Analysis of Long-Term Observations of the Groundwater Level in an Aseismic Region

A. A. Lyubushin[a,]*, O. S. Kazantseva[a,]**, and A. B. Manukin[a,]***

[a]*Schmidt Institute of Physics of the Earth, Russian Academy of Sciences, Moscow, 123242 Russia*
*e-mail: lyubushin@yandex.ru
**e-mail: o.kazantseva@yandex.ru
***e-mail: amanukin@yandex.ru

Abstract—The results of the analysis of continuous, precise time series of observations of the atmospheric pressure and groundwater level in a well drilled to a depth of 400 m and located in the territory of Moscow are presented. These observations are remarkable in terms of their duration which is more than 22 years (from February 2, 1993 to April 4, 2015) and by the sampling interval which is 10 min. Such a long duration of the observations allows exploring the question of how stationary the properties of hydrogeological time series are in a seismically quiet region, which is important from the methodological standpoint for the interpretation of similar observations conducted in seismically active regions for the purposes of earthquake prediction. As a result of applying the factor and cluster analysis to the sequence of multivariate vectors of the statistical properties of the time series of groundwater level observations in successive time windows with a length of 10 days, after the adaptive compensation for the atmospheric pressure effects, five different statistically significant states of the time series between which the transitions take place are distinguished. An attempt of geophysical interpretation of the revealed states is made. The spectral analysis of the sequence of times of the transitions between the clusters identified two significant periods 46 and 275 days.

INTRODUCTION

The changes in the groundwater level contain information about the volumetric deformations in the Earth's crust. If a well is drilled into a sufficiently thick fluid-saturated horizon, the water level variations will characterize the volumetric deformations that determine regional processes in the crust. By analyzing the water level fluctuations in wells, we can study lunar–solar tidal processes in the solid Earth (Bredehoft, 1967; Rojstaczer and Agnew, 1989; Bagmet et al., 1989; Lyubushin et al., 1997; Vinogradov et al., 2011) and obtain data on the variability of the crustal properties in a given region from the groundwater's response to the changes in the atmospheric pressure (Lyubushin and Malugin, 1993; Lyubushin and Lezhnev, 1995; Kopylova et al., 2009). Long-term observations in a seismically active region can yield information about the precursors of earthquakes (Roeloffs, 1988; Roeloffs et al., 1989; Igarashi and Wakita, 1991; Kopylova et al., 2000; Kopylova, 2001; 2006; Boldina and Kopylova, 2017), whereas the information acquired in the aseismic platform regions is necessary for detecting the growth of anomalous processes in seismically active regions by comparative analysis (Kissin and Gumen, 1994).

Parallel measurements at several wells drilled into different fluid-saturated horizons can provide information about the fluid dynamic processes in this region (Lyubushin et al., 1999; Besedina et al., 2015). Furthermore, the analysis of the measurement results will help identify the hydrodynamic regimes caused by industrial impacts: water withdrawal from a given or other water saturated horizon, fluid injection, the change in the soil load due to building operations, etc. The value of the information about the geophysical and hydrodynamic processes increases in the case of long-term (multiyear) continuous and highly sensitive observations with high-frequency sampling of the data. Under these conditions, it is possible to reveal fine effects in a broad range of periods.

The use of the instrumentation composing the system for the parallel measurements of the water level in a well and atmospheric pressure, which has a high resolution, wide dynamic range, sufficient long-term stability, reliability, and autonomy, made it possible to solve the problem of organizing and conducting long-term continuous observations.
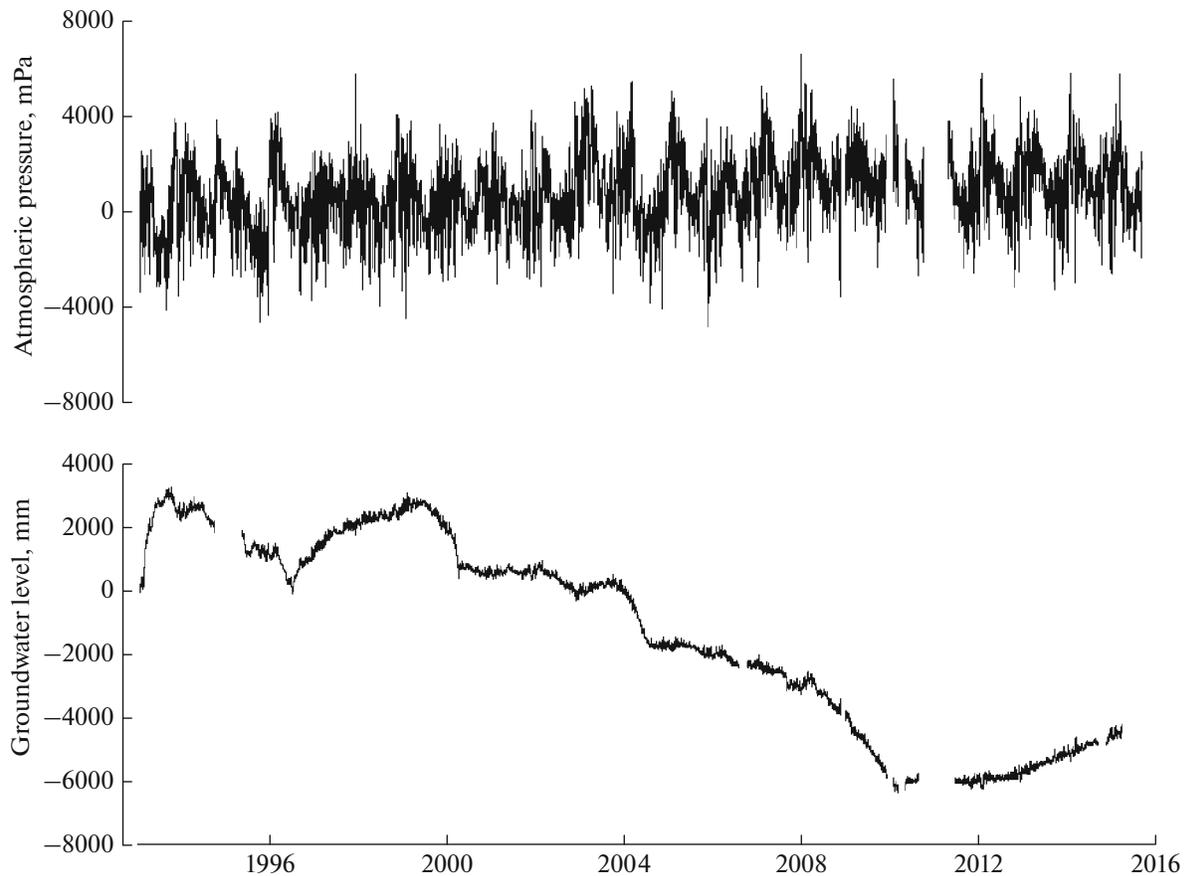
Fig. 1. Graphs of initial data: variations in atmospheric pressure and groundwater level, February 2, 1993 to April 4, 2015, time step 10 min.

INITIAL DATA

Continuous measurements of groundwater level fluctuations were conducted from February 2, 1993 to April 4, 2015 in a well located in the northwest of Moscow in the territory of the Central Institute for Traumatology and Orthopedics (CITO) at 55.8202° N, 37.5292° E. Drilled to a depth of 404 m, the well penetrates the Upper Famennian aquifer (*Gidrogeologiya...*, 1966) composed of fractured dolomites and limestones. The total thickness of water saturated rocks is above 200 m.

The well was drilled in 1969–1970 and intended for the balneotherapeutic treatment at the CITO; however, since the beginning of the 1980s, the well has not been used for the purposes for which it was intended to be used and was suspended. The chemical analysis of the water was conducted and water salinity was estimated at 5 g/L. A special observation point was equipped in a separate space with an area of 10 m² close to the wellhead.

Simultaneously with the groundwater level's fluctuations, variations in the atmospheric pressure were measured. The groundwater level and atmospheric pressure sensors are of the manometer type with a capacitance-to-frequency converter of the displacement of the sensitive element (a membrane) into the output signal. The sensitivity of the sensors is 0.1 mm for the water gauge for the groundwater level and 10 μbar for the atmospheric pressure. The groundwater level and atmospheric pressure sensors are installed in the well on the cables. The water column height above the level sensor is 2.5–3 m. The atmospheric pressure sensor is lowered to a depth of ~10 m. Data are sampled with an interval of 10 min and recorded in the digital form with the use of a special interface unit in the solid-state memory units. Figure 1 shows the graphs of the time series of the initial data. The total length of the records is 1 165 862 10-min readings. The records have gaps associated with failures of the recording system.

TRANSFER FUNCTIONS AND POWER SPECTRA

As shown by the estimates of the coherence spectra, the changes in the atmospheric pressure are the main external meteorological factor affecting the groundwater level variations. The long (22-year) observation time series allow statistically significantly
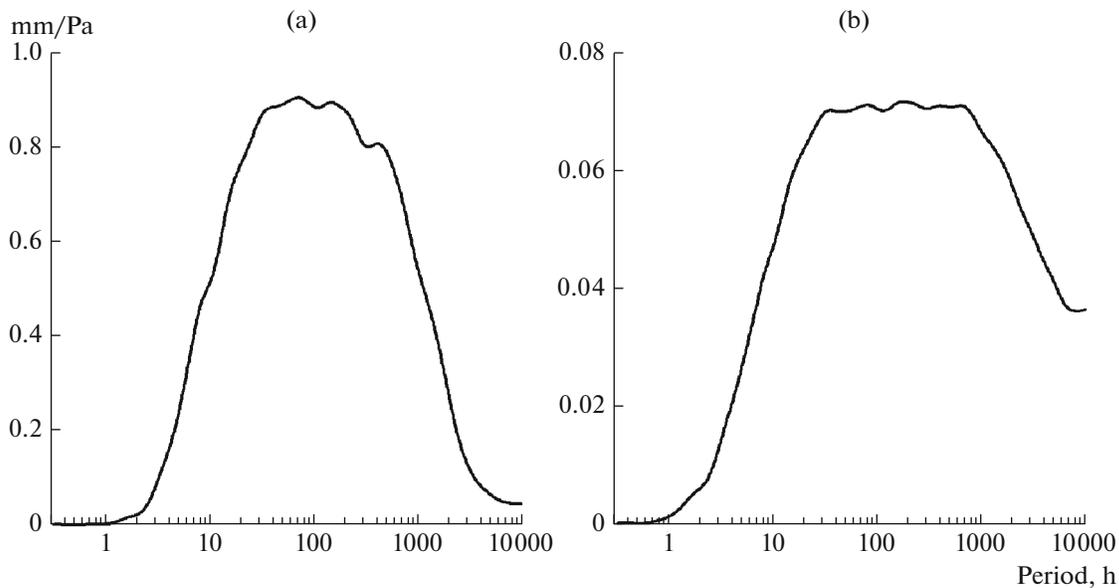
Fig. 2. Graphs of estimates of (a) quadratic coherence spectrum and (b) amplitude frequency transfer function.

estimating the amplitude frequency transfer function from the atmospheric pressure to the groundwater level fluctuations for a very wide range of periods, from 20 min which is the Nyquist frequency for the considered time series to one year. The very possibility of obtaining this estimate is unique and achieved due to the combination of the favorable factors listed in the Introduction.

