ISSN 0747-9239, Seismic Instruments, 2019, Vol. 55, No. 5, pp. 507–523. © Allerton Press, Inc., 2019.
Russian Text © The Author(s), 2019, published in Seismicheskie Pribory, 2019, Vol. 55, No. 1, pp. 5–28.

Statistical Analysis of Precision Water Level Data from Observations in a Seismoactive Region: Case Study of the YuZ-5 Well, Kamchatka

G. N. Kopylova[a], *, A. A. Lyubushin[b], and S. V. Boldina[a]

[a] *Kamchatka Branch, Geophysical Survey, Russian Academy of Sciences, Petropavlovsk-Kamchatsky, 683006 Russia*
[b] *Schmidt Institute of Physics of the Earth, Russian Academy of Sciences, Moscow, 123242 Russia*
**e-mail: gala@emsd.ru*

Abstract—A new method is presented for statistical analysis of long-term time series of water level observations aimed at distinguishing short-term disturbances; observation data from the YuZ-5 well, located in the Petropavlovsk Geodynamic Test Area, eastern Kamchatka, are considered. These data (from July 27, 2012, to February 1, 2018) are remarkable for their degree of detail: the sampling rate of the water level and atmospheric pressure measurements was 5 min and the sensitivity (accuracy) was ±0.1 cm for water level recording and ±0.1 hPa for atmospheric pressure. Also, five strong earthquakes with $M_w = 6.5–8.3$ occurred at epicentral distances of $d_e = 80–700$ km during the observation period. A thorough analysis of the hydrodynamic regime of the observation well over a long period and the high quality of observation data, together with the data on strong seismic events, allow us to consider the possibility of using formalized statistical methods of water level data processing for diagnostics of anomalous conditions. As a result of factor and cluster analysis applied to the sequence of multidimensional vectors of the statistical properties of water level time series in successive one-day-long time windows, after adaptive compensation for atmospheric pressure, four different statistically significant states of time series, replacing each other in time, are distinguished. Geophysical interpretation of the anomalous conditions of the water level time series (with a probability of 0.013) is carried out in comparison to strong earthquakes, technical conditions of observations, and seasonal features of the hydrodynamic regime in the observation well. It is shown that this method of water level data processing can detect short-term anomalies in the hydrogeodynamic regime of a well, significantly supplementing traditional processing of water level data aimed mostly at finding low-frequency trends in water level changes. This method can be applied in geophysical monitoring and prediction of earthquakes from online processing of water level data in wells.

Keywords: well, water level, earthquake, precursor, Kamchatka, time series, adaptive filtering, factor analysis, cluster analysis

DOI: 10.3103/S0747923919050086

INTRODUCTION

Precise measurements of water level in wells allow researchers to detect changes in groundwater pressure in the range of periods from seconds and minutes to tens and hundreds of days. The possible mechanisms causing and affecting pressure variations are quasielastic deformation of water-saturated rocks, development of fracture dilatancy in them, and other processes leading to changes in the capacity and permeability of rocks (Kissin, 1993, 2009; Kopylova, 2006b). The sensitivity of groundwater pressure to changes in the stress–strain state of rocks allows the application of water level data in a broad range of earth sciences problems. Water level variations in wells are used to study lunar and solar tides (Bredehoeft, 1967; Rojstaczer and Agnew, 1989; Lyubushin et al., 1997; Vinogradov et al., 2011). The possibility of assessing variability in crustal properties from the response of water level to variations of atmospheric pressure was studied in (Lyubushin and Malugin, 1993; Lyubushin and Lezhnev, 1995; Kopylova et al., 2001).

In seismoactive regions, water level measurements are used to study how seismicity affects water level variations through the effects of seismic waves, coseismic deformation of water-saturated rocks, and hydrogeodynamic earthquake precursors (Roeloffs, 1988; Roeloffs et al., 1989; Igarashi and Wakita, 1991; Kissin, 1993, 2009; Kopylova, 2006b; Kopylova et al., 2010; Wang and Manga, 2010). Recent studies in Kamchatka have shown that application of traditional processing methods to water level measurement data (Kopylova, 2001, 2006a), aimed at distinguishing low-frequency trends in water level variations, can aid in detecting hydrogeodynamic precursors (HPs) before strong earthquakes. HPs are expressed as water level changes in time reference intervals from a few tens of days to months and years before earthquakes with magnitudes of about 7–8 at epicentral distances of up

to a few hundred kilometers from observed wells (Kopylova et al., 2001; Kopylova, 2001, 2006a; Boldina and Kopylova, 2017). Examples of successful mid-term predictions of strong earthquakes in Kamchatka using HPs are provided in (Chebrov et al., 2011; Firstov et al., 2016).

At the same time, the problem of distinguishing relatively short-term anomalies in the hydrogeodynamic regimes of wells, hidden in noise variations of water level, is still disputable and needs to be solved as applied to problems of searching for short-term HPs and other geodynamic activity signals. The need to diagnose earthquake precursors in water level time series in a broad range of periods implies advances in experimental data processing methods with the use of formalized statistical analysis procedures for subsequent development of unified software applied for this type of geophysical observations.

In the present article, we use time series from the Yuz-5 well in Kamchatka to test the new statistical analysis method; the used time series of water level measurements has a sampling rate of 5 min, after adaptive compensation for atmospheric pressure variations in it, by means factor and cluster analysis of the sequence of multidimensional vectors corresponding to eight statistical properties of the observation data series in sequential time windows of one day long. Four different statistically significant states, alternating in time, have been identified in variations within the data series. Notably, three out of four of these states are considered background ones, while one is considered anomalous. Geophysical interpretation of variations corresponding to the anomalous state is made with respect to strong earthquakes, technical conditions of observations, and other factors.

TECHNICAL CONDITIONS OF OBSERVATIONS AND INITIAL DATA

Since 2003, the Kamchatka Branch, Geophysical Survey, Russian Academy of Sciences (KB GS RAS) has been carrying out observations of water level changes in the YuZ-5 well with an interval of 5 min using instrumentation manufactured by LLC Polinom, Khabarovsk (Kopylova et al., 2016).

The coordinates of the YuZ-5 well are 53.169° N, 158.414° E, and its depth is 800 m. In the depth range from 0 to 310 m, the well shaft is cased in a metallic pipe. At 310–800 m, the well shaft is open and directly connected to water-bearing rocks represented by interbedding of Late Cretaceous siltstones and shales. The water permeability of rocks is 7.8 m^2/day, and groundwater mineralization is 0.25 g/L. The water level is 1–1.5 m below the ground surface.

Water level changes demonstrated an intraannual seasonal character with an amplitude of up to 50 cm (Boldina and Kopylova, 2017), as well as barometric and tidal variations (Kopylova, 2006a). Local strong earthquakes cause coseismic and postseismic variations in water level changings. The hydrogeodynamic precursors were retrospectively distinguished in water level variations before the Kronotskoye (December 5, 1997, M_w = 7.8) and Zhupanovo (January 30, 2016, M_w = 7.2) earthquakes (Kopylova, 2006a; Kopylova and Boldina, 2012; Boldina and Kopylova, 2017). The passage of surface seismic waves from the strongest earthquakes with M_w of about 8–9 at epicentral distances of hundreds to thousands of kilometers due to water level variations (hydroseisms) with amplitudes of up to 9 cm and durations of several hours to one day. More detailed data on the well structure, elastic and filtering properties of water-saturated rocks, regularities of the hydrogeodynamic regime, and earthquake effects in water level changes are presented in our earlier publications (Kopylova, 2006a; Kopylova and Boldina, 2006; Kopylova et al., 2010, 2016; Boldina and Kopylova, 2016, 2017) and in publications by other authors.

The instruments used are characterized by high resolution (the recording accuracy is ±0.1 cm for water level and ±0.1 hPa for atmospheric pressure), broad dynamic range, and long-term stability and reliability of continuous observations. In order to ensure the integrity of the instruments, a protective reinforced concrete building with a metal door was constructed above the well head. Technical control of the observation system is carried out by the Laboratory of Geophysical Research and includes regular, at least quarterly, visits to the well in order to test the instruments and perform maintenance. Dates and times of visits, as well as the list of work done and test results, are documented in a special digital log. Processing and online analysis of current data transmitted to KB GS RAS via a cellular network are done by an operator on a daily basis, providing additional opportunities to monitor the technical state of the observation system.

In this study, we used time series of the water level and atmospheric pressure measured from July 27, 2012, to February 1, 2018 (in total, 2015 days) with a sampling rate (count interval) of 5 min (Fig. 1). The complete duration of water level and atmospheric pressure records is 580 605 five-minute counts. There are 424 singular gaps in the data (0.007% of the total duration of two time series); these gaps are related to technical failures of the recording system. They were taken into account by linear interpolation between the values of the nearest recorded counts.

During the observations, there were 32 visits. Five times operations were carried out to remove and install sensors recording groundwater parameters in the upper part of the well shaft. In particular, on April 16, 2014, and June 5, 2014, these operations entailed removal and installation of a device for measuring water temperature and electrical conductivity at a depth of 20 m; on March 1, 2016, hoisting and running of the water level sensor; on August 7, 2017, installation

