoothened in accordance with the recommendations (Deshcherevskii and Sidorin, 1999). The M4 algorithm is very effective if the phase and shape of the rhythm vary rather smoothly (for example, the annual course of diurnal variation). However, if the phase and oscillation changes occur quickly enough, the width of the window for average rhythm estimation should be minimized, which automatically leads to a deterioration in the noise reduction.

This problem is solved to some extent by the M5 algorithm. In this case, the rhythm calculated in the sliding window according to the M4 algorithm is considered as a reference **R**, which is then inserted into the regression model (1). This allows one to react much more quickly to the changes in the amplitude of rhythm. However, the assumption of a slow (insignificant within the window) change in the shape of the variation is still necessary. If the shape of the rhythm changes noticeably within the window, these changes will be treated as noise.

Algorithm M6 gives the possibility to track changes in both the amplitude and phase of the rhythm, and the shape of rhythm can be arbitrary, not necessarily harmonic. This is the only algorithm of those under consideration which makes it possible to fully implement the technique of phase capture; i.e., the period value is corrected in real time. However, for the correct operation of this algorithm, it is necessary that the shape of the variation is unchanging and the reference rhythm **R** is self-orthogonal for any phase shift within 2π. If this condition is violated, a side maximum can be captured. For example, if reference **R** consists of two identical half-periods, a high correlation between **R** and the actual signal **S** will be observed not only for an exact coincidence of phases **R** and **S**, but also when they are shifted by half of a period.[1] The M6 algorithm identifies the rhythm against the noisy background very poorly. These limitations are quite understandable, since the M6 algorithm has the greatest degree of adaptability. For a highly noisy signal, the introduction of additional degrees of freedom into the model makes it poorly defined and generally incapable of rhythm extraction.

Thus, all the algorithms considered have advantages and limitations. We emphasize that it is impossible to determine the best method in advance in order to use it in research work with monitoring data, since there is no a priori criterion for optimality. In one situation, the criterion for a minimum of the variance of the residual **Z** may be preferable, in another, accurately describing some features in the shape of rhythm, etc. To formalize this choice, it is necessary to know in advance what exactly we are looking for, that is, to proceed from a certain model of rhythm. However, usually such a model is not known in advance in research studies. On the contrary, one of the aims of an experiment is to formulate such a model.

[1] In such a situation, an additional triangular weight window can be used to allocate the main maximum, however, this reduces the noise stability of the algorithm.

Studying the Dynamics of Rhythm Parameters

Rhythms that are present in real signals are always masked by other components of the process (both deterministic and random) and by noise. To monitor changes in the rhythm parameters, it must be filtered (cleaned of noise). To do this, you can use the algorithms described in this paper or other similar algorithms. After this, various parameters of the rhythm are estimated from the refined signal (Deshcherevskii and Sidorin, 2003). Regression methods that are very resistant to missing data are well suited for tracking variations in the amplitude of rhythm. In those cases where the rhythm is approximated without using a regression model, the amplitude can be estimated by the span or variance of the variation.

To control the phase changes, we can monitor the position of extremes of the rhythm and its other characteristic features. However, if the shape of the variation can change, then such estimates are not very stable. For example, for a rhythm similar to the SOM4, the bathyphase value is very poorly defined, even for the filtered signal. In such a situation, the Rayleigh-Schuster hodograph method can be used to monitor the phase. An improved algorithm, free from some of the shortcomings of the classical method, was considered by us in (Deshcherevskii and Sidorin, 2015a, 2015b). This method can be used for any rhythms, including rhythms similar to the SOM4. Unfortunately, in this paper we were unable to illustrate this thesis in more detail because of size limitations.

Another useful characteristic of rhythm is the coefficient of its variability in time. The reverse parameter—rhythm stability—can be introduced as the correlation coefficient between **S** and $\hat{\mathbf{S}}$. With the direct evaluation of $\hat{\mathbf{S}}$ by epoch superposition or frequency filtering, an average rhythm **R**, which is constructed by averaging of the function $\hat{\mathbf{S}}$ over the entire observation period, is substituted into expression (1). Depending on the subject area, other parameters that characterize time dependence of signal can be evaluated for each rhythm.

CONCLUSIONS

In this work, some models of rhythms intended for an investigation of quasiperiodic processes are considered. Each such model is essentially a filter (algorithm), which makes it possible to extract the studied rhythm from the signal in the purest form, eliminating noise and interference. After isolating the refined rhythm, it becomes possible to study in detail characteristics such as amplitude, phase, form of variation, etc. In total, six algorithms (models) are considered. This is a frequency-filtering algorithm with a positively defined filter weight function (which significantly improves the tolerance of the method to data gaps), as well as algorithms that approximate the rhythm in a sliding window with a sinusoidal or some "reference rhythm" (that is evaluated separately). All these algorithms ensure efficient rhythm extraction even for very noisy signals and allow monitoring variation in rhythm parameters even with a significant number of missed observations.

The work of the presented algorithms is demonstrated using the example of two experimental realizations obtained with long-term monitoring of the motor activity of bioindicators at the Gharm geophysical test site. The circadian rhythm of motor and electrical activity of bioindicators was considered. Of course, all algorithms without any modifications can be used for an analysis of any other rhythms with other periods.

It is shown that the choice of the optimal algorithm for rhythm extraction depends both on properties of the rhythm and research objectives. For this reason, it is pointless to ask which of these models is better. In each specific situation, when solving certain scientific problems, the advantages and disadvantages of specific models can be more or less significant. The above results clearly show that even algorithms based on similar rhythm models often lead to markedly different estimates of the dynamics of the rhythm parameters. A comparison of the results that can be obtained with different models and their evaluation from the point of view of practical use in processing specific signals are the main criteria for choosing the best model.