The estimates of the transfer functions were constructed by the classical nonparametric method based on the Fourier transform of the input (atmospheric pressure) and output signals (water level variations), calculation of their cross-spectral function and the power spectrum of the "input" (by averaging the periodograms and the cross-periodogram) and the cross-spectrum to power spectrum ratio (Brillinger, 1975). The wide period range posed a technical challenge when obtaining an estimate at once over the entire sample. Therefore, the estimates were built by stitching three different estimates. The estimate for the periods from 20 min to 10 h was constructed by averaging (for each frequency of the cross-periodograms) the initial 10-min time series obtained in the moving time windows with a length of 2048 points (with the preliminary removal of the local linear trends in each window and the subsequent conversion to increments) and the finishing frequency averaging by the moving frequency window with a radius of 20 frequency values (9.7×10^{-4} min^{-1}). The estimate for the periods from 10 to 100 h was constructed by similar procedures but for the time series after the conversion to 1-h discretization (by averaging and six-fold decimation of the initial time series); the radius of the averaging frequency window became in this case equal to 9.7×10^{-3} h^{-1}. Finally, for estimating in the period range from 100 to 10000 h,

the data were preliminarily reduced (after averaging and 144-fold decimation of the initial time series) to 1-day discretization; the estimate was constructed by the same procedure; however, the length of the moving time window was 1024 points and the radius of the frequency averaging was 10 frequency values (9.7×10^{-3} day^{-1}). In all cases, the time windows were shifted by 1/8 of the window's length.

The missing values were processed in the following way. Initially, all the gaps were filled in with the artificial data synthesized in accordance with the behavior of the actual data on the left and on the right of the gap interval. The synthetic data were constructed on the time fragments of the same length as the gaps. Subsequently, since estimation was conducted by averaging the periodograms from the moving time windows, only the windows where the fraction of the synthetic (padded) values was at most 2% of the length of the time window were accepted for processing.

Figure 2 shows the graphs of the estimates of the square modulus of the coherence spectrum (frequency-dependent squared coefficient of correlation) and the amplitude frequency transfer function.

It should be borne in mind that the variance of the nonparametric estimate of the transfer function (asymptotic formula for a sufficiently large number of data points) has the form (Brillinger, 1975) $c(1 - \gamma_{xy}^2(\omega))(S_{yy}(\omega)/S_{xx}(\omega))$. Here, ω is frequency; $S_{yy}(\omega)$ is the power spectrum of the output signal (water level); $S_{xx}(\omega)$ is the power spectrum of the input signal (atmospheric pressure); $\gamma_{xy}^2(\omega)$ is their square modulus of the coherence spectrum; and c is the constant that depends on the method of averaging the

periodograms, in particular, $c \sim (k(2m + 1))^{-1}$ where m is the radius of frequency averaging of the periodograms in the units of the number of frequency values (in our case, 20 or 10) and k is the number of the independent time intervals over which the averaging was carried out. The variance of the estimate of the squared modulus of the coherence spectrum is $2c(1 - \gamma_{xy}^2(\omega))^2 \gamma_{xy}^2(\omega)$. Therefore, for the periods where the estimate $\gamma_{xy}^2(\omega)$ is small, the confidence interval for the estimate of the transfer function should be expected to be wide. However, the use of the very deep averaging of the periodograms, both over time intervals and over frequency (large k and m), makes the obtained estimate fairly robust to noise.

Atmospheric pressure can be used as a probing signal. Based on the changes in the frequency function of the groundwater level barometric response, it is possible to detect the time intervals and frequency bands of anomalous behavior (Lyubushin and Malugin, 1993; Lyubushin and Lezhnev, 1995; Kopylova et al., 2000). Below, the estimation of the variability of the frequency response function was conducted after reduction to 1-h discretization by calculating the average values over six neighboring 10-min values. We used a moving time window with a length of 672 h (28 days or a lunar month is the natural time scale in the analysis of low-frequency background geophysical processes in the Earth's crust). For each position of the window, the frequency transfer function of the variations in atmospheric pressure was estimated with the use of the parametric regression model

$$x(t) = \sum_{j=0}^{p} b_j u(t - j) + d + e(t). \qquad (1)$$

Here, $x(t)$ are the increments of the groundwater level time series; $u(t)$ are the increments of atmospheric pressure; p is the regression order; and $(b_j, j = 0, ..., p, d)$ are the model parameters, which were determined in each time window from the condition of the minimum sum of the squared residuals of identification $\sum e^2(t) \to \min$. The mutual shift of the windows was selected to be 14 days. Besides, before identifying the parameter of model (1), the samples were winsorized (Huber, 1981) for increasing the stability of the estimate to the presence of outliers. Winsorization means the iterative clipping of the extreme values that fall beyond the level of ± 3 sample standard deviations with the subsequent recalculation of the mean value and the variance. After determining the parameters of model (1), the frequency function of the groundwater level response to the variations in the atmospheric pressure depending on frequency ω was determined by the formula

$$H(\omega) = \sum_{j=0}^{p} b_j \exp(-i\omega j). \qquad (2)$$

In our calculations, the order of regression in model (1) was specified by $p = 5$ as the minimal value after the further increase of which the result of estimating the response functions changes insignificantly.

Figure 3 shows the results of estimating the amplitude transfer function of the response $A(\omega) = |H(\omega)|$. In Figure 3a there is the graph of the maximal value $\max_{\omega} A(\omega)$ for each time window, and Fig. 3b is the frequency–time diagram of the amplitude frequency transfer function. The shaded rectangles correspond to the time intervals in which the fraction of the synthetic (padded) values in the time series is above 2% of the window length.

From Fig. 3 it can be seen that the response function is stationary overall but contains short anomalous time intervals during which the maximum of the frequency amplitude function of the response undergoes sharp changes, both increasing and decreasing. Thus, latent tectonic life takes place even in a stable platform area. The existence of these short-lived anomalies was noted in (Lyubushin et al., 1999) where they were called slow events in the aseismic region.

For the subsequent analysis, we need to obtain the time series of the groundwater level variations with excluded influence of the atmospheric pressure. This was done with the use of the compensating adaptive frequency filter (Lyubushin, 2007). In the moving time windows with a length of 28 days (or 4032 time points with 10-min sampling interval) which were shifted with the minimal step by 1 point, we calculated the power spectrum $S_{uu}(\omega)$ of the pressure and the complex cross-spectrum $S_{xu}(\omega)$ between the groundwater level and pressure. These estimates are obtained by averaging the periodograms and cross-periodograms by the frequency window with a length of 1/32 of the window. Next, in each window we calculated the frequency transfer function $H_{xu} = S_{xu}(\omega)/S_{uu}(\omega)$. Here, prior to averaging the periodograms, the tidal frequency bands [1/11, 1/13] and [1/23, 1/27] h^{-1} were suppressed, whereas the estimates within these frequency bands were obtained by the interpolation of the estimates from the neighboring frequency values. The results of the compensation $\tilde{E}_x(\omega)$ in the frequency domain within each time window were calculated by the formula $\tilde{E}_x(\omega) = \tilde{X}(\omega) - H_{xu}(\omega)\tilde{U}(\omega)$, where $(\tilde{X}(\omega), \tilde{U}(\omega))$ is the discrete Fourier transform of the groundwater level and atmospheric pressure within a current window. The result of the compensation $e_x(t)$ in the time domain within each time window is determined as a result of the inverse Fourier transform of $\tilde{E}_x(\omega)$.
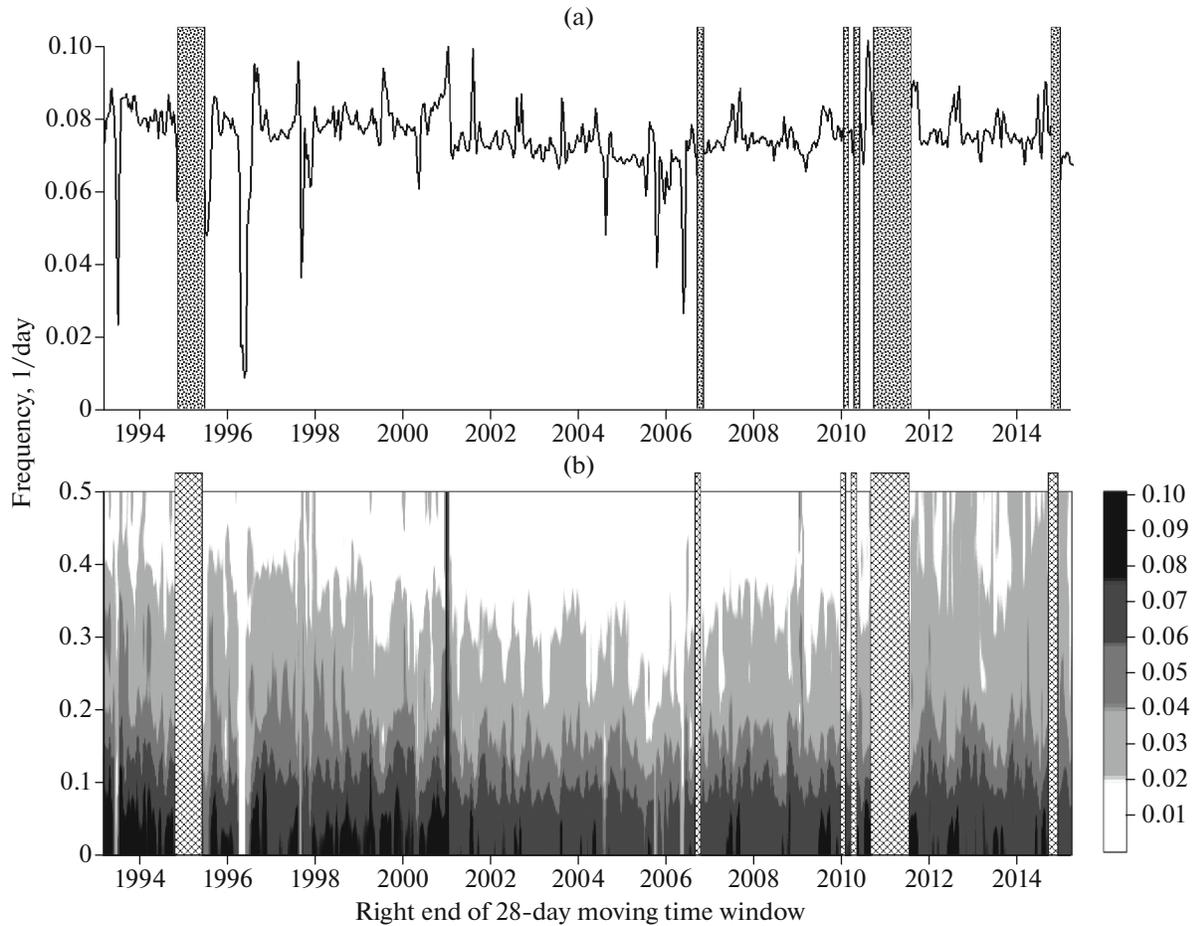
Fig. 3. (a) Graph of maximum value $\max\limits_{\omega} A(\omega)$ for each time window; (b) time-frequency diagram of amplitude frequency transfer function. Shaded rectangles correspond to time intervals in which content of artificial (padded) values of time series exceeds 2% of window length. Minimum frequency in diagram (b) corresponds to maximal period equal to window's length, i.e., 28 days.

Stitching the intrawindow results of the compensation into a single signal is the final procedure for obtaining a compensated signal. The first window contributes to this single signal the content of the part of the time window that corresponds to the first half of the window, whereas the last window contributes the content of its part corresponding to the second half. The other time windows, intermediate between the first and the last ones, contribute only one data point corresponding to the center of the window.

Figure 4 shows the graphs of the initial and compensated groundwater level for the 2-month time fragment for the beginning of 1999. It can be seen how much stronger the tidal variations of the groundwater level are pronounced in the compensated signal compared to the initial data where they sink in the variations caused by the atmospheric pressure.