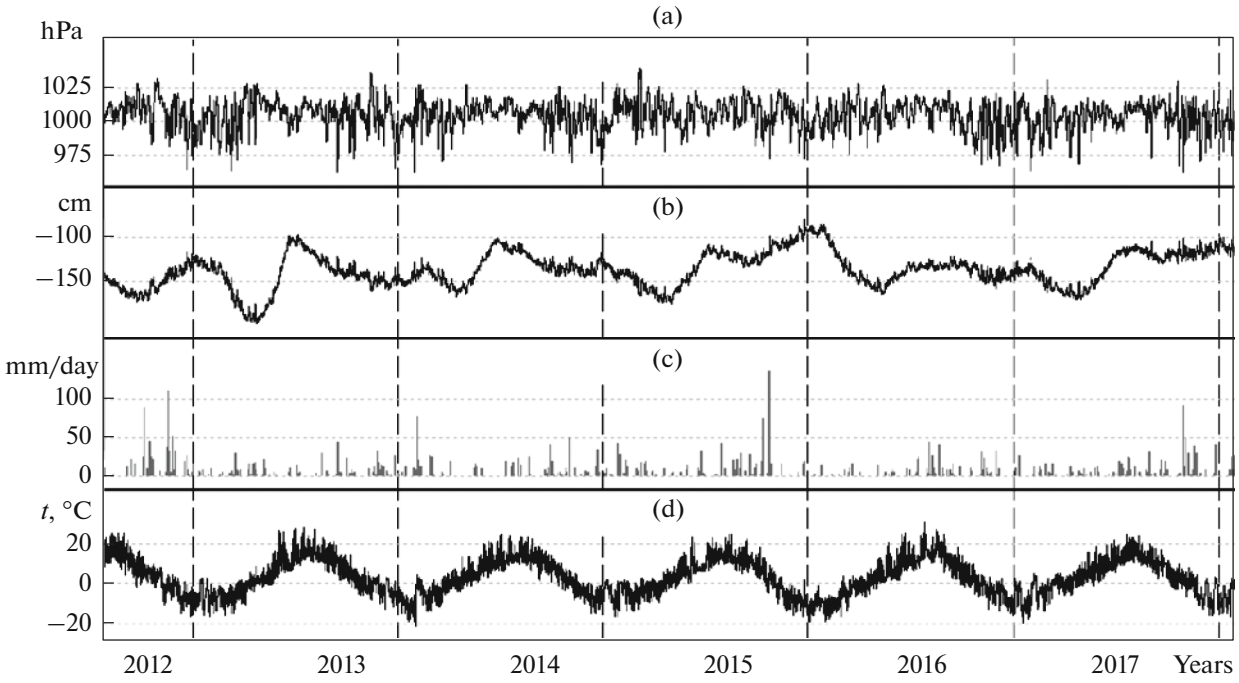

Fig. 1. Data of 5-min observations of (a) atmospheric pressure and (b) water level in YuZ-5 well from July 27, 2012, to February 1, 2018, compared with changes of (c) precipitation and (d) air temperature based on data from Pionerskaya meteorological station, Kamchatka Administration for Hydrometeorology and Environmental Control.

of an additional water pressure sensor at a depth of 5.6 m; and on October 4, 2017, repeated hoisting and running of the water level sensor with an amplitude of ≈20 cm. These technical operations were accompanied by changes in water level, either a rise or drop with an amplitude of 1–2 cm, with subsequent restoration of the stationary regime in 6–8 h.

Strong Local Earthquakes and Hydrogeoseismic Variations of Water Level

The considered time interval has been remarkable not only in the high quality of water level data, but also in that five earthquakes with $M_w = 6.5–8.3$ occurred in the Kamchatka and western Aleutian seismoactive zones at epicentral distances of d_e = 80–700 km (Table 1). The descriptions of events numbered 1 to 3 in Table 1, which occurred in 2013, are given in (*Sil'nye*..., 2014). In (Chebrov et al., 2016), data on the Zhupanovo earthquake are presented; in (Chebrov et al., 2017), data on the Near Aleutian earthquake. In the area of the well, these earthquakes were manifested as shaking with an intensity from V to II–III on the MSK-64 scale (Medvedev et al., 1965). All the mentioned earthquakes were reflected in the YuZ-5 well as various hydrogeoseismic variations of water level. Event nos. 1, 2, 4, and 5 (Table 1) were accompanied by coseismic effects during the first minutes after rupture at the earthquake source, as well as by longer-term postseismic changes of water level. Retrospective analysis revealed that event no. 4 (Table 1) was preceded by an HP in the form of a water level rise by 28 cm over 3.5 months (Boldina and Kopylova, 2017).

COMPENSATION OF BAROMETRIC VARIATIONS IN WATER LEVEL CHANGES AND POWER SPECTRA

A change of atmospheric pressure is the main meteorological factor affecting water level variation in the YuZ-5 well in the range of periods from hours to days (Fig. 1). Earlier, based on the behavior of the amplitude transfer function from atmospheric pressure variations to changes of water level, it was found that barometric response of water level is characterized by a constant value of barometric efficiency, E_b = 0.4 cm/hPa, in the range of periods from 6 h to a few tens of days. In periods of 2–6 h, E_b increases monotonically from 0.1 to 0.4 cm/hPa (Kopylova, 2006a, 2009).

Before statistical analysis of the measured water level data, we need to remove the influence of atmospheric pressure on the initial time series of water level changes. For this, we applied a compensating adaptive frequency filter (Lyubushin, 2007). In moving windows of 28 day long (8064 counts), with a 5-min step (1 count), we calculated power spectrum $S_{uu}(\omega)$ of atmospheric pressure and the complex cross-spectrum $S_{xu}(\omega)$ between the water level and atmospheric pressure, depending on frequency ω. These estimates

Table 1. Parameters of strong earthquakes based on data from KB GS RAS (http://www.emsd.ru/), Global CMT (http://www.globalcmt.org), and NEIS USGS (https://earthquake.usgs.gov/earthquakes/search/)

No.	Hypocenter					Energy		Epicentral distance, d_e, km/shaking intensity on the MSK-64 scale
	date (dd.mm.yyyy) and name of earthquake	time, hh:mm	coordinates, deg		H, km	class, K_S	magnitude, M_w	
			N	E				
1	28.02.2013	14:06	50.67	157.77	61	15.2	6.8	260/4–5
2	24.05.2013, Sea of Okhotsk	05:45	54.76	153.79	630	17.0	8.3	348/4
3	12.11.2013	07:04	54.63	162.45	72	15.0	6.5	300/3–4
4	30.01.2016, Zhupanovo	03:25	53.85	159.04	180	15.7	7.2	80/5
5	17.07.2017, Near Aleutian	23:34	54.35	168.90	7	16.1	7.8	700/2–3

are obtained by smoothing of periodograms and cross-periodograms by a frequency window having a length of 1/32 of the moving window length. Then, the frequency transfer function $H_{xu}(\omega) = S_{uu}(\omega)/S_{uu}(\omega)$ was calculated in every window. Before smoothing of periodograms, tidal frequency bands of [1/11, 1/13] and [1/23, 1/27] h^{-1} were suppressed and estimates within the limits of these frequency bands were obtained by interpolation of estimates from the adjacent frequency values. The compensation value $\tilde{E}_x(\omega)$ in the frequency zone within the limits of every time window was 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 from the water level and atmospheric pressure within the current window. The result of compensation $e_x(t)$ in the time zone within the limits of every time window was determined by inverse discrete Fourier transform from $\tilde{E}_x(\omega)$.

The final operation to obtain the compensated signal is sewing the compensation results within the limits of different windows together to obtain entire signal. The contribution of the first window to this signal is its first half, whereas the second half is taken from the last window. Regarding the remaining "intermediate" time windows, only their central counts are taken to form the entire signal.

Figure 2 shows the graphs of the initial and compensated water levels, hereinafter, $U_k(t)$) series for the first half of 2013 when the seismic events 1 and 2 (Table 1) occurred. We can see that tidal (Fig. 2c), coseismic, and postseismic variations of the level are distinguished much better in the compensated signal $U_k(t)$ (Fig. 2b) compared to the initial data (Fig. 2a), where they are hidden by variations mainly caused by atmospheric pressure.

Figure 3 shows the estimated power spectra of variations of the initial time series of water level and $U_k(t)$ series. The tidal harmonics of semidiurnal and diurnal groups are clearly seen in Fig. 3a. In the range of tidal periods in the power spectrum of $U_k(t)$ series, nine tidal waves are clearly seen (Fig. 3b). Note that weakly expressed spectral peaks at periods corresponding to the higher overtones of diurnal variations (6, 4, and 2 h) remained after the compensation of atmospheric pressure variations. These overtones occur as a result of (a) temporally nonuniform heating and cooling of air during the day and (b) the fact that the diurnal temperature variation differs from a pure sine shape. These temperature changes affect atmospheric pressure and are accompanied by the corresponding water level responses. We should also note how the power spectrum straightens (on a double logarithmic scale): after compensation of atmospheric pressure variations, the "hill" in the power spectrum at periods of 10 to 1000 h disappeared.

STATISTICAL PROPERTIES OF THE TIME SERIES OF COMPENSATED WATER LEVEL $U_k(t)$

Further analysis implies estimation of statistical properties describing the behavior of water level time series $U_k(t)$ in sequential time frames and using of the obtained values for identifying different states of the time series. The length of time frame was chosen at $N = 288$ counts (each 5 min long) corresponding to one day.

Below we provide a brief description of eight used statistics. They were selected taking into consideration their previous applications to analyze other time series in monitoring systems, notably, not only geophysical ones (see the respective links below). It is remarkable that these statistics are not linked to physical nature of the analyzed signals. They describe very general properties of such time series as entropy, degree of difference from chaotic behavior, predictability, degree of nonstationarity of the behavior, shape of power spectrum, etc. All estimates were made for the time series representing augmentation of the compensated groundwater level $U_k(t)$ after application of winsorization (Huber and Ronchetti, 2009) in an interval of $\pm 3\sigma$ (hereinafter, $x(t)$ series) in order to ensure the stability (robustness) of the obtained statistical estimates to various surges.

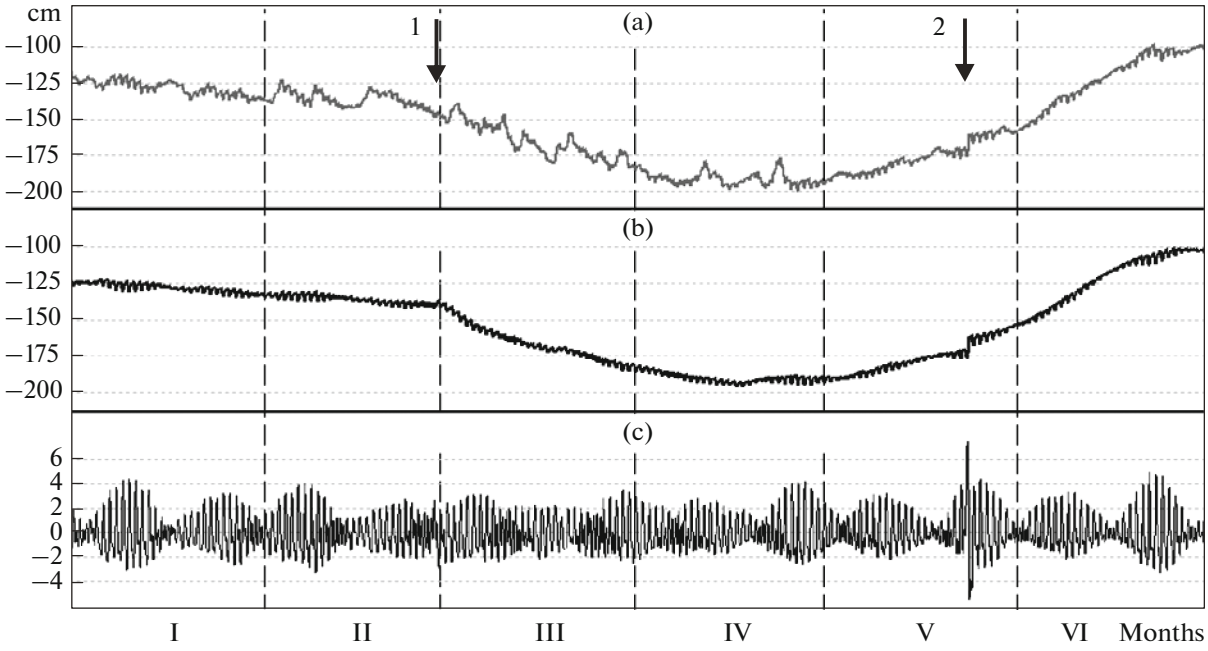

Fig. 2. Changes of water level in YuZ-5 well in January–June 2013: (a) initial data recorded every 5 min; (b) data after compensation for atmospheric pressure variations ($U_k(t)$ series); (c) tidal variations. Arrows denote earthquakes; numerals correspond to those in Table 1.