All algorithms are implemented in the WinABD package (Deshcherevskii et al., 2016b–2016d).

REFERENCES

Anderson, T., *The Statistical Analysis of Time Series*, New York: John Wiley and Sons, 1971; Moscow: Mir, 1976.

Deshcherevskii, A.V., Zhuravlev, V.I., and Sidorin, A.Ya., Some filtering algorithms for geophysical temporal series, *Izv., Phys. Solid Earth*, 1996, vol. 32, no. 2, pp. 138–148.

Deshcherevskii, A.V., Zhuravlev, V.I., and Sidorin, A.Ya., Spectral–temporal features of seasonal variations in apparent resistivity, Izv., Phys. Solid Earth, 1997a, vol. 33, no. 3, pp. 217–226.

Deshcherevskii, A.V., Lukk, A.A., and Sidorin, A.Ya., Flicker noise structure in the time realizations of geophysical fields, *Izv., Phys. Solid Earth*, 1997b, vol. 32, no. 7, pp. 515–529.

Deshcherevskii, A.V. and Sidorin, A.Ya., *Nekotorye voprosy metodiki otsenki srednesezonnykh funktsii dlya geofizicheskikh dannykh* (Some Problems of the Method for Estimating the Average Seasonal Functions for Geophysical Data), Moscow: OIFZ RAN, 1999.

Deshcherevskii, A.V. and Lukk, A.A., Identification of regular components in time variations of geophysical parameters by the method of decomposition into nonharmonic components, *Vulkanol. Seismol.*, 2002, no. 5, pp. 65–78.

Deshcherevskii, A.V. and Sidorin, A.Ya., Database of biological monitoring in the Garm polygon, *Geofiz. Protsessy Biosfera*, 2002, vol. 1, no. 2, pp. 3–15.

Deshcherevskii, A.V. and Sidorin, A.Ya., Parametrization of time series of animal activity for geophysical studies, in *Modelirovanie geofizicheskikh protsessov* (Modeling of Geophysical Processes), Moscow: OIFZ RAN, 2003, pp. 137–155.

Descherevsky, A.V. and Sidorin, A.Ya., Seasonal variations in natural processes and atmospheric precipitation, *Ann. Geophys.*, 2004, vol. 47, no. 1, pp. 72–81.

Deshcherevskii, A.V. and Sidorin, A.Ya., Improvement of robustness and stability in estimating Rayleigh–Schuster's hodograph parameters using different procedures of vector normalization, *Seism. Instrum.*, 2015, vol. 51, no. 2, pp. 79–97.

Deshcherevskii, A.V. and Sidorin, A.Ya., Testing Rayleigh–Schuster hodographs using time series models and earthquake flows, *Seism. Instrum.*, 2016, vol. 51, no. 3, pp. 232–252.

Deshcherevskii, A.V., Zhuravlev, V.I., Nikol'skii, A.N., and Sidorin, A.Ya., Program package ABD—a universal tool for analysis of monitoring observation data, *Nauka Tekhnol. Razrab.*, 2016a, vol. 95, no. 4, pp. 35–48.

Deshcherevskii, A.V., Zhuravlev, V.I., Nikol'skii, A.N., and Sidorin, A.Ya., Problems of analysis of time series with gaps and their solution methods by the WinABD software, *Geofiz. Protsessy Biosfera*, 2016b, vol. 15, no. 3, pp. 5–34.

Deshcherevskii, A.V., Zhuravlev, V.I., Nikol'skii, A.N., and Sidorin, A.Ya., Tekhnologii analiza geofizicheskikh ryadov. Ch. 1. Software requirements, *Seism. Instrum.*, 2017a, vol. 53, no. 1, pp. 46–59.

Deshcherevskii, A.V., Zhuravlev, V.I., Nikolsky, A.N., and Sidorin, A.Ya., Technology for analyzing geophysical time series: Part 2. WinABD—A software package for maintaining and analyzing geophysical monitoring data, *Seism. Instrum.*, 2017b, vol. 53, no. 3, pp. 203–223.

Kanasewich, E.R., *Time Sequence Analysis in Geophysics*, Edmonton: Univ. of Alberta Press, 1981.

Härdle, W., *Applied Nonparametric Regression*, Cambridge, Cambridge University Press, 1989; Moscow: Mir, 1993.

Hemming, R.W., *Digital Filters*, Englewood Cliffs: Prentice Hall, 1983; Moscow: Nedra, 1987.

Komarov, F.I., *Khronobiologiya i khronomeditsina* (Chronobiology and Chronomedicine). Moscow: Meditsina, 1989.

Lagutin, M.B., *Naglyadnaya matematicheskaya statistika* (Illustrative Mathematical Statistics), Moscow: BINOM, 2009.

Rytov, S.M., *Vvedenie v statisticheskuyu radiofiziku* (Introduction to Statistical Radiophysics), vol. 1: *Sluchainye protsessy* (Random Processes), Moscow: Nauka, Glavnaya redaktsiya fiziko-matematicheskoi literatury, 1976.

Sidorin, A.Ya., *Garmskii geofizicheskii poligon* (The Garm Geophysical Polygon), Moscow: IFZ AN SSSR, 1990.

Terebizh, V.Yu., *Analiz vremennykh ryadov v astrofizike* (Analysis of Time Series in Astrophysics), Moscow: Nauka, 1992.

Translated by I. Ptashnik