Figure 5 shows the power spectrum estimates of groundwater level variations for the initial time series and after the compensation for the effects of atmospheric pressure. In Fig. 5b it can be seen that the 12-

and 24-h tidal harmonic groups are identified highly reliably and the spectral peaks at the periods corresponding to the highest overtones of diurnal variations (6, 4 h, etc.) which are present in the power spectrum of the initial signal have vanished after the compensation. This indicates that these peaks are associated with the effects of the nonharmonic form of the diurnal pressure variation. Moreover, we note the straightening of the log–log graph of the power spectrum after compensation for the pressure effect: the hump in the power spectrum for the periods from 10 to 3000 h completely disappeared.

STATISTICAL PROPERTIES OF THE COMPENSATED GROUNDWATER LEVEL TIME SERIES

The plan of the further analysis is to estimate the set of the statistical properties describing the behavior of the groundwater level time series after the compensation for the atmospheric pressure effects in successive
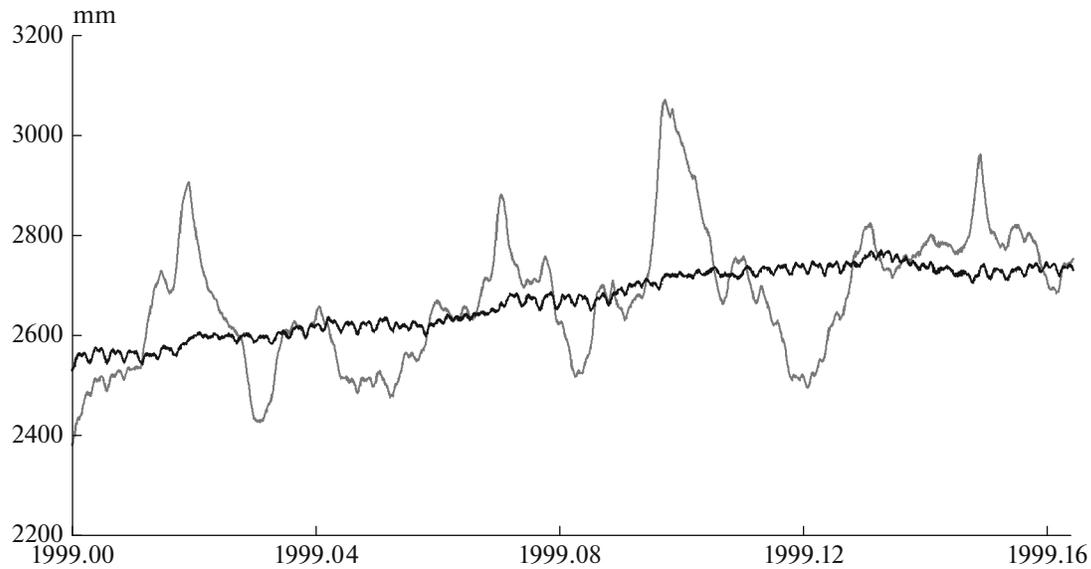
Fig. 4. Comparison of initial groundwater level data (gray line) and result of adaptive compensation of atmospheric pressure in moving window with length of 28 days (4032 time points with time step of 10 min, black line) for time fragment of first two months of 1999.
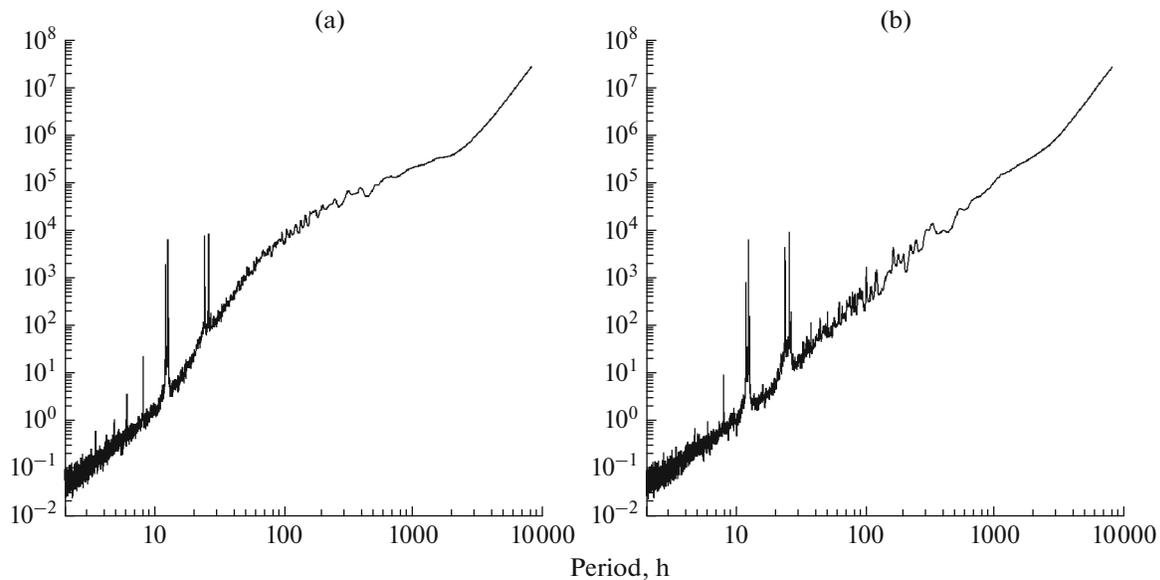

Fig. 5. Comparison of power spectra of (a) initial groundwater level data and (b) after adaptive compensation of effect of atmospheric pressure in 28-day moving window.

time fragments and to use their values for identifying different states of the set of long-term observations of the dynamics of a water-bearing horizon. The length of the time fragment was selected to be $N = 1440$ points, which corresponds to 10 days with a time step of 10 min. Next, in this section we briefly describe the statistics we used in the study. There are ten of these statistics overall. All the estimates were conducted for the time series $x(t)$ of the groundwater level increments after compensation for the influence of the atmospheric pressure.

Minimal normalized entropy of wavelet coefficients En. Let $x(t)$ be some finite sample of a certain random signal and let $t = 1,...,N$ be the index that enumerates the successive data points (the discrete time). We determine the normalized entropy of the finite sample by the following formula:

$$En = -\sum_{k=1}^{N} p_k \log(p_k)/\log(N),$$

$$p_k = c_k^2 \Big/ \sum_{j=1}^{N} c_j^2, \quad 0 \le En \le 1. \tag{3}$$

Here, c_k, $k = 1, N$ are the coefficients of the orthogonal wavelet decomposition with a certain basis. Below we used 17 orthogonal Daubechies wavelets: ten ordinary bases with the minimal support width with 1 to 10 vanishing moments and seven so-called Daubechies symlets (Mallat, 2005) with 4 to 10 vanishing moments. For each basis, the normalized entropy of the distribution of the squared coefficients (3) was calculated and the basis that provides the minimum of quantity (3) was determined. We note that due to the orthogonality of the wavelet transform, the sum of the squared coefficients is equal to the variance (energy) of the signal $x(t)$. Thus, quantity (3) calculates the entropy of the energy distribution of the oscillations on the different frequency and time scales. The *En* statistics was used in (Lyubushin, 2012; 2014; 2018) in the study of the prognostic properties of seismic noise on the Islands of Japan.

Donoho–Johnstone Index γ. After the wavelet basis is determined for a given signal from the minimum entropy condition, we can find the set of the wavelet coefficients which are smallest by the absolute value. In wavelet filtering, these coefficients can be zeroed before the inverse wavelet transform in order to reduce the noise (Donoho and Johnstone, 1995; Mallat, 1999). We assume that the noise is mainly concentrated in the variations at the first detail level . Due to the orthogonality of the wavelet transform, the variance of the wavelet coefficients is equal to the variance of the initial signal. Thus, we estimate the standard deviation of the noise as the standard deviation of the wavelet coefficients at the first level of detail. This estimate should be stable, i.e., insensitive to the outliers in the values of the wavelet coefficients at the first level. For this purpose we can use the robust median estimate of standard deviation for a normal random quantity

$$\sigma = \text{med}\{|c_k^{(1)}|, \quad k = 1, ..., N/2\}/0.6745, \quad (4)$$

where $c_k^{(1)}$ are the wavelet coefficients at the first level of detail and $N/2$ is the number of these coefficients. The estimate of the standard deviation σ from formula (2) determines the quantity $\sigma\sqrt{2\ln N}$ as the natural threshold for separating the noise wavelet coefficients. The quantity $\sigma\sqrt{2\ln N}$ is known in wavelet analysis as the Donoho–Johnstone threshold and the expression for this quantity is based on the formula for the asymptotic probability of the maximal deviations of Gaussian white noise (Mallat, 1999). As a result, it is possible to determine the dimensionless characteristic of the signal $\gamma, 0 < \gamma < 1$ as the ratio of the number of the most informative wavelet coefficients for which inequality $|c_k| > \sigma\sqrt{2\ln N}$ is satisfied to the total number N of all the wavelet coefficients. Formally, the larger the index γ the more informative (the less noisy) the signal.

Autoregression model. Below, we widely use the autoregression (AR) model (Box and Jenkins, 1970; Kashyap and Rao, 1976) for the time series $x(t)$. The general form of this AR model is

$$x(t) + \sum_{k=1}^{p} a_k^{(p)} x(t - k) = e^{(p)}(t) + d^{(p)}. \quad (5)$$

Here, integer $p \geq 1$ denotes the order of the AR and vector $c = (a_1^{(p)}, ..., a_p^{(p)}, d^{(p)})^T$ is the vector of the unknown parameters. The upper index (p) in formula (5) indicates that the AR model of order p is used. Here, $a_k^{(p)}$ are autoregressive coefficients, $d^{(p)}$ is the static shift parameter, and $e^{(p)}(t)$ is the residual signal with the zero mean and variance σ_p^2. Model (5) can be written out in a compact form:

$$x(t) = c^T Y(t) + e^{(p)}(t),$$
$$Y(t) = (-x(t-1), ..., -x(t-p), 1)^T. \quad (6)$$

Let there be a finite sample $\{x(t), t = 1, ..., N\}$. Then estimating parameter vector c from the condition of the minimum sum of the squared residuals $\sum_{t=p+1}^{N} (e^{(p)}(t))^2 \to \min_c$ is reduced to solving the system of normal equations with a symmetric positive definite matrix A:

$$Ac = R, \quad A = \sum_{t=p+1}^{N} Y(t)Y^T(t),$$
$$R = \sum_{t=p+1}^{N} x(t)Y(t). \quad (7)$$

The total vector of the parameters of model (5) is $\theta^{(p)} = (c^T, \sigma_p)^T$.

Hereinafter, as the characteristics of the fragments of the time series, together with the other parameters, we will use the values of coefficient $a_1^{(1)}$ of the first-order AR model and the logarithm of the variance of the residual in this model, $\log \sigma_1^2$.

Index of linear predictability c_{Pred}. The index of linear predictability was introduced in (Lyubushin, 2010), see also (Lyubushin, 2012). Let us consider quantity $c_{\text{Pred}} = V_0/V_{AR} - 1$. Here, V_0 is the variance of the error $\varepsilon_0(t + 1)$ of the one-step-ahead trivial prediction $\hat{x}_0(t + 1)$ for the signal $x(t)$, which is equal to the mean over the previous short time window with a length of n data points: $\hat{x}_0(t + 1) = \sum_{s=t-n+1}^{t} x(t)/n$. Thus, $\varepsilon_0(t + 1) = x(t + 1) - \hat{x}_0(t + 1)$, and $V_0 = \sum_{t=n+1}^{N} \varepsilon_0^2(t)/(N - n)$, where $N > n$ is the number of data points in the successive long time fragments. Quantity V_{AR} is calculated

by the similar formula $V_{AR} = \sum_{t=n+1}^{N} \varepsilon_{AR}^2(t)/(N-n)$ in which $\varepsilon_{AR}(t+1) = x(t+1) - \hat{x}_{AR}(t+1)$ is the error of the linear one-step-ahead prediction $\hat{x}_{AR}(t+1)$ with the use of the second-order AR model whose coefficients are also estimated from the previous short time window with a length of n data points.