Minimum Normalized Entropy of Wavelet Factors En

Let $x(t)$ be a finite sampling of some random signal, $t = 1, ...,$ and N be a discrete time index (count marker). Let $c_j^{(k)}$ be the wavelet coefficient of the analyzed signal, whose superscript index k is the number denoting the level of detail of orthogonal wavelet decomposition and subscript index j denotes the sequence of centers of the time intervals in vicinities the vicinities of which signal convolution c is calculated by the finite elements of the basis. We used 17 Daubechies orthogonal wavelets: ten normal bases with the minimum carrier having from one to ten clearable moments, and seven Daubechies simlets (Mallat, 1999) with four to ten clearable moments. For each basis, the normalized entropy of the distribution the squared coefficients was calculated and the basis providing the minimum entropy was found:

$$En = -\sum_{k=1}^{m}\sum_{j=1}^{M_k} p_j^{(k)} \ln p_j^{(k)} \Big/ \ln N_r \to \min,$$
$$p_j^{(k)} = \left|c_j^{(k)}\right|^2 \Big/ \sum_{l,i}\left|c_i^{(l)}\right|^2, \quad (1)$$

where m is the number of considered levels of detail; M_k, the number of wavelet coefficient at level of detail k. The number of levels m depends on the length N of the analyzed sampling. For example, if $N = 2^n$, then $m = n$, $M_k = 2^{(n-k)}$. The condition $N = 2^n$ is necessary for fast wavelet transform to be applied. If the length N is not a power of two, the signal $x(t)$ is supplemented with zeros to attain the minimum length L which is greater than or equal to N: $L = 2^n \geq N$. In this case, among all values of $2^{(n-k)}$ for all wavelet coefficients at level k, only $N \times 2^{-k}$ coefficients correspond to decomposition of the real signal, while the remaining coefficients are zero due to addition of zeros to signal $x(t)$. Thus, in formula (1) $M_k = N \times 2^{-k}$ and the entropy is calculated using only "real" coefficients $c_j^{(k)}$. Number N_r in formula (1) is equal to the number of real coefficients, i.e., $N_r = \sum_{k=1}^{m} M_k$. By construction, $0 \leq En \leq 1$. The En statistics was used in (Lyubushin, 2012, 2014) to study the prognostic properties of seismic noise in the region of the Japanese Islands.

Donoho–Johnstone Index γ

When a wavelet basis is found for a given signal from the minimum entropy condition, it is possible to determine the set of least-modulus wavelet coefficients. In wavelet filtering, these wavelet coefficients can be cleared before inverse wavelet transform in order to "decrease noise" (Donoho and Johnstone, 1995; Mallat, 1999). We assume that noise is concentrated mainly in variations at the first level of detail. Recall that the first level of detail corresponds to the highest-frequency variations in the time series, characterized by periods from $2\Delta t$ to $4\Delta t$, where Δt is the sample spacing. Due to the orthogonal character of wavelet transform, the dispersion of the wavelet coef-

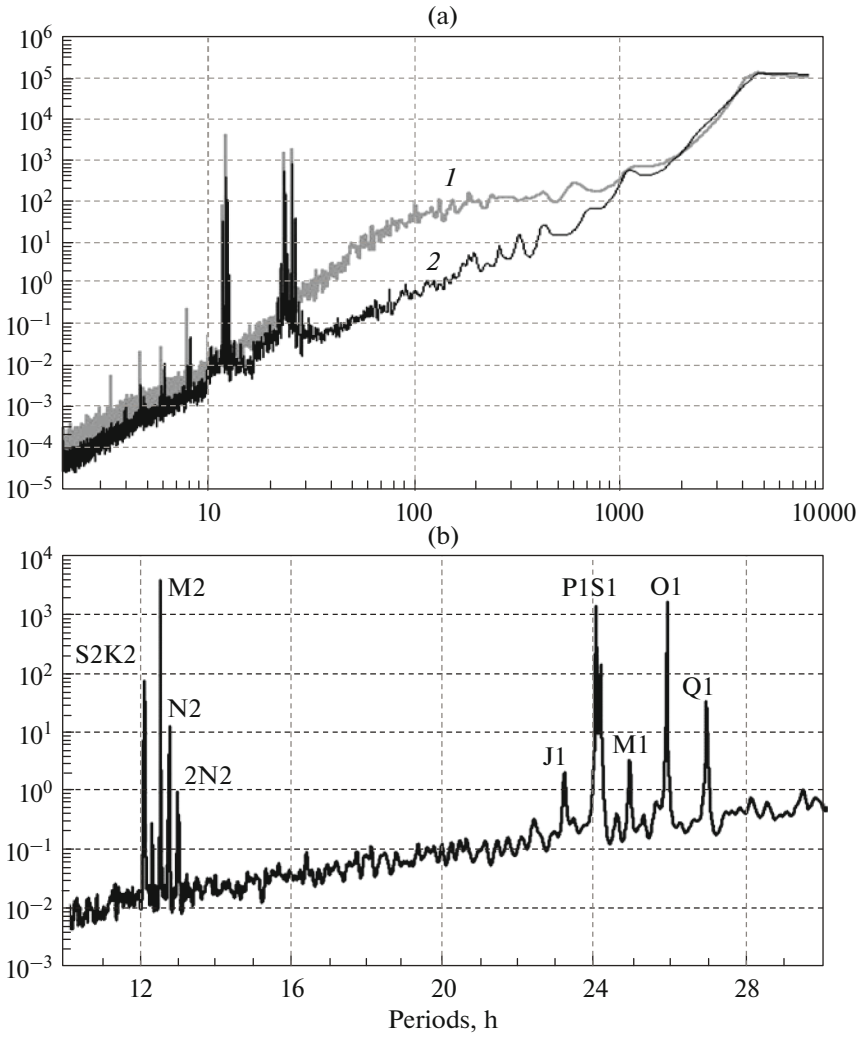

Fig. 3. Power spectra of water level variations in YuZ-5 well: (a) initial 5-min time series (*1*) and compensated $U_k(t)$ time series (*2*); (b) variations in $U_k(t)$ series in range of tidal waves from diurnal (J1, P1S1, M1, O1, and Q1) and semidiurnal (S2K2, M2, N2, and 2N2) groups. Names of tidal waves are after (Melchior, 1966).

ficients is equal to one initial signal. Hence, we estimate the standard deviation of noise as the standard deviation of the wavelet coefficients at the first level of detail. This estimate should be stable, i.e., insensitive to surges in the values of wavelet coefficients at the first level. For this, we can use the robust median estimate of the standard deviation for a normal random value:

$$\sigma = med\left\{\left|c_k^{(1)}\right|, k = 1,\ldots, N/2\right\}/0.6745, \qquad (2)$$

where $c_k^{(1)}$ is the wavelet coefficient at the first level of detail and $N/2$ is the number of these coefficients. The estimated standard deviation σ from formula (2) determines quantity $\sigma\sqrt{2\ln N}$ as the natural threshold for distinguishing noise-related wavelet coefficients. Quantity $\sigma\sqrt{2\ln N}$ is known in wavelet analysis as the Donoho–Johnstone threshold, and its complete expression is based on the formula for the asymptotic probability of maximum evasion of Gaussian white noise (Mallat, 1999). As a result, we can define the dimensionless characteristics γ of the signal, $0 < \gamma < 1$, as the ratio of the number of the most informative wavelet coefficients, for which the inequality $|c_k| > \sigma\sqrt{2\ln N}$ is satisfied, to the total number N of all wavelet coefficients. Formally, the larger the index γ, the more informative (less noisy) the signal.

Wavelet Spectral Exponent β

For an optimal orthogonal wavelet, the mean values of the squared wavelet coefficients can be calcu-

lated at each level of detail: $S_k = \sum_{j=1}^{M_k} \left|c_j^{(k)}\right|^2 \Big/ M_k$. The mean value of the squared wavelet coefficients is the part of the oscillation energy corresponding to level of detail k. In other words, this value can be considered an estimate of the power spectrum for signal $x(t)$ in the frequency band corresponding to level of detail k (Mallat, 1999): $[f_{\min}^{(k)}, f_{\max}^{(k)}] = [1/(2^{(k+1)}\Delta s), 1/(2^k \Delta s)]$, where Δs is the length of the sampling interval (in this case, $\Delta s = 5$ min). The values of periods corresponding to the centers of frequency bands are $T_k = 2/(f_{\min}^{(k)} + f_{\max}^{(k)}) = 2\Delta s/(2^{-k} + 2^{-(k+1)})$. Quantities $S_k = S(T_k)$, $k = 1, ..., m$, are analogous to ordinary Fourier power spectra, with the only difference being that the S_k values are much smoother—this feature is convenient when calculating the spectral exponent (slope of the curve of the logarithm of the power spectrum as a function of the logarithm of the period). To calculate the spectral exponent, let us consider the model $\ln(S(T_k)) = \beta \ln(T_k) + c + \varepsilon_k$, where ε_k is a sequence of independent random values with zero average. Parameter β can be called the wavelet spectral exponent, the value of which can be found by the least squares method: $\sum_{k=1}^{m} \varepsilon_k^2 \to \min\limits_{\beta,c}$.