The choice of the second order of AR is due to the fact that this is the minimal order for the AR model that describes oscillatory motion and allows the frequency position of the maximum of the spectral density of the AR model between the Nyquist frequency and zero. The AR-prediction utilizes the property of the neighboring values of being correlated with each other, and if the correlation takes place, then $V_{AR} < V_0$ and $c_{\mathrm{Pred}} > 0$. With the length of the long window $N = 1440$, the length of the short window was specified to be $n = 144$.

Autoregressive measure of nonstationarity of the signal R^2. Let $x(t)$ be the signal under study and n be the half-length of the moving time window which will be hereinafter referred to as a short window. Let τ be the center of the double moving time window which, thus, contains the time points t satisfying the condition $\tau - n \le t \le \tau + n$. For the left and right parts of the short window, we construct the scalar autoregressive model (5) of order $p = 2$ of signal $x(t)$. By estimating the model independently from the samples falling in the left and right parts of the moving time window, we obtain two vectors of parameters $\theta_1^{(p)}$ and $\theta_2^{(p)}$, respectively. We denote the difference between the vectors of estimates on the right and left parts of the moving window by $\Delta\theta = \theta_2^{(p)} - \theta_1^{(p)}$.

If the behavior of the studied signal strongly differs on the left and right parts of the studied signal, the difference $\Delta\theta$ will increase. As a metric matrix for weighting the vector $\Delta\theta$, it is reasonable to use the Fisher matrix as it determines the rate of change of the log-likelihood function in the vicinity of the point of the maximum with respect to the parameters:

$$B = -\frac{\partial^2 \ln(\Phi)}{\partial\theta\partial\theta}, \quad \ln(\Phi) = -(n-p)\ln(\sigma_p)$$
$$-\frac{1}{2\sigma_p^2}\sum_t (x(t) - c^T Y(t))^2 \qquad (8)$$

are the matrices of the parametric second derivatives of the conditional log-likelihood function of the autoregressive model. Let $B^{(1)}$ and $B^{(2)}$ be the matrices calculated from the left and right parts of the moving window, respectively. Then, the measure of nonstationarity in the behavior of the process $x(t)$ in the symmetric vicinity of point τ will be the quantity

$$r^2(\tau) = (\Delta\theta^T B^{(1)} \Delta\theta + \Delta\theta^T B^{(2)} \Delta\theta)/(2(n-p)). \qquad (9)$$

In formula (9), the half-sum of the lengths of the parameter difference vector $\Delta\theta$, which are measured by the metric matrices $B^{(1)}$ and $B^{(2)}$, is divided by $(n-p)$, i.e. the number of the data points in the left and right parts of the moving window minus the number of AR parameters. This metrics provides a natural dimensionless measure of nonstationarity in the behavior of the studied signal. After fairly simple transformations, we obtain the following expression:

$$\Delta\theta^T B \Delta\theta = \frac{2(\Delta\sigma_p)^2}{\sigma_p^2} + \frac{\Delta c^T (\sum_t Y(t) Y^T(t)) \Delta c}{\sigma_p^2(n-p)}$$
$$+ \frac{4\Delta c^T \Delta\sigma_p \sum_t e^{(p)}(t) Y(t)}{\sigma_p^3(n-p)}, \qquad (10)$$

which is useful for calculating the nonstationarity measure (9). The nonstationarity measure (9)−(10) was introduced in (Lyubushin et al., 1999), see also (Lyubushin, 2007).

Using formulas (9) and (10), we can define another, more stable measure of the nonstationary behavior of the studied signal within the long time interval consisting of N successive time points. For this purpose, we consider a short window with a radius of n points, $2n + 1 < N$, and calculate the measure of nonstationary behavior $r^2(\tau)$ for all possible positions of the central point τ within the long window at which the short window lies entirely inside the long one. It can be easily calculated that the number τ of these allowable positions of the central point is $N - 2n$. We determine the integral nonstationarity measure R^2 for the long window as the median of the $r^2(\tau)$ values for all the allowable positions of central point τ of the short window inside the long one. In our calculations we used the window lengths $N = 1440$ and $n = 144$. Hereinafter, we will use the logarithm of the nonstationarity measure $\log(R^2)$.

Multifractal parameters $\Delta\alpha$, α^* and $\alpha_{\min}$. We consider a certain random oscillation $x(t)$ on the time interval $[t - \delta/2, t + \delta/2]$ with length δ centered at the time point t. We analyze the peak-to-peak amplitude $\mu(t, \delta)$ of the random oscillation on this interval, i.e., the difference between its maximal and minimal values:

$$\mu(t, \delta) = \max_{t-\delta/2 \le s \le t+\delta/2} x(s) - \min_{t-\delta/2 \le s \le t+\delta/2} x(s). \qquad (11)$$

If we let $\delta \to 0$, then $\mu(t, \delta)$ will also tend to zero; however, here the rate of this decay is important. If the decay rate is determined by $\delta^{h(t)}$: $\mu(t, \delta) \sim \delta^{h(t)}$ or if the limit $h(t) = \lim_{\delta \to 0} \frac{\log(\mu(t, \delta))}{\log(\delta)}$ exists, then quantity $h(t)$ is referred to as the Hölder−Lipschitz exponent. If $h(t)$

does not depend on time t, $h(t) = \text{const} = H$, then the random oscillation $x(t)$ is referred to as monofractal and quantity H as the Hurst exponent. In the case when the Hölder–Lipschitz exponents $h(t)$ differ for different time instants t, the random oscillation is referred to as a multifractal, and for this multifractal, it is possible to define the notion of the singularity spectrum $F(\alpha)$ (Feder, 1991). With this end in view, we select a set $C(\alpha)$ of the time instants that have the same value α of the Hölder–Lipschitz exponent: $h(t) = \alpha$. Sets $C(\alpha)$ are not empty not for all α values; i.e., there are certain minimal $\alpha_{\min}$ and maximal $\alpha_{\max}$ values such that the sets $C(\alpha)$ contain some elements only for $\alpha_{\min} < \alpha < \alpha_{\max}$. The multifractal singularity spectrum $F(\alpha)$ is a fractal dimension of the set of points $C(\alpha)$. Parameter $\Delta\alpha = \alpha_{\max} - \alpha_{\min}$ which is referred to as the singularity spectrum support width appears to be an important multifractal characteristic. Argument α^* that provides the maximum of the singularity spectrum $F(\alpha^*) = \max\limits_{\alpha_{\max} \le \alpha \le \alpha_{\min}} F(\alpha)$ and is referred to as the generalized Hurst exponent is also fairly interesting. The maximum of the singularity spectrum cannot exceed 1, i.e., the dimension of the embedding set or the time axis, $0 < F(\alpha^*) \le 1$; typically, $F(\alpha^*) = 1$. We note that in the case of a monofractal signal, $\Delta\alpha = 0$ and $\alpha^* = H$.

Below, for estimating the multifractal characteristics of the signals, we used the method that is based on the analysis of the fluctuation after the elimination of scale-dependent trends (Kantelhardt et al., 2002). In (Lyubushin, 2010; 2012; 2013; 2014; 2018), multifractal parameters $\Delta\alpha$ and α^* were used in seismic hazard assessment from the properties of seismic noise in Japan (see also (Lyubushin, 2009; 2012; 2014)).

Kurtosis κ is defined by the formula $\kappa = M(x^4)/(M(x^2))^2$ (Cramer, 1999). It characterizes the sharpness of the graph of the probability density distribution of the zero-mean random quantity x and gives the measure of the probability density deviation from the normal law for which $\kappa = 3$. Here, operation $M(...)$ denotes the calculation of the mathematical expectation, in the considered case simply the sample mean of a random quantity. The kurtosis is typically understood as the quantity κ that was introduced above minus three in order that the kurtosis for the normal distribution is zero. However, hereinafter, we will consider the logarithmic kurtosis $\log(\kappa)$; therefore, the value 3 is not subtracted in order that κ is guaranteed to be positive.

Thus, for each time window there are ten parameters characterizing the statistical properties of the time series within this window: the minimal normalized entropy of wavelet coefficients En, the Donoho–Jonstone index γ, coefficient $a_1^{(1)}$ and log variance $\log \sigma_1^2$ in the AR model of the first order, linear predictability index c_{Pred}, and the log nonstationarity measure $\log(R^2)$, which are based on the use the second-order AR model, multifractal parameters $\Delta\alpha$, α^*, $\alpha_{\min}$, and log kurtosis $\log(\kappa)$.

We denote the ten-dimensional vector of parameters characterizing the statistical properties of the time series within the successive time fragments with a length of 10 days (1440 data points with a time step of 10 min) by

$$\zeta = (En, \gamma, a_1^{(1)}, \log \sigma_1^2, c_{\text{Pred}}, \\ \log(R^2), \Delta\alpha, \alpha^*, \alpha_{\min}, \log(\kappa)). \tag{12}$$

Figure 6 shows the graphs of the variations in the components of the ten-dimensional vector ζ of the properties of the time series as a function of the position of the right end of the successive time windows with a length of 10 days.

FACTOR ANALYSIS OF THE TIME SERIES' VECTOR OF PROPERTIES

Let us try to separate different states in the 22-year history of the groundwater level observation time series using the cluster analysis of the 10-dimensional vector of properties (12). We note that separate components of vector ζ clearly show that the entire observation history can be divided into several intervals with the dissimilar behavior, e.g., of $a_1^{(1)}, \log \sigma_1^2, c_{\text{Pred}}, \alpha^*, En$. For formally partitioning the obtained cloud of vectors ζ into clusters, we preliminarily perform the dimensionality reduction procedure using factor analysis. The factor analysis model (Harmon, 1967) in this case is described by the formula

$$z = \Lambda f + e, \tag{13}$$

where the 10-dimensional vector z is obtained from vector ζ by the normalization procedure, which consists of eliminating the sample mean and division by the sample estimate of standard deviation for each component of vector ζ. After performing the normalization procedure, the correlation matrix R_{zz} is calculated.

In formula (13), f is the vector of dimension $q < p = 10$ composed of the hidden (latent) factors, i.e., certain random factors that govern the values of the scalar components of the multidimensional vector z through multiplication by the matrix of factor loadings Λ with p rows and q columns. The matrix elements $\Lambda = (\lambda_{j\alpha})$, $j = 1, ..., p$; $\alpha = 1, ..., q$ are the unknown model parameters that it is required to determine from the known sample estimate R_{zz} of the correlation matrix of the initial data. We assume for now that the number of the hidden parameters q is known. Concerning the properties of the random vector f, it is assumed that its mean is zero, $M\{f\} = 0$, and its cova-
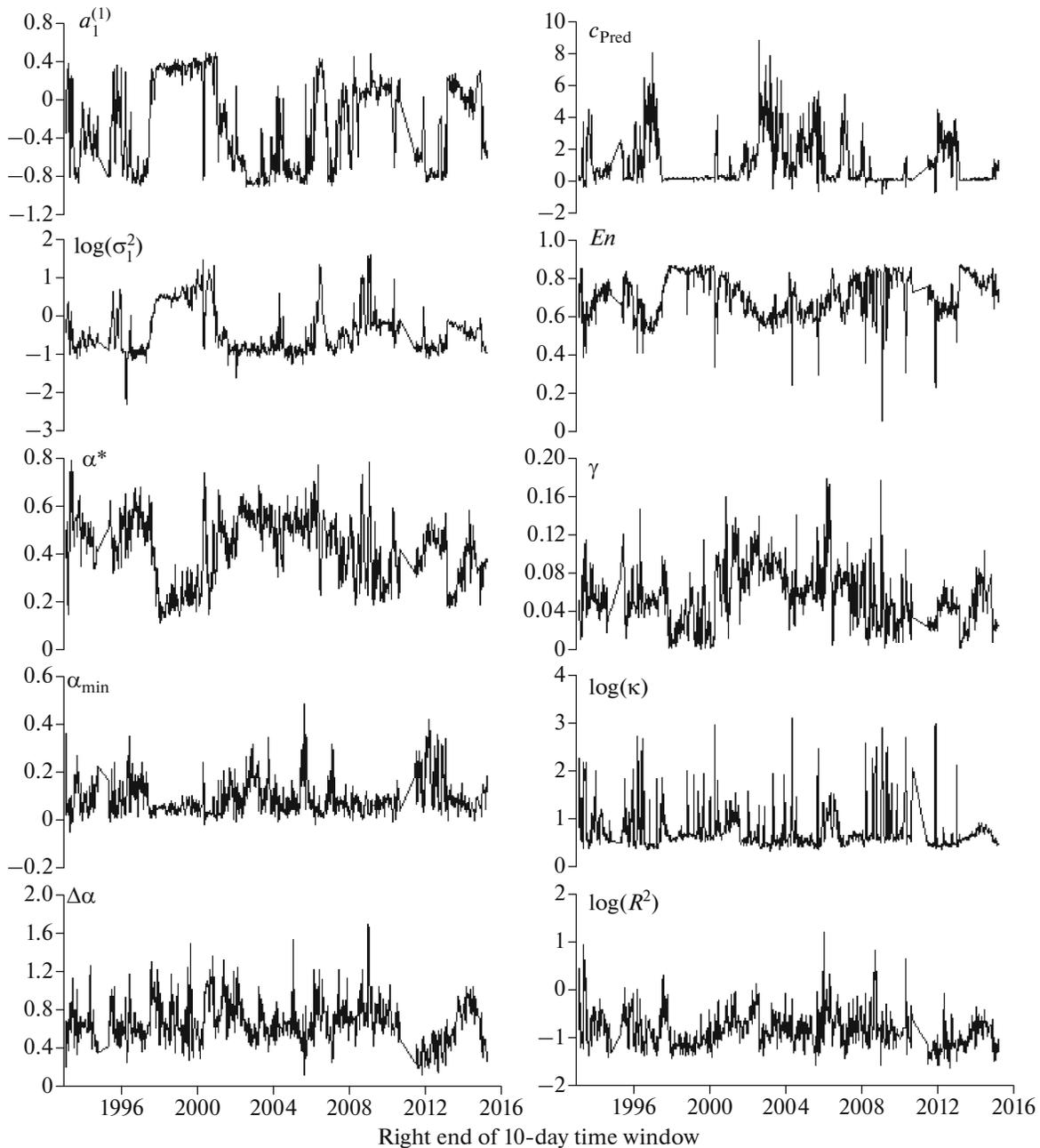
Fig. 6. Graphs of ten properties of groundwater level time series after compensating effect of atmospheric pressure in successive 10-day windows.

riance matrix is an identity (unit) matrix $M\{ff^T\} = I_q$, where I_q is the identity matrix of size q.

This condition means the orthogonality of the factors (in the Gaussian case, their independence). The condition that the variances of the orthogonal factors are equal to 1 is a sort of normalization since otherwise this could be achieved by scaling the elements of matrix Λ. Vector e in formula (13) has the same dimension as the initial vector z and consists of random quantities that describe the noise in each compo-

nent of vector z, i.e., that do not contain useful information. Since the noise in the different components should be independent, it is assumed for vector e that it is centered and that its covariance matrix is diagonal: $M\{ee^T\} = \Psi^2 = \text{diag}(\psi_1^2, ..., \psi_p^2)$, where $\psi_j^2, j = 1, ..., p$ are the so called residual variances or variances of the noise. The elements of the diagonal matrix Ψ^2 are also the parameters of model (13).

The parameters of model (13) can be identified by several methods; however, the method of minimal

residuals (Harman, 1967) is the most reliable and simplest one among them. From the condition that the covariance matrices of vectors f and e are diagonal it can be easily derived that, due to model (13), the covariance matrix of vector z is

$$\Sigma = M\{zz^T\} = \Lambda\Lambda^T + \Psi^2. \qquad (14)$$

The method of minimal residuals consists in determining the elements of matrix Λ from the condition of the minimum sum of squared differences between the sample estimates and the theoretical values of the pairwise coefficients of correlation. Thus, the criterion of the model's closeness to the data is the closeness of all the theoretical correlation coefficients to their sample estimates. We denote matrix elements R_{zz} by r_{ij}. Then, it is necessary to minimize the following function of the elements of the matrix of factor loadings:

$$\Phi(\Lambda) = \sum_{j=1}^{p-1} \sum_{i=j+1}^{p} \left(r_{ij} - \sum_{\alpha=1}^{q} \lambda_{j\alpha}\lambda_{i\alpha}\right)^2 \to \min_{\Lambda}. \qquad (15)$$

In this case, the elements of matrix Λ should obey the conditions

$$\sum_{\alpha=1}^{q} \lambda_{j\alpha}^2 \le 1, \quad j = 1,...,p \qquad (16)$$

following from the condition that the diagonal elements of the theoretical correlation matrix (14) are unities. It is worth noting that the problem of determining matrix Λ is independent of determining the diagonal matrix of the residual variances Ψ^2. After the solution of the minimization problem (15) under constraints (16), the residual variances are found automatically:

$$\psi_j^2 = 1 - \sum_{\alpha=1}^{q} \lambda_{j\alpha}^2, \quad j = 1,...,p. \qquad (17)$$

When the matrix of factor loadings is determined, the analysis comes to its final step which consists in calculating the realization of the orthogonal factors—the cloud of q-dimensional vectors f. The simplest estimate follows from the condition that the noise vector e is distributed by the p-dimensional normal distribution with the covariance matrix Ψ^2. In this case, the maximum likelihood estimate will be the weighted least squares estimate:

$$e^T\Psi^{-2}e \to \min_{f}, \quad f = (\Lambda^T\Psi^{-2}\Lambda)^{-1}\Lambda^T\Psi^{-2}z. \qquad (18)$$

However, estimate (18) gives the vector of common factors with the nondiagonal covariance matrix. In order for the components of the factors' vector to be orthogonal, the modification of estimate (18) proposed in (Anderson and Rubin, 1956) should be used:

$$f = (\Lambda^T\Psi^{-2}\Sigma\Psi^{-2}\Lambda)^{-1/2}\Lambda^T\Psi^{-2}z,$$
$$\Sigma = \Lambda\Lambda^T + \Psi^2. \qquad (19)$$

The purpose of obtaining the realization of the vector of common factors f is to reduce the dimension of the problem (Ayvazyan et al., 1989). The question of what the number q of the common factors (the dimension of vector f) should be is most difficult in factor analysis. For solving this question, Lawley and Maxwell (1971) proposed using the Rippe criterion, which is based on the assumption of the normal distribution of vectors z. However, this criterion has shown itself to be highly sensitive to small deviations from normality, which makes it impracticable. If the a priori information about number q is absent, the estimate of the maximum permissible number of common factors can be obtained by starting to solve the problem from the minimum value $q = 1$ and then gradually increasing q by unity until the factor analysis model degenerates (the total number of the parameters becomes redundant). After this occurs, q can be specified as the last maximal value before the degeneracy of the problem. The degeneracy of the problem of factor analysis is referred to as the Heywood case (Harman, 1967) and consists in zeroing the residual variance ψ_j^2 for one or a few components of vector z. In practice, a sharp drop (by several orders of magnitude) of the residual variance ψ_j^2 for some component compared to the other components is observed instead of zeroing,

This method for selecting q was used in our case. It determined the maximum allowable value of the number of common orthogonal factors at $q = 4$. Figure 7 shows the graphs of four common orthogonal factors for the set of 10-dimensional vectors of the properties of the groundwater level time series after compensating for the effect of atmospheric pressure.

CLUSTER ANALYSIS OF ORTHOGONAL COMMON FACTORS

After reducing the dimension of the set of the vectors of static parameters for a sequence of time intervals of the time series by considering four orthogonal common factors, we identify the clusters in the space of common factors using the popular method of k-means (also known as ISODATA) (Ayvazyan et al., 1989; Duda and Hart, 1973). In our case, the classification objects are the points in 4-dimensional Euclidean space and each component of these vectors has zero mean and unit standard deviation. Therefore, it is logical to introduce the ordinary Euclidean distance between the vectors. Let us consider a cloud of 4-dimensional vectors f of the common orthogonal factors. Within the minimal parallelepiped that contains the points f that are to be classified, there are randomly located centers of the trial clusters, and the number $q \ge 2$ of these clusters is fixed. Let Γ denote the initial random position of the trial clusters. For a given arrangement of the centers of the clusters, a test partitioning of the set of points is carried out according
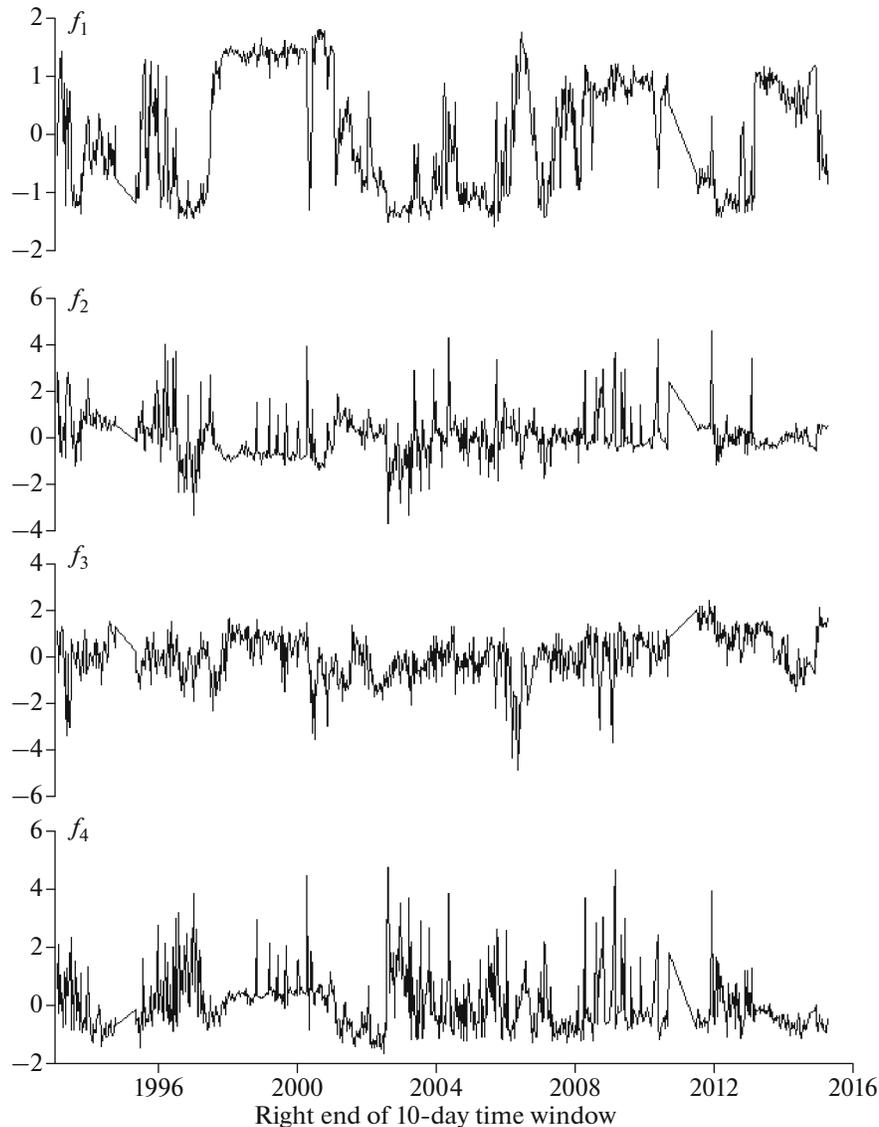
Fig. 7. Graphs of four orthogonal common factors for set of ten properties of groundwater level time series after compensating effect of atmospheric pressure in successive 10-day windows.