Autoregression Model

Now let us use the autoregression model (Box and Jenkins, 1970; Kashyap and Rao, 1976) for the $x(t)$ time series. Let us write it in general form:

$$x(t) + \sum_{k=1}^{p} a_k^{(p)} x(t-k) = e^{(p)}(t) + d^{(p)}. \qquad (3)$$

Here, the integer $p \geq 1$ denotes the order of autoregression and vector $c = (a_1^{(p)}, ..., a_p^{(p)}, d^{(p)})^T$ is the vector of unknown parameters. The superscript (p) in formula (3) emphasizes that the used autoregression model is of order p. Here, $a_k^{(p)}$ are the autoregression coefficients, $d^{(p)}$ is the static displacement parameter, and $e^{(p)}(t)$ is the residual signal with zero average and dispersion σ_p^2. Model (3) can be written in concise form:

$$\begin{gathered} x(t) = c^T Y(t) + e^{(p)}(t), \\ Y(t) = (-x(t-1), ..., -x(t-p), 1)^T. \end{gathered} \qquad (4)$$

Let there be a finite sampling $\{x(t), t = 1, ..., N\}$. Then, the estimate of parameter vector c from the condition of minimum sum of squared residuals $\sum_{t=p+1}^{N} \left(e^{(p)}(t)\right)^2 \to \min\limits_{c}$ is reduced to solution of a system of normal equations with a symmetric positively determined matrix A:

$$\begin{gathered} Ac = R, \quad A = \sum_{t=p+1}^{N} Y(t) Y^T(t), \\ R = \sum_{t=p+1}^{N} x(t) Y(t). \end{gathered} \qquad (5)$$

The complete parameter vector of model (4) is $\theta^{(p)} = (c^T, \sigma_p)^T$.

Below, we use, along with other parameters, the values of coefficient $a_1^{(1)}$ from the first-order autoregression model and the logarithm of residual dispersion in this model, $\lg \sigma_1^2$, to characterize fragments of the time series.

Index of Linear Predictability c_{Pred}

The index of linear predictability was introduced in (Lyubushin, 2010; see also (Lyubushin, 2012)). Let us consider the value $c_{\text{Pred}} = V_0/V_{AR} - 1$. Here, V_0 is the dispersion of error $\varepsilon_0(t+1)$ in a trivial prediction $\hat{x}_0(t+1)$ by one step forward for signal $x(t)$ which equals the mean value on the previous "small" time window having length of n counts: $\hat{x}_0(t+1) = \sum_{s=t-n+1}^{t} x(t)/n$. 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 counts in sequential "big" time fragments. The value V_{AR} is calculated by the analogous formula $V_{AR} = \sum_{t=n+1}^{N} \varepsilon_{AR}^2(t)/(N-n)$, where $\varepsilon_{AR}(t+1) = x(t+1) - \hat{x}_{AR}(t+1)$ is the error in linear prediction $\hat{x}_{AR}(t+1)$ by one step forward using second-order autoregression (AR) model, whose coefficients are also estimated by the previous "small" time window with a length of n counts.

Second-order autoregression was chosen because this was the minimum order for the AR model for which oscillating motion is described and the maximum of the spectral density of the AR model can fall within the frequency band between Nyquist and zero frequency. AR prediction employs the correlation property for adjacent values, and if there is correlation, $V_{AR} < V_0$ and $c_{\text{Pred}} > 0$. The length of "big" time window $N = 288$; the length of "small" one was $n = 48$.

Autoregression Measure of Signal Nonstationarity R^2

Let $x(t)$ be the studied signal; n, half-length of the moving window (hereinafter, short); τ, the center of the double moving window, which, as a result, includes counts t satisfying the condition $\tau - n \leq t \leq \tau + n$. Let us construct the scalar autoregression model (5) of order $p = 2$ for signal $x(t)$ for the left and right halves of the short window. Estimating the model independently of samplings that fall to the left and right halves of the double moving window, we obtain

parameter vectors $\theta_1^{(p)}$ and $\theta_2^{(p)}$, respectively. Let $\Delta\theta = \theta_1^{(p)} - \theta_2^{(p)}$ be the difference between the estimated vectors in the left and right halves of the moving time window.

If the behavior of the studied signal in the left and right halves differs considerably, the difference $\Delta\theta$ will increase. In order to "weigh" vector $\Delta\theta$, it is logical to use Fisher matrix s a metric one, because it defines the rate of change in the likelihood logarithmic function in the vicinity of the point with maximal matrix parameters, including second derivatives from the conditional likelihood logarithmic function of the autoregression model:

$$B = -\frac{\partial^2 \ln\Phi}{\partial\theta\partial\theta}, \quad \ln\Phi = -(n-p)\ln\sigma_p - \frac{1}{2\sigma_p^2}\sum_t (x(t) - c^T Y(t))^2. \tag{6}$$

Let us denote $B^{(1)}$ and $B^{(2)}$ matrices calculated on the left and right halves of the moving window, respectively. Then, the measure of nonstationary behavior for process $x(t)$ in the symmetric vicinity of point τ will be

$$r^2(\tau) = (\Delta\theta^T B^{(1)}\Delta\theta + \Delta\theta^T B^{(2)}\Delta\theta)/2(n-p). \tag{7}$$

In formula (7), the half-sum of the lengths of the parameter difference vector $\Delta\theta$, measured with metric matrices $B^{(1)}$ and $B^{(2)}$, is divided into $(n-p)$ counts in the left and right halves of the moving window, with the number of autoregression parameters being subtracted. Such a metric provides a natural dimensionless measure of nonstationarity in the behavior of the studied signal. Through simple manipulations, we obtain the following expression:

$$\Delta\theta^T B\Delta\theta = \frac{2(\Delta\sigma_p)^2}{\sigma_p^2} + \frac{\Delta c^T\left(\sum_t Y(t)Y^T(t)\right)\Delta c}{\sigma_p^2(n-p)} + \frac{4\Delta c^T \Delta\sigma_p \sum_t e^{(p)}(t)Y(t)}{\sigma_p^3(n-p)}, \tag{8}$$

which is useful when calculating the nonstationarity measure (7). The measure of nonstationary behavior was introduced in (Lyubushin et al., 1999; see also (Lyubushin, 2007)). In (Osorio et al., 2011) statistics (7) were used to analyze electroencephalograms during epilepsy studies, while in (Lyubushin and Farkov, 2017), it was used to analyze financial time series.

Using formulas (7) and (8), we can find another, more stable measure of nonstationary behavior of the studied signal within the limits of a long time interval, including N sequential counts. Let us take a short window having a radius of n counts, $2n + 1 < N$, and compute the measure of nonstationary behavior $r^2(\tau)$ for all possible positions of central point τ within the limits of a long window, for which a short window falls completely inside the long one. It is easy to find that the number of such positions of central point τ equals $N - 2n$. Let us find the integral nonstationarity measure R^2 for a long window as the median of $r^2(\tau)$ values for all acceptable values of central point τ of a short window within the limits of a long one. In our calculations, we used windows with lengths $N = 288$ and $n = 48$. Hereinafter, we will consider the logarithm of the nonstationarity measure, $\lg R^2$.

Excess Coefficient

Excess coefficient κ is determined by the formula $\kappa = M(x^4)/(M(x^2))^2$ (Cramer, 1999). It characterizes the acuteness of the probability density graph in the distribution of random value x with a nonzero average, yielding the measure of deviation of the probability density from the normal law with $\kappa = 3$. Here, the operation $M(...)$ means calculation of mathematical expectation—in this case, the simple sample average of a random value. The excess coefficient is usually understood as value of κ (see above) with subtraction of 3 so that excess will be zero for the normal law. However, hereinafter, we will consider the logarithm of excess $\log\kappa$ instead; therefore, no subtraction of 3 is done in order to ensure positive values of this logarithm

Thus, for each one-day-long time window, there are eight parameters characterizing various statistical properties $U_k(t)$ of the time series within the limits of this window: minimum normalized entropy of wavelet coefficients En, Donoho–Johnstone index γ; coefficient $a_1^{(1)}$ and dispersion logarithm $\lg\sigma_1^2$ in the first-order autoregression model; index of linear predictability c_{Pred} and the logarithm of the nonstationarity measure $\lg R^2$, which are based on the second-order autoregression model; wavelet spectral exponent β; and the logarithm of excess coefficient $\lg\kappa$.

Let us denote the 8D attribute vector, whose attributes characterize the statistical properties of the augmentation time series $U_k(t)$ within the limits of sequential one-day-long fragments (288 counts with a 5-min step) as follows:

$$\zeta = (En, \gamma, a_1^{(1)}, \lg\sigma_1^2, c_{\mathrm{Pred}}, \lg R^2, \beta, \lg\kappa). \tag{9}$$

Figure 4 shows changes in individual components of the 8D vector ζ of attributes of the augmentation series $U_k(t)$ as a function of the position of the right end of sequential one-day-long time windows.

FACTOR ANALYSIS OF THE ATTRIBUTE VECTOR OF THE TIME SERIES

Now let us attempt to identify different states in the history of the time series of observed changes of groundwater level using cluster analysis of the 8D