to the principle of the minimum distance to some center. Let $c_k, k = 1, ..., q$ be vectors of the clusters' centers, n_k be the number of points in the kth cluster, and $\sum_{k=1}^{q} n_k = M$ be the total number of points in the set under clustering. In our case, $M = 746$, which corresponds to the number of successive time intervals with a length of 10 days that contain at most 2% of the synthetic (padded) data replacing the gaps. Let B_k be the set of vectors pertaining to the kth cluster. Let us calculate the vectors of the centers of gravity of the resultant clusters: $r_k = \sum_{f \in B_k} \xi / n_k$. If for all $c_k = r_k$, the partition is stopped. Otherwise, the vectors of the clusters' centers c_k are moved to the centers of gravity r_k, new partitioning into clusters is carried out, the new

centers of gravity are calculated, the termination criterion of partitioning is checked, etc. The procedure converges rapidly. However, the partition that is obtained upon the termination of the iterations depends on the random positions of the centers of the trial clusters Γ at the beginning of the iterations. The quality of the final partitioning is estimated by the criterion of compactness of the clusters:

$$J(q|\Gamma) = \sum_{k=1}^{q} \sum_{f \in B_k} |f - c_k|^2. \qquad (20)$$

For a given number of clusters q, it is natural to try to find the random initial arrangement Γ for which quantity (20) is minimal. This is achieved by the Monte Carlo method: random experiments on scat-

tering the centers of trial clusters inside the cloud of points are repeated many times (below, we used 10^4 trials in the analysis of the specific data), after which the partition that provided the minimum with respect to Γ is selected.

The problem now arises to determine the optimal number of clusters into which the set of properties should be partitioned. Let $J_0(q) = \min_\Gamma J(q|\Gamma)$. If we successively reduce the number of trial clusters q from a certain fairly large number to the minimal $q = 2$, quantity $J_0(q)$ will grow monotonically; however, it will have a bend at the optimal number of clusters (if such a number exists). A more efficient method for revealing the optimal number of clusters consists in using the pseudo F-statistics (Vogel and Wong, 1979); this method is adopted from the dispersion analysis:

$$\text{PFS}(q) = (M - q)$$
$$\times \sum_{k=1}^{q} n_k |c_k - r_0|^2 \Big/ ((q - 1)J_0(q)), \tag{21}$$

where $r_0 = \sum f \big/ M$ is the common center of gravity of the entire set of points that we are trying to classify. The optimal number of clusters corresponds to the point of the maximum of function (21).

Figure 8 shows the dependence of the pseudo F-statistics on the number of trial clusters, from which it can be seen that the optimal number of the clusters in the space of the orthogonal common factors is 5.

SPECTRAL ANALYSIS OF THE SEQUENCE OF TRANSITIONS BETWEEN CLUSTERS

Figures 9a and 9b show the sequence of transitions between the selected five states and the sequence of 279 events of the change of a cluster's number. Let us examine the intensity of these transitions for the presence of periodic components. Below we use the method of identifying the periodic components of a point process based on calculating the difference between the maximum values of the log likelihood functions for two models: a Poisson process with constant intensity and the process with a periodic component proposed in (Lyubushin et al., 1998).

Let $t_i, i = 1, \ldots, N$ be the times of the sequence of events observed on the interval $(0, T]$. We consider the following intensity model with a periodic component:

$$\lambda(t) = \mu(1 + a\cos(\omega t + \varphi)), \tag{22}$$

where frequency ω, amplitude $a, 0 \le a \le 1$, phase angle φ, $\varphi \in [0, 2\pi]$, and factor $\mu > 0$ (describing the Poisson part of intensity) are the model parameters. Thus, the Poisson part of the intensity is modulated by harmonic oscillation.

We fix some frequency value ω. The log likelihood function of model (22) in this case for the series of observed events is
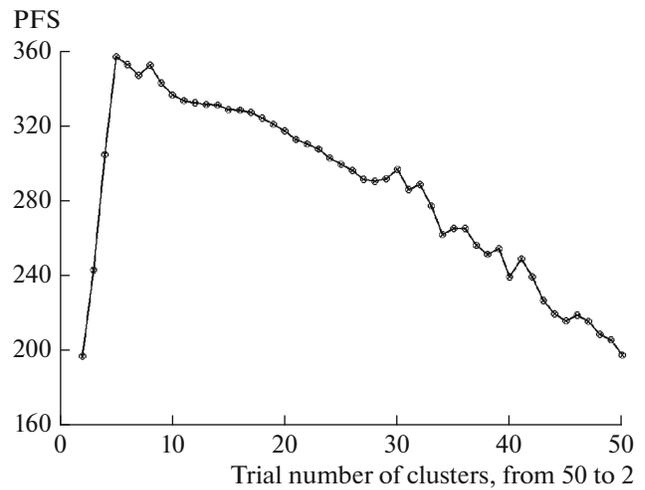

Fig. 8. Graph of pseudo F-statistics for clustering of four orthogonal common factors.

$$\ln L(\mu, a, \varphi | \omega) = \sum_{t_i} \ln(\lambda(t_i)) - \int_0^T \lambda(u)du$$
$$= N\ln(\mu) + \sum_{t_i} \ln(1 + a\cos(\omega t_i + \varphi)) \tag{23}$$
$$- \mu T - \frac{\mu a}{\omega}[\sin(\omega T + \varphi) - \sin(\varphi)].$$

Taking the maximum of expression (23) with respect to parameter μ, we find that

$$\mu = \hat{\mu}(a, \varphi | \omega)$$
$$= \frac{N}{T + a(\sin(\omega T + \varphi) - \sin(\varphi))\big/\omega}. \tag{24}$$

Substituting (24) in formula (23), we obtain

$$\ln(L(\hat{\mu}, a, \varphi | \omega)) = \sum_{t_i} \ln(1 + a\cos(\omega t_i + \varphi))$$
$$+ N\ln(\hat{\mu}(a, \varphi | \omega)) - N. \tag{25}$$

Since our intensity model with a harmonic component with the given frequency ω is richer than the model for a purely random event flow, the increment of the log likelihood function is

$$\Delta \ln L(a, \varphi | \omega) = \sum_{t_i} \ln(1 + a\cos(\omega t_i + \varphi))$$
$$+ N\ln(\hat{\mu}(a, \varphi | \omega)/\hat{\mu}_0). \tag{26}$$

Assume that

$$R(\omega) = \max_{a, \varphi} \Delta \ln L(a, \varphi | \omega),$$
$$0 \le a \le 1, \varphi \in [0, 2\pi]. \tag{27}$$

Function (27) can be considered as a generalization of the spectrum for the sequence of events. The graph of this function shows the extent to which the periodic model of intensity is advantageous compared to the purely random model. The maximal value functions (27)
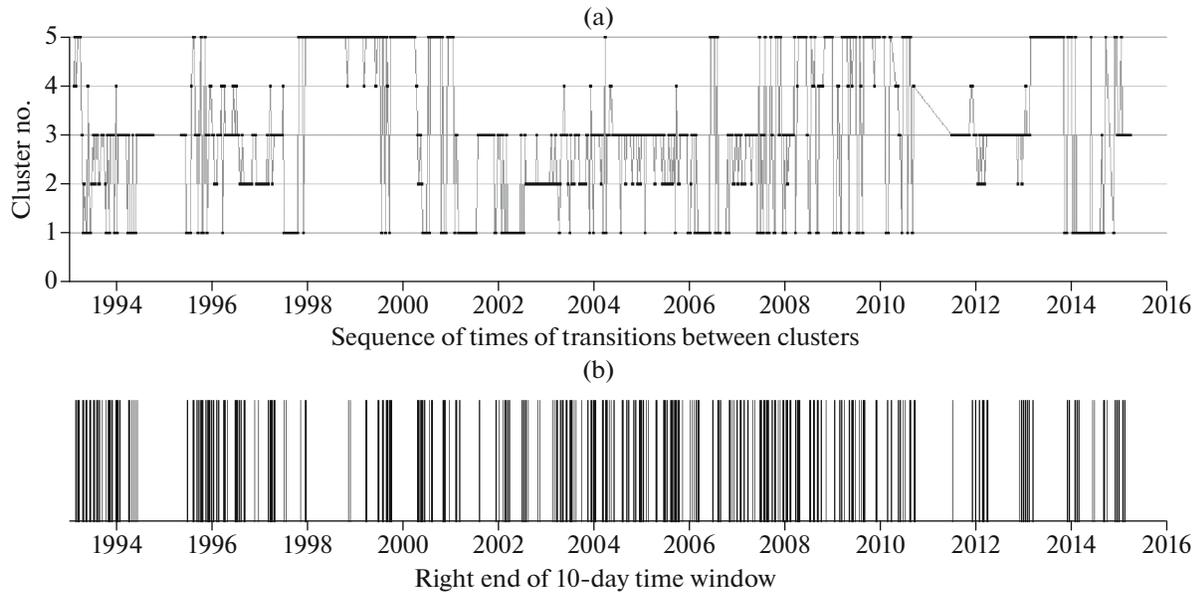
Fig. 9. (a) Sequence of five clusters in space of four common factors of statistical properties of 10-day time fragments of groundwater level time series after compensating for effect of atmospheric pressure (Fig. 7); (b) sequence of 279 transitions between clusters.

distinguish the frequencies present in the flow of events. As shown in the work (Lyubushin et al., 1998)

$$P\{R(\omega) < x\} = 1 - e^{-x}, \quad N \to \infty. \qquad (28)$$

The asymptotic distribution (28) can be used for estimating the statistical significance of the maxima of function (27). From formula (28) it follows that the probability that quantity (27) exceeds the threshold

Fig. 10. Graph of maximum difference between log likelihood functions versus period for time sequence of transitions between five clusters (Fig. 9b) in time interval from June 22, 1995 to September 10, 2010 (217 events) for trial periods from 20 to 1000 days. Two periods with significance of 98% at 46 and 275 days are indicated.

value 4 is 0.02; i.e., the asymptotic significance of frequencies ω for which $R(\omega) > 4$ is 98%.

Direct application of the described method of the spectral analysis of the sequence of events presented in Fig. 9b is impeded by the presence of fairly long data gaps in 1995 and at the end of 2010. These data gaps can lead to the emergence of false low-frequency components. Therefore, we consider the sequence of 217 transitions between clusters only for the time interval from June 22, 1995 to September 10, 2010 when there were no long recording breaks. Figure 10 shows a graph of function (27) calculated for 2000 trial period values on a uniform logarithmic scale from 20 to 1000 days. From this graph it follows that the transitions between clusters have two significant periods (with the probability of the hypothesis of the existence of these periodicities of at least 98%) with periods of 46 and 275 days.
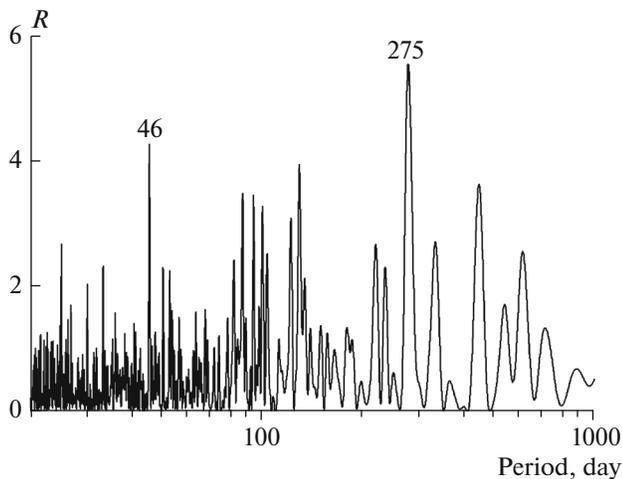
TESTING THE HYPOTHESIS OF LINKAGE WITH THE STRONGEST EARTHQUAKES

The frequentist estimates of the probability that the properties of the time series occur in each of the five identified states presented in Fig. 9a are 0.217, 0.120, 0.341, 0.055, and 0.267 for states 1 to 5, respectively. Thus, the fourth state can be named anomalous because the probability that the properties of the time series are in this cluster is 5.5%. Let us test the hypothesis that the occurrence of the properties of the time series in this anomalous cluster is somehow associated with the sequence of the 27 strongest earthquakes with a magnitude of at least 9, which occurred during the 22 years of observations (1993–2015). In Figs. 11a and 11b,
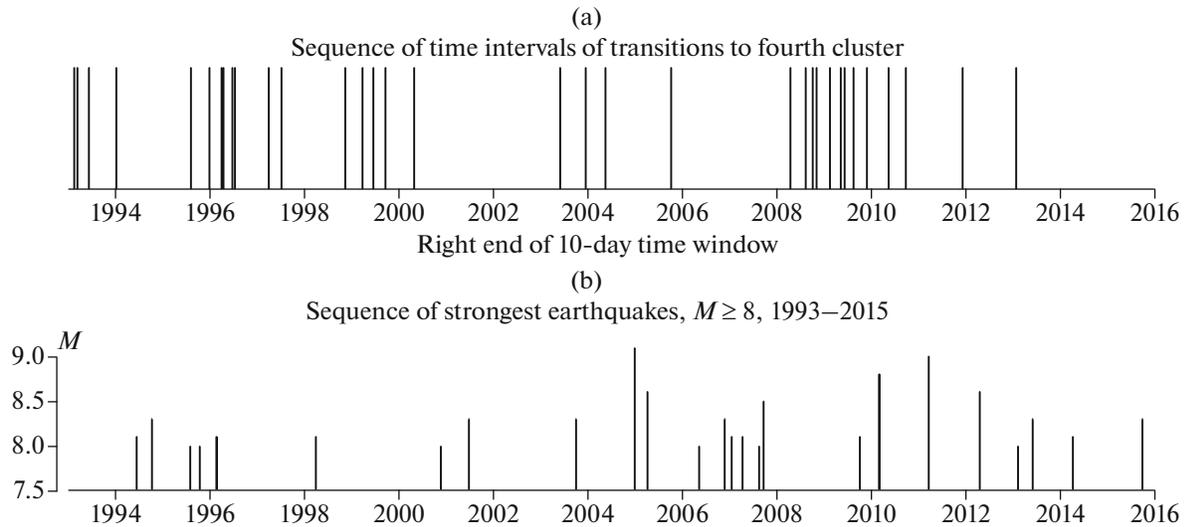
(a)
Sequence of time intervals of transitions to fourth cluster

Right end of 10-day time window

(b)
Sequence of strongest earthquakes, $M \geq 8$, 1993−2015

Fig. 11. (a) Sequence of 10-day time intervals when transition to fourth anomalous cluster takes place; (b) sequence of strongest earthquakes with magnitudes of at least 8.

the vertical lines show two sequences of the events: 34 time instants corresponding to the right end of the 10-day time window after which the properties of the time series pass into the fourth cluster (Fig. 11a) and the 27 strongest earthquakes with a magnitude of at least 8 that occurred during the 22 years of observations from 1993 to 2005 (Fig. 11b).

For testing the hypothesis, we use the method of influence matrices based on a linear model of the interaction between the flows of events. This method was suggested in (Lyubushin and Pisarenko, 1993) (see also (Lyubushin, 2007)), where the authors considered the model of interaction between the sequences of seismic events from several regions and utilized the influence function of events with the power-law decay, which is used in the ETAS model (epidemic-type aftershock sequence) (Ogata and Katsura, 1993). Here, the power exponent of the decay law is included in the list of the model parameters to be determined. Below, the method is simplified and described for the particular case of two sequences of events and the exponentially decaying influence function of the event.

Let $t_j^{(\alpha)}$, $j = 1,...,N_\alpha$; $\alpha = 1,2$ be the time instants (points) of two event flows. We consider the intensity of a certain process in the form

$$\lambda^{(\alpha)}(t) = b_0^{(\alpha)} + \sum_{\beta=1}^{2} b_\beta^{(\alpha)} g^{(\beta)}(t), \qquad (29)$$

where $b_0^{(\alpha)} \geq 0$, $b_\beta^{(\alpha)} \geq 0$ are parameters and $g^{(\beta)}(t)$ is the influence function of the events of the flow with number β. We represent the influence function of an event in the following form:

$$g^{(\beta)}(t) = \sum_{t_j^{(\beta)} < t} \exp(-(t - t_j^{(\beta)})/\tau), \qquad (30)$$

where $\tau > 0$ is the characteristic time scale in the analysis of the interaction between the flows of the events. Thus, in accordance with formula (30), the weight of the event with number j becomes nonzero for the times $t > t_j^{(\beta)}$ and exponentially decays with the characteristic time τ with the increase of the current time t. The sum of all these decaying exponentials forms the influence function $g^{(\beta)}(t)$ of the flow with number β. Parameter $b_\beta^{(\alpha)}$ is a scaling factor and it determines the degree of influence of flow β on flow α: $\beta \rightarrow \alpha$. Parameter $b_\alpha^{(\alpha)}$ determines the degree of influence of flow α on itself (self-excitation), and parameter $b_0^{(\alpha)}$ reflects a purely random intensity component for which the influence function is constant and identical to 1.

We fix parameter τ and consider the problem of determining the parameters $b_0^{(\alpha)}$ and $b_\beta^{(\alpha)}$. The log likelihood function for a nonstationary Poisson process is (Cox, Lewis, 1966)

$$\ln(L_\alpha) = \sum_{j=1}^{N_\alpha} \ln(\lambda^{(\alpha)}(t_j^{(\alpha)})) - \int_0^T \lambda^{(\alpha)}(s)ds, \qquad (31)$$

$$\alpha = 1, 2,$$

where $[0, T]$ is the observation interval. Thus, it is necessary to find the maximum of function (31) with respect to the parameters $b_0^{(\alpha)}$ and $b_\beta^{(\alpha)}$. Taking into account formula (29) and using the rule of differentiation of a complex function, we easily obtain the following expression:

Table 1. Influence matrixes for analyzing relationship between strongest earthquakes and transition times to anomalous cluster for two decay times of exponential influence function of events, 10 and 100 days

Decay time τ, days	Event sequence	Poisson part	Self-excitation	Influence of other sequence
10	Earthquakes	0.954	0.046	0.000
	Transition to fourth cluster	1.000	0.000	0.000
100	Earthquakes	1.000	0.000	0.000
	Transition to fourth cluster	0.784	0.216	0.000

$$b_0^{(\alpha)} \frac{\partial \ln(L_\alpha)}{\partial b_0^{(\alpha)}} + \sum_{\beta=1}^{2} b_\beta^{(\alpha)} \frac{\partial \ln(L_\alpha)}{\partial b_\beta^{(\alpha)}} \quad (32)$$
$$= N_\alpha - \int_0^T \lambda^{(\alpha)}(s)ds.$$

Since the parameters $b_0^{(\alpha)}$ and $b_\beta^{(\alpha)}$ should be non-negative, each term in the left-hand side of formula (32) is zero at the maximum point of function (31), either because of the necessary extremum conditions (if the parameters are positive) or, if the maximum is reached at the boundary, the parameters themselves are zero. Hence, at the maximum point of the log likelihood function (31), the following equality is true:

$$\int_0^T \lambda^{(\alpha)}(s)ds = N_\alpha. \quad (33)$$

We substitute expression (29) into (33) and divide by the length of the observation interval. Then we obtain another record of formula (33):

$$b_0^{(\alpha)} + \sum_{\beta=1}^{m} b_\beta^{(\alpha)} \overline{g}^{(\beta)} = \lambda_0^{(\alpha)} \equiv \frac{N_\alpha}{T}, \quad (34)$$

where $\overline{g}^{(\beta)} = \frac{1}{T} \int_0^T g^{(\beta)}(s)ds$ is the mean value of the influence function. Substituting $b_0^{(\alpha)}$ from (34) into (31), we obtain the following maximum problem, which is equivalent to the maximization problem (31):

$$\Phi^{(\alpha)}(b_1^{(\alpha)}, b_2^{(\alpha)})$$
$$= \sum_{j=1}^{N_\alpha} \ln(\lambda_0^{(\alpha)} + \sum_{\beta=1}^{2} b_\beta^{(\alpha)} \Delta g^{(\beta)}(t_j^{(\alpha)})) \rightarrow \max, \quad (35)$$

where $\Delta g^{(\beta)}(t) = g^{(\beta)}(t) - \overline{g}^{(\beta)}$, under the constraints

$$b_1^{(\alpha)} \geq 0, \quad b_2^{(\alpha)} \geq 0, \quad \sum_{\beta=1}^{2} b_\beta^{(\alpha)} \overline{g}^{(\beta)} \leq \lambda_0^{(\alpha)}. \quad (36)$$

Function (35) is convex with a negatively definite Hessian (Lyubushin and Pisarenko, 1993) and, therefore, problem (35) and (36) has a unique solution. This problem is solved numerically by the gradient projection method (Moiseev et al., 1978).

Having solved problem (35) and (36) for the given τ, we can introduce the fractions of intensity $\mu_\beta^{(\alpha)}, \alpha = 1, 2$ and $\beta = 0, 1, 2$ according to the formulas

$$\kappa_0^{(\alpha)} = \frac{b_0^{(\alpha)}}{\lambda_0^{(\alpha)}} \geq 0, \quad \kappa_\beta^{(\alpha)} = \frac{b_\beta^{(\alpha)} \overline{g}^{(\beta)}}{\lambda_0^{(\alpha)}} \geq 0, \quad (37)$$

which can be referred to as the elements of the influence matrix.

The interpretation of these quantities is quite natural: $\kappa_0^{(\alpha)}$ is part of the average intensity of the process with number α, which is purely stochastic, the part $\kappa_\alpha^{(\alpha)}$ is due to the influence of self-excitation $\alpha \rightarrow \alpha$, and the part $\kappa_\beta^{(\alpha)}$, $\beta \neq \alpha$ is associated with an external influence. From formula (33), the normalization condition follows:

$$\kappa_0^{(\alpha)} + \sum_{\beta=1}^{2} \kappa_\beta^{(\alpha)} = 1, \quad \alpha = 1, 2. \quad (38)$$

Table 1 presents the results of calculating the elements of the influence matrix for two decay times, 10 and 100 days. It can be seen that the considered sequences of events are Poisson for a decay time of 10 days; however, in the case of a decay time of 100 days, a small self-excited component appears in the time sequence of the transitions in the fourth cluster. However, for all the decay times considered, there is no link between the event flows.