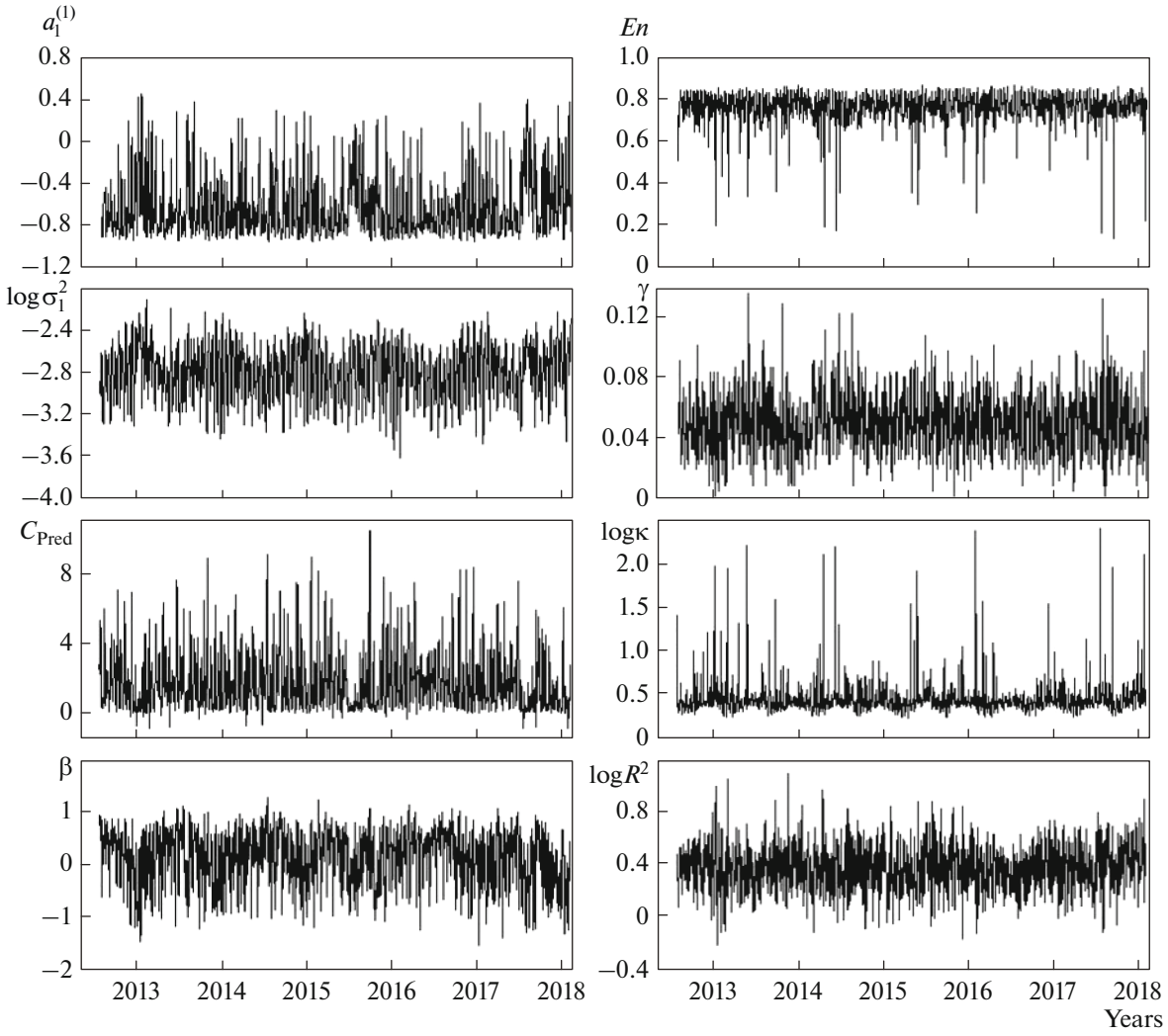

Fig. 4. Diagrams of eight properties of time series of water level increments, with atmospheric pressure compensated, $U_k(t)$, in one-day-long sequential windows.

attribute vector (9). In order to formally subdivide the obtained cloud of vectors ζ into clusters, let us preliminarily reduce the dimension using factor analysis. In our case, the factor analysis model (Harman, 1967) is described by the following formula:

$$z = \Lambda f + e, \tag{10}$$

where 8D vector z is obtained from vector ζ through normalization, implying that the sample average is removed and every component of vector ζ is divided by the sample estimate of the standard deviation. After normalization, the correlation matrix R_{zz} is calculated.

In formula (10), f means the vector of dimension $q < p = 8$, consisting of underlying factors (certain random vectors governing the values of scalar components of multidimensional vector z via multiplication by the matrix of factor loads Λ having p lines and q columns. The elements of matrix $\Lambda = (\lambda_{ja})$, $j = 1,..., p$; $a = 1, ..., q$ are unknown parameters of the model, and they need to be found, having a sample estimate of the correlation matrix R_{zz} of the initial data. Let us assume that the number of underlying parameters q is known. Regarding the random vector f, we also assume that its average value is zero, $M\{f\} = 0$, and its covariation matrix is unity, $M\{f \cdot f^T\} = I_q$, where I_q is a q-dimensional unity matrix.

This condition means orthogonality of factors (their independence in the Gaussian case). The condition of the dispersion of orthogonal factors being equal to unity is, in a certain sense, a normalization, otherwise this can be attained by scaling the elements of matrix Λ. Vector e in formula (10) has the same dimension as the initial vector z and consists of random values describing noise on every component of vector z, i.e., not carrying the desired signal. Since noise on different components should be independent, it is assumed that vector e is centered and its

covariation matrix is diagonal: $M\{e \cdot e^T\} = \Psi^2 = \mathrm{diag}\{\Psi_1^2,..., \Psi_p^2\}$, where $\Psi_j^2, j = 1, ..., p$ are the so-called residual variances or noise dispersions. The elements of diagonal matrix Ψ^2 are also the parameters of model (10).

The most reliable and simplest way to identify the parameters of model (10) is the minimum residual method (Harman, 1967). It is easy to find from the conditions of diagonality of the covariation matrices of vectors *f* and *e* that, due to model (10), the covariation matrix of vector *z* is

$$\Sigma = M\left\{z \cdot z^T\right\} = \Lambda \cdot \Lambda^T + \Psi^2. \quad (11)$$

The minimum residual method implies determination of the elements of matrix Λ from the condition of the minimum sum of the squared differences between sample estimates and theoretical values of paired correlation factors. Thus, the criterion of model proximity to the data is the proximity of all theoretical correlation factors to their sample estimates. Let r_{ij} denote elements of matrix R_{zz}. Then, it is necessary to minimize the next function of elements of the factor loading matrix:

$$\Phi(\Lambda) = \sum_{j=1}^{p-1}\sum_{i=j+1}^{p}(r_{ij} - \sum_{\alpha=1}^{q}\lambda_{j\alpha}\lambda_{i\alpha})^2 \to \min_{\Lambda}. \quad (12)$$

Note that elements of matrix Λ should have the following limitations imposed

$$\sum_{\alpha=1}^{q}\lambda_{j\alpha}^2 \le 1, \quad j = 1,..., p, \quad (13)$$

which follow from the condition that the diagonal elements of the theoretical correlation matrix be equal to unity (13). It should be noted that the problem of determining matrix Λ is independent of determining the diagonal matrix of residual dispersions Ψ^2. After the problem of finding the minimum (12) under limitations (13) is solved, residual dispersions can be found automatically:

$$\psi_j^2 = 1 - \sum_{\alpha=1}^{q}\lambda_{j\alpha}^2, \quad j = 1,..., p. \quad (14)$$

After the factor loading matrix is found, at the final step of analysis, it is necessary to calculate the realizations proper of orthogonal factors, namely, clouds of *q*-dimensional vectors *f*. The simplest estimate follows from the condition that noise vector *e* is distributed in accordance with a *p*-dimensional normal distribution with covariation matrix Ψ^2. In this case, the maximum likelihood estimator will be the estimate of the weighted least squares method:

$$e^T \cdot \Psi^{-2} \cdot e \to \min_f, \quad f = (\Lambda^T\Psi^{-2}\Lambda)^{-1}\Lambda^T\Psi^{-2} \cdot z. \quad (15)$$

However, estimate (15) yields the vector of general factors with a nondiagonal covariation matrix. In order that the components of factor vector be orthogonal, it is necessary to apply a modified version of (15), proposed in (Anderson and Rubin, 1956):

$$f = (\Lambda^T\Psi^{-2}\Sigma\Psi^{-2}\Lambda)^{-1/2}\Lambda^T\Psi^{-2} \cdot z, \quad \Sigma = \Lambda\Lambda^T + \Psi^2. \quad (16)$$

The aim of obtaining realizations of general factor vector *f* is to decrease the dimension of the problem (Aivazyan et al., 1989). The most complicated problem in factor analysis is selecting the number *q* of general factors (i.e., the dimension of vector *f*). We propose to solve it using the Rippe criterion (Lawley and Maxwell, 1971), which is based on the assumption of a normal distribution of vectors *z*. However, this criterion has demonstrated a high sensitivity to small deviations from normality, making it almost inapplicable. If there is no a priori information about number *q*, the estimate of maximum acceptable number of general factors can be obtained by solving the problem starting from the minimum value, $q = 1$, and gradually adding unity until the factor analysis model degenerates (i.e., the total number of parameters becomes excessive). After this, the last maximum value before degeneration can be taken as the value of *q*. Degeneration of the factor analysis problem is referred to as the Haywood case (Harman, 1967) and implies muting of the residual dispersion Ψ_j^2 for one or several components of vector *z*. In fact, instead of muting, an abrupt decrease (by several orders of magnitude) in residual dispersion is observed for some component in comparison to the other components.

We used precisely this method of selecting the *q* value, and the maximum acceptable number of general orthogonal factors appeared to be $q = 3$. Figure 5 shows the diagrams of three general orthogonal factors F1, F2, and F3 for the set.

CLUSTER ANALYSIS OF ORTHOGONAL COMMON FACTORS

After decreasing the dimension of the set of vectors of statistical properties for the $U_k(t)$ time series by transition to consideration of three general orthogonal factors, let us identify clusters in the space of general factors F1, F2, and F3 by applying the method of *k*-means (also known as ISODATA) (Aivazyan et al., 1989; Duda and Hart, 1973). In our case, the classification objects are points in three-dimensional Euclidean space, and each component of these vectors has a zero average and standard deviation of unity. Therefore, it is logical to introduce ordinary Euclidean distance between vectors. Let us consider the cloud of three-dimensional vectors *f* of general orthogonal factors. Inside the minimum parallelepiped containing the points *f* being classified, the centers of test clusters are randomly located, and the number $q \ge 2$ of such clusters is fixed. Let Γ denote the initial random position of test clusters. For the given arrangement of cen-

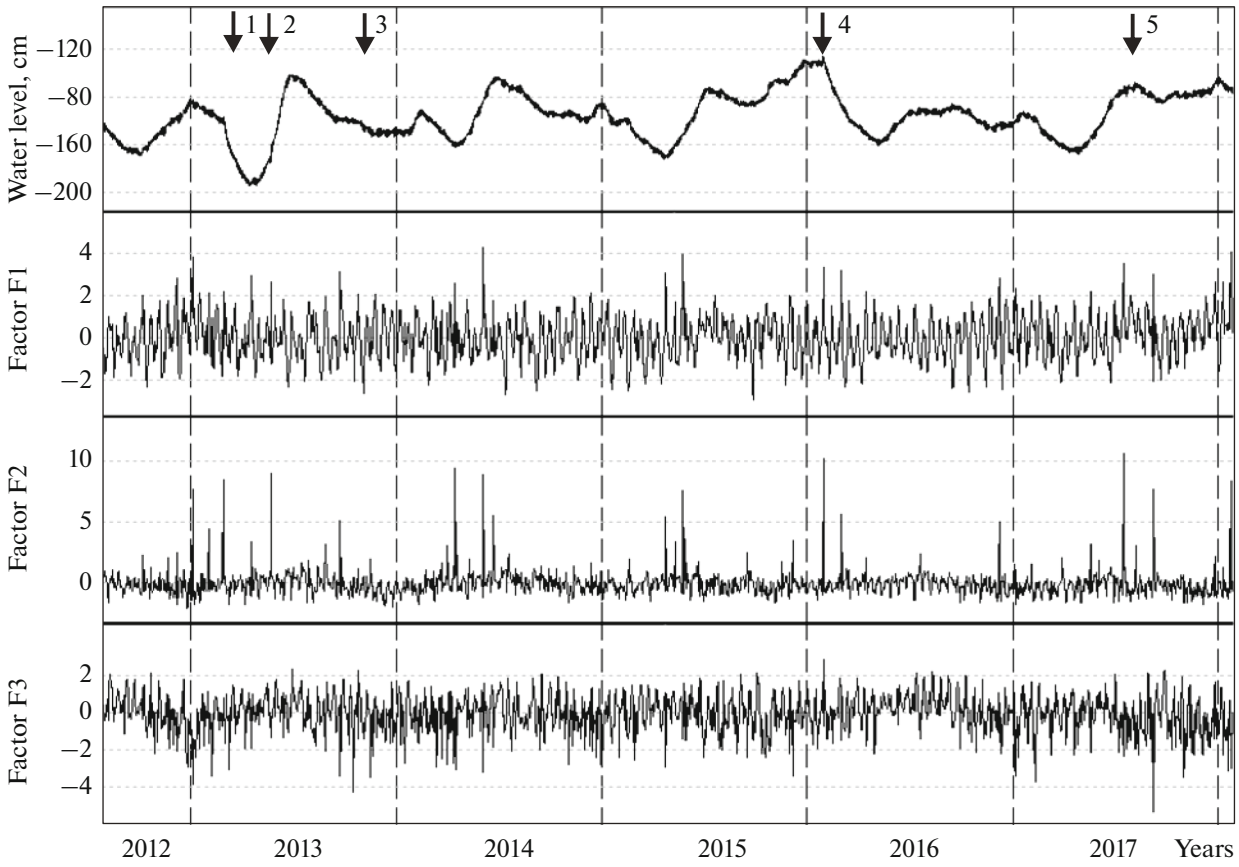

Fig. 5. Diagrams of orthogonal general factors F1, F2, and F3 for set of eight properties of water level time series with atmospheric pressure compensated, $U_k(t)$ (uppermost panel), in one-day-long sequential windows. Arrows denote earthquakes, numerals correspond to those in Table 1.

ters of clusters, test partitioning of the set of points is performed by the principle of minimum distance to some center. Let c_k with $k = 1,..., q$ be the vectors of the centers of clusters; n_k, the number of points in the kth cluster; and $\sum_{k=1}^{q} n_k = M$, the total number of points in the partitioned set.

In our case, $M = 2015$, which corresponds to the number of sequential time intervals that are one day long. Let B_k be the set of vectors belonging to the kth cluster. Let us calculate the vectors of the centers of gravity of the obtained clusters: $r_k = \sum_{f \in B_k} \xi / n_k$. If $c_k = r_k$ for all vectors, then partitioning ends; otherwise, the vectors of the centers of clusters c_k are shifted to centers of gravity r_k, another partitioning into clusters is done, new centers of gravity of clusters are found, the condition for the end of partitioning is checked, and so on. The procedure converges quickly; however, the partitioning of clusters obtained after all iterations depends on the random positions of centers of test clusters Γ before the start of iterations. The number of final partitioning is estimated by the cluster density criterion:

$$J(q|\Gamma) = \sum_{k=1}^{q} \sum_{f \in B_k} |f - c_k|^2. \quad (17)$$

For the set number of clusters q, we will find the random initial position of Γ for which the value of (17) is minimum. This is attained by the Monte Carlo method: random experiments on placing centers of test clusters within the limits of the cloud of points are repeated many times (below, when analyzing the correctness of data, there are 10^4 tests), and then the partitioning with the minimum Γ is chosen.

Next, we solve the problem of determining the optimal number of clusters into which the set of attributes should be partitioned. Let $J_0(q) = \min_{\Gamma} J(q|\Gamma)$. If we sequentially reduce the number of test clusters q from some quite large value to the minimum $q = 2$, then $J_0(q)$ will decrease monotonically; however, for the optimal value of clusters (if it exists), it will be disrupted. A more effective method of finding the optimal number of clusters is borrowed from dispersion analysis and implies that pseudo-F-statistics are used (Vogel and Wong, 1979):

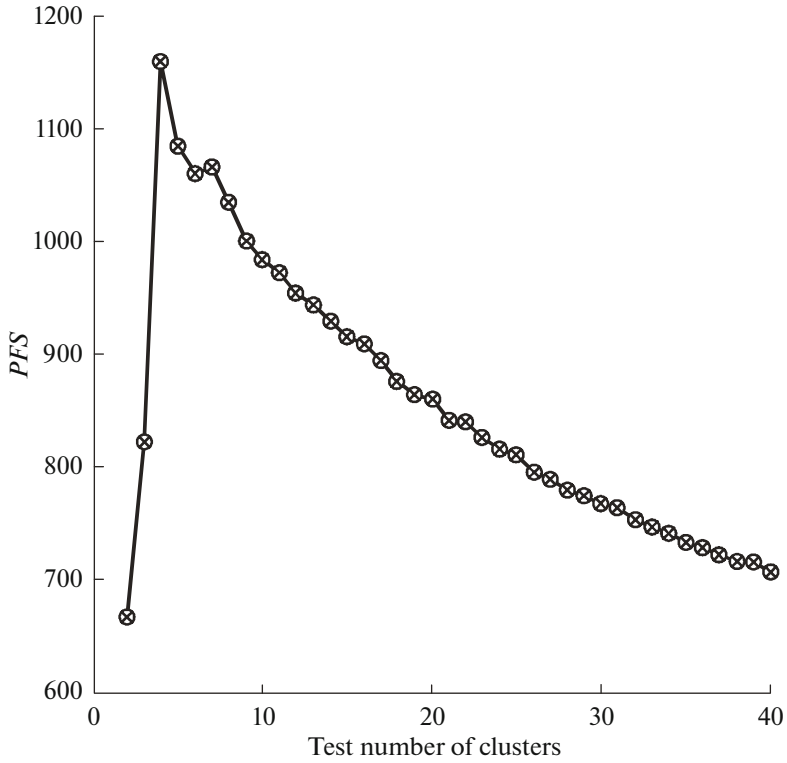

Fig. 6. Diagram of pseudo-F-statistics for clustering of three orthogonal general factors F1, F2, and F3.

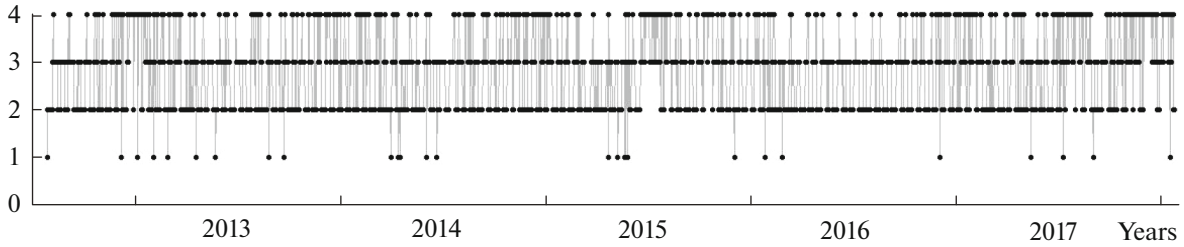

Fig. 7. Sequential transitions between four clusters of statistical properties of one-day-long fragments of time series of water level increments, $U_k(t)$, in YuZ-5 well.

$$PFS(q) = (M - q)\sum_{k=1}^{q} n_k \left|c_k - r_0\right|^2 \Big/ ((q-1)J_0(q)), \quad (18)$$

where $r_0 = \sum f \Big/ M$ is the common center of gravity of the entire set of points being classified. The optimal number of clusters corresponds to the point where function (18) is maximum.

Figure 6 shows the dependence of pseudo-F-statistics on the number of test clusters; it is seen that the optimal number of clusters in the space of orthogonal general factors F1, F2, and F3 is four.

Figure 7 shows the sequence of transitions between the four distinguished states of the $U_k(t)$ time series. Out of 2015 values (day), state 1 distinguished by cluster 1 was manifested 27 times; state 2, 810 times; state 3, 722 times; and state 4, 456 times. Thus, the probability of finding the time series in every distinguished state (Fig. 7) is 0.013, 0.402, 0.358, and 0.226, respectively. Hence, state 1, distinguished by cluster 1, can be considered anomalous, whereas the rest of the three states distinguished by clusters 2–4 can be considered background states.

GEOPHYSICAL INTERPRETATION OF THE STATISTICAL ANALYSIS RESULTS

As a result of the processing of water level data discussed above, the time series of statistical parameters of the time series $U_k(t)$ were obtained, including the set