CONCLUSIONS

The diagram of the transitions between the statistically significant clusters shown in Fig.9a is the final result of the suggested multivariate analysis of the properties of the time series of the groundwater level's observations. The spectral analysis of the point process of the transition times between different clusters (Fig. 9b) identified significant periodic components in the intensity of these transitions with periods of 46 and 275 days (Fig. 10). The presence of these periods probably reflects both the regional and global factors affecting a large underground horizon. Testing the linkage between the transitions into an anomalous cluster and the strongest earthquakes in the world

(Fig. 11) gave a negative result: this connection for the observations in the aseismic region is absent (Table 1).

The further interpretation of the obtained results needs information about the other geophysical fields in the studied region. Nevertheless, some general considerations about the presence of several states of the time series of long-term observations can be inferred.

In the aseismic regions, which include the territory of Moscow, the release of the energy flow from the Earth's interior is smooth. It does not occur as a sharp excitation of seismic waves but rather as the enhancement of slow movements of the Earth's crust along relatively narrow deformation localization zones—the lineaments. Intensification of these crustal movements can enhance the groundwater migration capable of weakening the crustal blocks and strengthen the karst-suffosion processes. In turn, the latter increases the probability of landslides, cracking of the foundations of large buildings, breakthrough of underground tunnels, and corrosion of subway lines. Thus, the identified sequence of time fragments having different statistical characteristics can be used for correlating the time intervals pertaining to the different clusters to the intensity of manmade accidents in the territory of the megalopolis.

FUNDING

REFERENCES

Aivazyan, S.A., Bukhshtaber, V.M., Enyukov, I.S., and Meshalkin, L.D., *Prikladnaya statistika. Klassifikatsiya i snizhenie razmernosti* (Applied Statistics. Classification and Dimension Reduction), Moscow: Finansy i statistika, 1989.

Anderson, T.W. and Rubin, H., Statistical inference in factor analysis, *Proc. Third Berkley Symp. on Math. Statistics and Probability*, 1956, vol.5, pp.111−150.

Bagmet, A.L., Bagmet, M.I., Barabanov, V.L., et al., Studying the solid-earth tidal fluctuations of the groundwater level in well "Obninsk," *Izv. Akad. Nauk SSSR*, 1989, no. 11, pp. 84−95.

Besedina, A.N., Vinogradov, E.A., Gorbunova, E.M., Kabychenko, N.V., Svintsov, I.S., Pigulevskiy, P.I., Svistun, V.K., and Shcherbina, S.V., The response of fluid-saturated reservoirs to lunisolar tides: Part 1. Background parameters of tidal components in ground displacements and groundwater level, *Izv., Phys. Solid Earth*, 2015, vol. 51, no. 1, pp. 70−79.

Boldina, S.V. and Kopylova, G.N., Effects of the January 30, 2016, Mw = 7.2 Zhupanovsky earthquake on the water level variations in wells YuZ-5 and E-1 in Kamchatka, *Geodinam. Tektonofiz.*, 2017, vol. 8, no. 4, pp. 863−880.

Box, G.E.P., and Jenkins, G.M., *Time Series Analysis: Forecasting and Control*, San Francisco: Holden-Day, 1970.

Bredehoft, J.D., Response of a well-aquifer system to earth tides, *J. Geophys. Res.*, 1967, vol. 12, pp. 3075−3087.

Brillinger, D.R., *Time Series: Data Analysis and Theory*, New York: Holt, Rinehart and Winston, 1975.

Cox, D.R. and Lewis, P.A.W., *The Statistical Analysis of Series of Events*, London: Methuen, 1966.

Cramer, H., *Mathematical Methods of Statistics*, Princeton: Princeton Univ. Press, 1999.

Donoho, D.L. and Johnstone, I.M., Adapting to unknown smoothness via wavelet shrinkage, *J. Am. Stat. Assoc.*, 1995, vol. 90, no. 432, pp. 1200−1224.

Duda, R.O., and Hart, P.E., *Pattern Classification and Scene Analysis*, New York: Wiley, 1973.

Feder, J., *Fractals*, New York: Plenum, 1988.

Gidrogeologiya SSSR (Hydrogeology of the USSR), Moscow: Nedra, 1966.

Harman, H.H., *Modern Factor Analysis*, Chicago: Univ. of Chicago Press, 1967.

Huber, P.J., *Robust Statistics*, New York: Wiley, 1981.

Igarashi, G. and Wakita, H., Tidal response and earthquake-related changes in the water level of deep wells, *J. Geophys. Res.*, 1991, vol. 96, pp. 4269−4278.

Kantelhardt, J.W., Zschiegner, S.A., Konscienly-Bunde, E., Havlin, S., Bunde, A., and Stanley, H.E., Multifractal detrended fluctuation analysis of nonstationary time series, *Phys. A (Amsterdam, Neth.)*, 2002, vol. 316, nos. 1−4, pp. 87−114.

Kashyap, R.L. and Rao, A.R., *Dynamic Stochastic Models from Empirical Data*, New York: Academic, 1976.

Kissin, I.G. and Gumen, A.M., Hydrogeological indicators of recent crustal movements in an aseismic region, *Dokl. Akad. Nauk*, 1994, vol. 334, no. 6, pp. 768−772.

Kopylova, G.N., Variations of water level in Elizovskaya-1 well, Kamchatka due tolarge earthquakes: 1987−1998 observations, *Vulkanol. Seismol.*, 2001, no. 2, pp. 39−52.

Kopylova, G.N., Earthquake-induced water level changes in the YuZ-5 well, Kamchatka, *Vulkanol. Seismol.*, 2006, no. 6, pp. 52−64.

Kopylova, G.N., Lyubushin, A.A., Malugin, V.A., Smirnov, A.A., and Taranova, L.N., Groundwater observations at the Petropavlovsk test site, Kamchatka, *Vulkanol. Seismol.*, 2000, no. 4, pp. 69−79.

Kopylova, G.N, Gorbunova, E.M., Boldina, S.V., and Pavlov, D.V., Estimation of deformational properties of a stratum−borehole system based on analysis of barometric and tidal responses of the water level in a borehole, *Izv., Phys. Solid Earth*, 2009, vol. 45, no. 10, pp. 905−913.

Lawley, D.N. and Maxwell, A.E., *Factor Analysis as a Statistical Method*, New York: Elsevier, 1971.

Lyubushin, A.A., *Analiz dannykh sistem geofizicheskogo i ekologicheskogo monitoringa* (Analysis of the Data of Geophysical and Ecological Monitoring), Moscow: Nauka, 2007.

Lyubushin, A.A., Synchronization trends and rhythms of multifractal parameters of the field of low-frequency microseisms, *Izv., Phys. Solid Earth*, 2009, vol. 45, no. 5, pp. 381−394.

Lyubushin, A.A., The statistics of the time segments of low-frequency microseisms: trends and synchronization, *Izv., Phys. Solid Earth*, 2010, vol. 46, no. 6, pp. 544−554.

Lyubushin, A., Multifractal parameters of low-frequency microseisms, in *Synchronization and Triggering: from Fracture to Earthquake Processes*, de Rubeis, V. et al., Eds., Geo-

Planet: Earth and Planetary Sciences 1, Berlin: Springer, 2010, Chapter 15, pp.253-272.

Lyubushin, A.A., Forecast of the Great Japan earthquake, *Priroda (Moscow, Russ. Fed.),* 2012a, no. 8, pp. 23−33.

Lyubushin, A., Prognostic properties of low-frequency seismic noise, *Nat. Sci.*, 2012b, vol. 4, no. 8A, pp. 659−666.

Lyubushin, A., How soon would the next mega-earthquake occur in Japan, *Nat. Sci*, 2013, vol. 5, no. 8, A1, pp. 1−7.

Lyubushin, A.A., Dynamic estimate of seismic danger based on multifractal properties of low-frequency seismic noise, *Nat. Hazards*, 2014a, vol. 70, no. 1, pp. 471−483.

Lyubushin, A.A., Prognostic properties of the stochastic fluctuations of geophysical characteristics, *Biosfera*, 2014b, no. 4, pp. 319−338.

Lyubushin A. Synchronization of geophysical fields fluctuations, in *Complexity of Seismic Time Series: Measurement and Applications*, Chelidze, T., Telesca, L., and Vallianatos, F., Eds., Amsterdam: Elsevier, 2018, Chapter 6, pp. 161−197. https://doi.org/10.1016/B978-0-12-813138-1.00006-7.

Lyubushin, A.A. and Lezhnev, M.Yu., Analysis of the variability of the function of groundwater level response to baric variations in South Kuriles (Shikotan Island), *Izv. Ross. Akad. Nauk, Fiz. Zemli*, 1995, no. 8, pp. 302−307.

Lyubushin, A.A. and Malugin, V.A., Statistical analysis of the groundwater level response to the variations in the atmospheric pressure, *Izv. Ross. Akad. Nauk, Fiz. Zemli*, 1993, no. 12, pp. 74−80.

Lyubushin, A.A. and Pisarenko, V.F., Studying the seismic regime with the use of a linear model of the intensity of the interacting point processes, *Izv. Ross. Akad. Nauk, Fiz. Zemli*, 1993, no. 12, pp. 81−87.

Lyubushin, A.A., Malugin, V.A., and Kazantseva, O.S., Monitoring of tidal variations of the underground water level in a group of water-bearing horizons, *Izv., Phys. Solid Earth*, 1997, vol. 33, no. 4, pp. 52−64.

Lyubushin, A.A., Pisarenko, V.F., Ruzhich, V.V., and Buddo, V.Yu., A new method for identifying seismicity periodicities, *J. Volcanol. Seismol.*, 1998, vol. 20, no. 1, pp. 73−89.

Lyubushin, A.A., Malugin, V.A., and Kazantseva, O.S., Recognition of "slow events" in an aseismic region, *Izv., Phys. Solid Earth*, 1999, vol. 35, no. 3, pp. 195−203.

Mallat, S., *A Wavelet Tour of Signal Processing*, second ed., San Diego: Academic, 1999.

Moiseev, N.N., Ivanilov, Yu.P., and Stolyarova, E.M., *Metody optimizatsii* (Optimization Methods), Moscow: Nauka, 1978.

Ogata, Y. and Katsura, K., Analysis of temporal and spatial heterogeneity of magnitude frequency distribution inferred from earthquake catalogues, *Geophys. J. Int.*, 1993, vol. 113, pp. 727−738.

Roeloffs, E., Hydrologic precursors to earthquakes: a review, *Pure Appl. Geophys.*, 1988, vol. 126, nos. 2/4, pp. 177−209.

Roeloffs, E., Burford, S., Riley, F., and Records, A., Hydrologic effects on water level changes associated with episodic fault creep near Parkfield, California, *J. Geophys. Res.*, 1989, vol. 94, no. B9, pp. 12387−12402.

Rojstaczer, S. and Agnew, D.C., The influence of formation material properties on the response of water levels in wells to Earth tides and atmospheric loading, *J. Geophys. Res.*, 1989, vol. 94, pp. 12403−12411.

Vinogradov, E.A., Gorbunova, E.M., Kabychenko, N.V., Kocharyan, G.G., Pavlov, D.V., and Svintsov, I.S., Groundwater level monitoring by precision measurements, *Geoekol. Inzh. Geol. Gidrogeol. Geokriol.*, 2011, no. 5, pp. 439−449.

Vogel, M.A. and Wong, A.K.C., PFS clustering method, *IEEE Trans. Pattern Anal. Mach. Intell.*, 1979, vol. 1, pp. 237−245. doi 10.1109/TPAMI.1979.4766919

Translated by M. Nazarenko