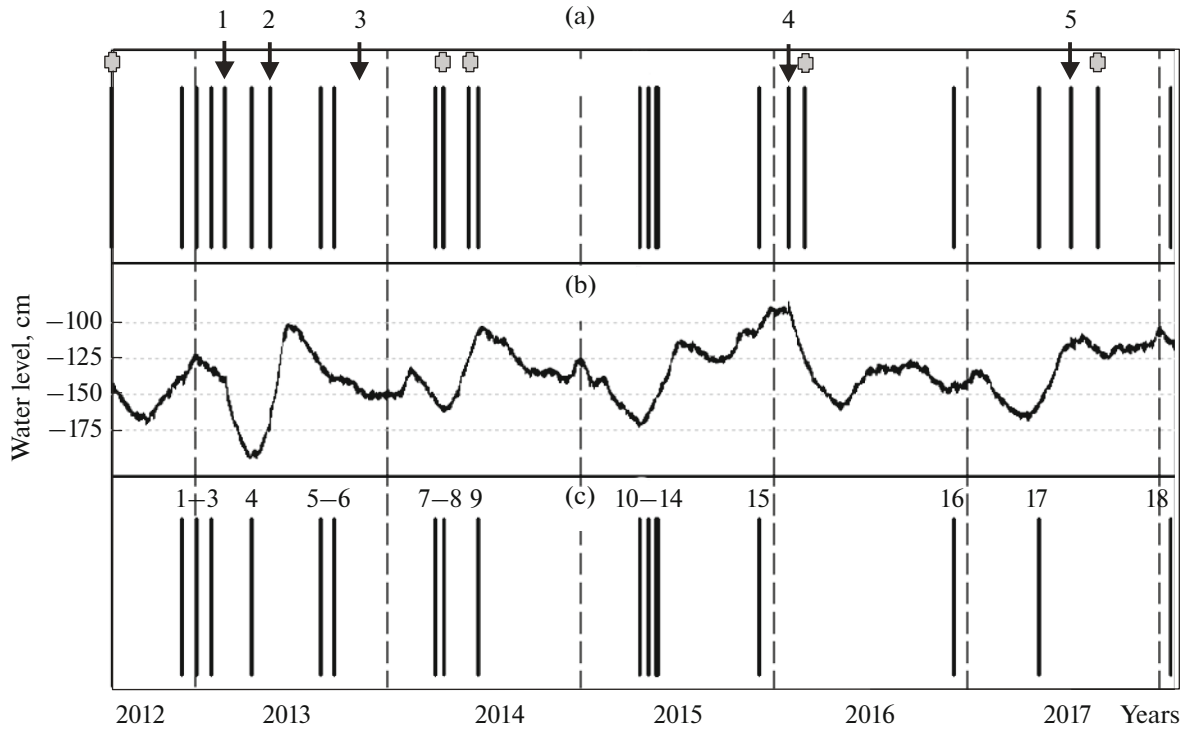

Fig. 8. Time distribution of (a) cluster 1 in comparison with (b) $U_k(t)$ time series, strong earthquakes (indicated with arrows, numerals correspond to those in Table 1), and dates of maintenance works with maintenance to well (grey crosses). Explanations on panel (c) are in text.

of eight properties of time series (Fig. 4), three orthogonal general factors F1, F2, and F3 (Fig. 5), and manifestations of four clusters of statistical properties of one-day-long fragments (Fig. 7). The sampling rate of all statistical parameters is one day for a time series length of 2015 days.

Further analysis of the obtained statistical parameters was aimed at assessing their sensitivity to various natural and technogenic effects on the hydrogeodynamic regime of the well. The main focus was to analyze the relationship between manifestations of anomalous cluster 1 (Fig. 7) and disturbed states of the studied object. Dates of disturbances in hydrogeodynamic regime due to coseismic variations of water level during strong earthquakes (Table 1) and five cases of methodical works carried out, associated to the maintenance to the well shaft, are known and their total number is ten (Fig. 8).

Out of five dates of strong earthquakes (Table 1), four cases (all earthquakes except for no. 3) were marked by the display of cluster 1 (Fig. 8a), whereas the earthquake of November 12, 2013 was not accompanied by manifestation of cluster 1. Note that this earthquake had the minimum magnitude and its intensity shaking in the area of the well was I = III–IV, coseismic and postseismic effects did not manifest in 5-min variations of water level.

In all five cases of technical operations associated with maintenance, cluster 1 was manifested. Thus, out of ten known cases of a disturbed hydrogeodynamic regime of the well, nine cases (90%) were characterized by anomalous cluster 1, indicating a sufficient sensitivity of the considered method of processing of the water level data in order to diagnose short-term disturbances in hydrogeodynamic regime of the YuZ-5 well.

In cases of perceptible local and strong distant earthquakes, durations of coseismic and postseismic effects in water level variations are usually no longer than a few tens of minutes or a few hours at maximum. After these earthquakes, the hydrogeodynamic regime of the well is restored relatively quickly because of the quite high water permeability of water-bearing rocks (7.8 m^2/day). For the same reason, restoration of the water level disturbed during technical maintenance is also restored in no more than 6–8 h. Thus, the one-day-long interval of assessing the statistical properties of the studied time series is sufficient for diagnosing relatively short-term disturbances of the hydrogeodynamic regime of the well.

In cases of the three strongest local earthquakes accompanied by shaking of I = V or more on the MSK-64 scale, postseismic variations of water level evolved for as long as several months, e.g., the drop of water level after the earthquake of March 28, 2013,

lasted for a month and a half (Fig. 2b). The considered method of processing water level data does not allow us to distinguish hydrogeoseismic variations of that length. They are diagnosed by the traditional method of finding low-frequency trends in water level changes after compensation for atmospheric pressure variations and suppressing diurnal and semidiurnal tidal variations of water level (Kopylova, 2001, 2006a, 2009).

If we exclude from consideration nine cases when cluster 1 was manifested due to disturbance of hydrogeodynamic regime of the well during coseismic effects and technical maintenance (33%), there are still 18 other cases (67%) where it was manifested (Fig. 8c). We think that possible causes of cluster 1 manifestation may be short-term fluctuations of hydrodynamic pressure, water level variations caused by seismic waves traveling from strong distant earthquakes (hydroseisms), preparation of strong local earthquakes, extreme meteorological conditions (cyclones, strong gusts of wind, abrupt changes in atmospheric pressure), and technogenic effects on the regime of the well located in the residential zone of the town of Yelizovo, near its airport and highway.

Fluctuations of the Hydrostatic Pressure Head

The eight cases of manifestation of cluster 1 (vertical lines 7–9 and 10–14 in the lower panel of Fig. 8) fall in time intervals from April to May in 2014 and 2015. These were years when no strong earthquakes occurred in Kamchatka and normal seasonal variations of water level were observed. This shows that anomalous cluster 1 can be related to the processes forming groundwater pressure. Earlier, peculiarities in the seasonal formation of water pressure in the area of the YuZ-5 well were considered in (Boldina and Kopylova, 2017), where the intraannual curve of changes of hydrostatic pressure based on interannual observations was provided. In accordance with this curve, the period of April–May is when the minimum water pressure transits to growth due to the change of the phase characterized by predominant run-off in depth (seasonal drop of water level) to the phase characterized by infiltration recharge and growth of pressure (increase in water level). The most probable cause of the appearance of cluster 1 in April–May is short-term variations in water pressure due to infiltration recharge of groundwaters and local effects of groundwater flow between particular layers of water-bearing rocks with different filtering properties.

Hydroseisms

Variations of water level during the passage of seismic waves from strong distant earthquakes can also cause manifestation of cluster 1. For example, cluster 1, numbered 18 in Fig. 8c, coincides with the date (January 23, 2018) of the earthquake in Alaska ($M_w = 7.9$, according to the USGS), which occurred at an epicentral distance of $d_e = 3400$ km. This earthquake was accompanied by water level oscillations in the YuZ-5 well (7 cm in amplitude for a duration of 15 min).

According to the USGS, there were other strong earthquakes that occurred in the world during the observations at the YuZ-5 well: April 1, 2014, Chile, $M_w = 8.2$, $d_e = 13\,300$ km; April 25, 2015, Nepal, $M_w = 7.8$, $d_e = 6100$ km; September 16, 2015, Chile, $M_w = 8.3$, $d_e = 14\,600$ km; and September 8, 2017, Mexico, $M_w = 8.1$, $d_e = 7400$ km. Seismic waves from these earthquakes were accompanied by weak variations of water level of the YuZ-5 well for 1–6 h, with amplitudes ranging from 0.4 to 2 cm. Notably, cluster 1 was not manifested on the days of these earthquakes.

Influence of Earthquake Preparation

The influence of processes related to earthquake preparation as a possible cause of anomalous states in the well regime and manifestation of cluster 1 was considered by comparison of the time when strong local earthquakes occurred (Table 1) and manifestations of cluster 1 before them in the period of up to several months. Emphasize that nine cases when cluster 1 was related to coseismic and known technogenic effects were excluded from consideration.

We assumed that anomalous cluster 1 could have been manifested in a period of up to several months before earthquakes (Table 1 and arrows in Fig. 8a) due to short-term disturbances of the hydrodynamic regime of the well owing to preparation of seismic events. This assumption is based on spatiotemporal regularities of how hydrogeological (hydrogeodynamic and hydrogeochemical) precursors were manifested before the strong earthquakes in Kamchatka region, as revealed from observations of groundwater parameters in deep wells of the Petropavlovsk Geodynamic Test Area. It was shown in (Kopylova, 2006a, 2006b) that precursors related to the water level and chemical composition of groundwater are observed before earthquakes with $M_w = 6.6–7.8$ at epicentral distances of up to a few hundred kilometers over a few weeks to nine months. For example, before the Zhupanovo earthquake, a hydrogeodynamic precursor was observed in well YuZ-5 over the course of 3.5 months (Boldina and Kopylova, 2017).

Taking into account such factors as the sensitivity of cluster 1 to short-term disturbances of the hydrodynamic regime of the well and general regularities of the relationship between hydrogeological precursors and strong Kamchatka earthquakes, we can think that nine cases when cluster 1 manifested (nos. 1–3, 4, 5–6, 15, and 16–17) could be related to the earthquake preparation processes from Table 1. The forecast time interval of cluster 1 manifestation before earthquakes

(Table 1) was from 25 to 220 days, 72 days or 2.4 months on average (Fig. 8c).

Brief analysis of possible causes of manifestation of cluster 1 shows that it may be related to both seismic and other natural and technogenic factors. This should be taken into consideration when applying the considered method of statistical processing of water level data in problems of geophysical monitoring and searching for earthquake precursors.

CONCLUSIONS

The presented method of statistical processing of water level data implies detection of an anomalous cluster of three orthogonal factors in variations of the set including eight statistical parameters in one-day-long intervals of 5-min series of water level variations after compensation for atmospheric pressure influence and allows short-term disturbances of hydrogeodynamic regime of the studied well to be diagnosed. This method significantly augments the traditional approach, which entails processing of water level data with the identification of a low-frequency trend in water level changes and the respective low-frequency signals in changes of groundwater pressure.

The test of sensitivity of the method for distinguishing short-term disturbances of the hydrogeodynamic regime of a well using known coseismic and technogenic effects indicates its sufficient reliability. The majority (nine out of ten) of these disturbances were diagnosed by the appearance of an anomalous cluster; i.e., these disturbances were distinguished using the considered experimental data processing method. This shows that the proposed method can be applied to geophysical monitoring in the Kamchatka region to identify short-term disturbances of the hydrogeodynamic regime of the YuZ-5 and other monitoring wells, including the possible effects in water level changes at the preparation stage of strong earthquakes. We emphasize that the important conditions for applying the method are the continuous character of observation data, high quality of initial data, and application of a procedure to compensate for atmospheric pressure variations in water level changes.

It has been found that a considerable fraction (67%) of cases when anomalous cluster was manifested had no clear relationship to either seismic or technogenic processes, and this should be taken into consideration when applying the considered method in problems of geophysical monitoring and searching for earthquake precursors. Nevertheless, the undoubted advantage of the proposed method is its ability to reveal noisy short-term anomalous states of observation wells in the real-time processing of large datasets of water level measurements, which is not provided by traditional methods.

The problem on the optimal choice of statistical parameters characterizing the properties of $U_k(t)$ time series within sequential one-day-long time fragments—i.e., the dimension and composition of parameter vector ζ in formula (9)—remains open. To verify the applicability of the method for distinguishing short-term disturbances of the well regime, we carried out experiments with different sets of statistics, from eight (present work) to 12, with inclusion of one to four time series of additional parameters in ζ and all other conditions of calculations remaining the same. In all variants of these calculations, we obtained the same results: three general orthogonal factors describing the used statistical sets are distinguished, and then four clusters are obtained in the space of orthogonal general factors; one of these clusters is manifested with a low probability and could be considered anomalous.

FUNDING

The work was supported by the Russian Foundation for Basic Research (project nos. 18-05-00133 and 18-05-00337).

CONFLICT OF INTERESTS

The authors declare they have no conflict of interests.

REFERENCES

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

Anderson, T.W. and Rubin, H., Statistical inference in factor analysis, in *Proceedings of 3rd Berkley Symposium on Mathematical Statistics and Probability*, 1956, vol. 5, pp. 111–150.

Boldina, S.V. and Kopylova, G.N., Coseismic effects of the 2013 strong Kamchatka earthquakes in well YuZ-5, *Vestn. KRAUNTs. Nauki Zemle*, 2016, vol. 30, no. 2, pp. 66–76.

Boldina, S.V. and Kopylova, G.N., Effects of the January 30, 2016, $M_w = 7.2$ Zhupanovsky earthquake on the water level variations in wells YuZ-5 and E-1 in Kamchatka, Kamchatka, *Geodin. Tektonofiz.*, 2017, vol. 8, no. 4, pp. 863–880.
https://doi.org/10.5800/GT-2017-8-4-0321

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

Bredehoeft, J.D., Response of well-aquifer systems to earth tides, *J. Geophys. Res.*, 1967. vol. 72, no. 12, pp. 3075–3087.

Chebrov, V.N., Kugaenko, Yu.A., Abubakirov, I.R., Droznina, S.Ya., Ivanova, E.I., Matveenko, E.A., Mityushkina, S.V., Ototyuk, D.A., Pavlov, V.M., Raevskaya, A.A., Saltykov, V.A., Senyukov, S.L., Serafimova, Yu.K., Skorkina, A.A., Titkov, N.N., and Chebrov, D.V., The January 30th, 2016 earthquake with $K_s = 15.7$, $M_w = 7.2$, $I = 6$ in the Zhupanovsky region (Kamchatka), *Vestn. KRAUNTs. Nauki Zemle*, 2016, vol. 29, no. 1, pp. 5–16.

Chebrov, V.N., Saltykov, V.A., and Serafimova, Yu.K., *Prognozirovanie zemletryasenii na Kamchatke. Po materialam raboty Kamchatskogo filiala Rossiiskogo ekspertnogo*

soveta po prognozu zemletryasenii, ocenke seismicheskoi opasnosti i riska v 1998–2009 gg. (Prediction of Earthquakes in Kamchatka Based on the Activities of the Kamchatka Branch of the Russian Expert Council for Earthquake Prediction and Seismic Risk and Hazard Assessment in 1998–2009), Moscow: Svetoch Plyus, 2011.

Chebrov, D.V., Kugaenko, Yu.A., Abubakirov, I.R., Lander, A.V., Pavlov, V.M., Saltykov, V.A., and Titkov, N.N., The July 17th, 2017 Nizhne-Aleutian earthquake with M_w = 7.8 on the border of the Komandor seismic gap (western part of the Aleutian Arc), *Vestn. KRAUNTs. Nauki Zemle*, 2017, vol. 35, no. 3, pp. 22–25.

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.

Firstov, P.P., Kopylova, G.N., Solomatin, A.V., and Serafimova, Yu.K., Strong earthquake forecast near the Kamchatka peninsula, *Vestn. KRAUNTs. Nauki Zemle*, 2016, vol. 32, no. 4, pp. 106–114.

Harman, H.H., *Modern Factor Analysis*, Chicago: Univ. Chicago Press, 1967, 2nd ed.

Huber, P.J. and Ronchetti, E.M., *Robust Statistics*, New York: Wiley, 2009, 2nd ed., ch. 1.
https://doi.org/10.1002/9780470434697.ch1

Igarashi, G. and Wakita, H., Tidal responses and earthquake-related changes in the water level of deep wells, *J. Geophys. Res., [Solid Earth Planets]*, 1991, vol. 96, pp. 4269–4278.

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

Kissin, I.G., Hydrogeological monitoring of the Earth's crust, *Fiz. Zemli*, 1993, no. 8, pp. 58–69.

Kissin, I.G., *Flyuidy v zemnoi kore: geofizicheskie i tektonicheskie aspekty* (Fluids in the Earth's Crust: Geophysical and Tectonic Aspects), Moscow: Nauka, 2009.

Kopylova, G.N., Variations of water level in Elizovskaya-1 well, Kamchatka, due to large earthquake: 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.*, 2006a, no. 6, pp. 52–64.

Kopylova, G.N., Seismicity as a factor of setting the regime of ground water level rise, *Vestn. KRAUNTs. Nauki Zemle* 2006b, vol. 7, no. 1, pp. 50–66.

Kopylova, G.N., The application of water level observations in wells for searching earthquakes precursors (on the example of Kamchatka), *Geofiz. Issled.*, 2009, vol. 10, no. 2, pp. 56–68.

Kopylova, G.N. and Boldina, S.V., Estimation of poreelastic parameters for a reservoir of ground water (based on water level observation at YuZ-5 well, Kamchatka), *Vulkanol. Seismol.*, 2006, no. 2, pp. 17–28.

Kopylova, G.N. and Boldina, S.V., On the mechanism of hydrogeodynamic precursor of the Kronoki earthquake of December 5, 1997, M_w = 7.8, *Tikhookean. Geol.*, 2012, vol. 31, no. 5, pp. 104–114.

Kopylova, G.N., Boldina, S.V., Smirnov, A.A., and Chubarova, E.G., Experience in registration of variations caused by strong earthquakes in the level and physicochemical parameters of ground waters in the piezometric wells: the case of Kamchatka, *Sesim. Instrum.*, 2016, vol. 53, no. 4, pp. 286–295.
https://doi.org/10.3103/S0747923917040065

Kopylova, G.N., Lyubushin, A.A., Malugin, V.A., Smirnov, A.A., and Taranova, L.N., Observations at the Petropavlovsk Test Site, Kamchatka, *Volcanol. Seismol.*, 2001, vol. 22, no. 4, pp. 453–468.

Kopylova, G.N., Steblov, G.M., Boldina, S.V., and Sdel'nikova, I.A., The possibility of estimating the coseismic deformation from water level observations in wells, *Izv., Phys. Solid Earth*, 2010, vol. 46, no. 1, pp. 47–56.
https://doi.org/10.1134/S1069351310010040

Lawley D.N. and Maxwell, A.E., *Factor Analysis as a Statistical Method*, London: Butterworth, 1971, 2nd ed.

Lyubushin, A.A., *Analiz dannykh sistem geofizicheskogo i ekologicheskogo monitoringa* (Analyzing Data from Systems 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., Prognostic properties of low-frequency seismic noise, *Nat. Sci.* 2012, vol. 4, no. 8A, pp. 659–666.
https://doi.org/10.4236/ns.2012.428087

Lyubushin, A.A., Prognostic properties of stochastic variations in geophysical parameters, *Biosfera*, 2014, no. 4, pp. 319–338.

Lyubushin, A.A., Jr. and Lezhnev, M.Yu., Variation in the response function of the groundwater table with atmospheric pressure in the Southern Kuriles, Shikotan Island, *Phys. Solid Earth*, 1995, vol. 31, no. 8, pp. 710–715.

Lyubushin, A.A. and Malugin, V.A., Statistical analysis of ground water level response to atmospheric pressure variations, *Fiz. Zemli*, 1993, no. 12, pp. 74–80.

Lyubushin, A.A., Jr., 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. 302–313.

Lyubushin, A.A., Jr., 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.

Lyubushin, A.A. and Farkov, Yu.A., Synchronous components of financial time series, *Komp'yut. Issled. Modelir.*, 2017, vol. 9, no. 4, pp. 639–655.
https://doi.org/10.20537/2076-7633-2017-9-4-639-655

Mallat, S., *A Wavelet Tour on Signal Processing*, San Diego: Academic Press, 1999.

Medvedev, S.V., Shponkhoier, V., and Karnik, V., *Shkala seismicheskoi intensivnosti MSK-64* (MSK-64 Scale of Seismic Intensity), Moscow: Mezhved. Geofiz. Kom Akad. Nauk SSSR, 1965.

Melchior, P., *Earth Tides*, London: Pergamon, 1966.

Osorio, I., Lyubushin, A., and Sornette, D., Automated seizure detection: Unrecognized challenges, unexpected insights, *Epilepsy Behav.*, 2011, vol. 22, no. 1, pp. S7–S17. https://doi.org/10.1016/j.yebeh.2011.09.011

Roeloffs, E.A., Hydrologic precursors to earthquakes: A review, *Pure Appl. Geophys.*, 1988, vol. 126, pp. 177–209.

Roeloffs, E.A., Burford, S.S., Riley, F.S., and Records, A.W., Hydrologic effects on water level changes associated with episodic fault creep near Parkfield, California, *J. Geophys. Res., [Solid Earth Planets]*, 1989, vol. 94, pp. 12387–12402.

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

Sil'nye kamchatskie zemletryaseniya 2013 goda (Strong Earthquakes of Kamchatka in 2013), Chebrov, V.N., Ed., Petropavlovsk-Kamchatsky: Novaya kniga, 2014.

Vinogradov, E.A., Gorbunova, E.M., Kabychenko, N.V., Kocharyan, G.G., Pavlov, D.V., and Svintsov, I.S., Monitoring of the ground water level based on the data of 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, no. 3, pp. 237–245. https://doi.org/10.1109/TPAMI.1979.4766919

Wang, C.Y. and Manga, M., *Earthquakes and Water*, vol. 114 of *Lecture Notes in Earth Sciences*, Berlin: Springer, 2010. https://doi.org/10.1007/978-3-642-00810-8

Translated by N. Astafiev