Izvestiya, Physics of the Solid Earth, Vol. 34, No. 3, 1998, pp. 238–243. Translated from Fizika Zemli, No. 3, 1998, pp. 69–74.

SCIENTIFIC COMMUNICATIONS

An Aggregated Signal of Low-Frequency Geophysical Monitoring Systems

A. A. Lyubushin, Jr.

Institute of Experimental Geophysics, Schmidt Joint Institute of Physics of the Earth, Russian Academy of Sciences, ul. B. Gruzinskaya 10, Moscow, 123810 Russia

Received December 13, 1995

Lyubushin [1, 2] proposed an approach to the combined analysis of long-term routine observations of the background variations in geophysical parameters (crustal deformations, groundwater levels in wells, groundwater concentrations of chemical elements, gas emanation intensity, etc.) measured synchronously at monitoring network stations. The monitoring is aimed at the detection of the so-called *synchronization signal* in these variations (possibly differing in their physical origin), as well as its duration and frequency range. The algorithm of the method consists in the time evolution estimation of the spectral matrix eigenvalue in a moving time window and in fact is a multidimensional extension of the 1D time–frequency analysis widely used in geophysics. The synchronization signal is associated with bursts (increases) of the maximum eigenvalue. Supposedly, such a signal may be a precursor of strong earthquakes [1–4].

This method is developed for the purpose of the construction and subsequent analysis of the so-called *aggregated signal* of geophysical monitoring systems. This signal is represented by a scalar time series which includes the maximum number of features common to all components of the multidimensional time series studied, and no feature characteristic of only one component is involved in this signal.

The efficiency of the method is tested on a four-dimensional time series of Cl^- concentration variations measured by the system of wells and springs of the Petropavlovsk, Kamchatka, geodynamic research site.

INTRODUCTION

The goal of this paper is the development of an algorithm that constructs an aggregated signal (scalar time series) from a multidimensional time series of observations. The spectrum of this signal includes features common to all of the series analyzed, whereas the signals specific only to one series are maximally suppressed. The original multidimensional series can be represented by a set of scalar time series obtained from the measurement of either one characteristic at different stations of a geophysical monitoring network or various characteristics at one station. Various combinations of these two cases are also possible. Importantly, the scalar components of the original multidimensional time series may differ in their physical dimensions and scales. Their general aspect is the synchronism of measurements and *a priori* assumption that there are features common to all of them which reflect universal factors affecting the state of physical fields in the crust, e.g., strong earthquake preparation processes. An important problem arises concerning the method by which these general factors and tendencies can be recognized against the background of noisy signals of either local or external (with respect to the crust), often anthropogenic origin. Moreover, as compared with the noise, the desired signal may have a comparable or even lower amplitude and a similar frequency range. If external noise, like the atmospheric pressure and temperature variations, can be measured, its effect can be eliminated using the multivariate noise suppression algorithm [1, 5]. The problem of anthropogenic or local noise (i.e., caused by either the station siting conditions or specific features of the process measured) is more complicated, because such noise is commonly inaccessible to direct measurement. The only way to eliminate this noise is the use of either the observations at other stations or physically different processes, which are also affected by the general factors but have their own, uncorrelated local and anthropogenic noise properties. The same ideas are employed in the processing technique of seismic signals recorded by a spread, involving the construction of the beam-forming or Capon filters to define the arrival direction [6–8]. However the wave terminology cannot be used in the processing of the data analyzed in this paper (low-frequency time series of observations), which invalidates the application of the classical methods of general signal recognition, developed in seismology.

In addition to general signal recognition, the construction of the aggregated signal aims at simplifying the comparable analysis of data flows recorded by several different monitoring systems. For example, consider two observation networks, each covering its own region and forming a ten-dimensional series of combined observations. If the problem of interaction between the geophysical fields of these regions is to be solved, the aggregated signals constructed for each monitoring network reduces the problem to the analysis of the interrelation between two scalar series. Otherwise, a 20-dimensional series should be studied.

ALGORITHM OF THE AGGREGATED SIGNAL CONSTRUCTION

We will consider an m-dimensional ($m \geq 3$) time series

$$\mathbf{Z}(t) = (Z_1(t), \ldots, Z_m(t))^T, \quad (1)$$

where t is the discrete time corresponding to consecutive measurements, $t = 1, \ldots, N$, and the superscript "T" implies transposition. For convenience, the first component $Z_1(t)$ will be temporarily denoted as $x(t)$, so that the remaining components of the series $\mathbf{Z}(t)$ form the $(m-1)$-dimensional temporal series $\mathbf{Y}(t) = (Z_2(t), \ldots, Z_m(t))^T$.

We want to find the spectral correlated signals that are common to both series $\mathbf{Y}(t)$ and $x(t)$. For this purpose, an $(m-1)$-dimensional filtering procedure is applied to $\mathbf{Y}(t)$, resulting in scalar series $\eta(t)$. The algorithmic description of the $(m-1)$-dimensional filtering may be represented as a sequence of the following operations. First, we obtain the discrete Fourier transform of $\mathbf{Y}(t)$, denoted by $\tilde{\mathbf{Y}}(\omega)$, where ω are discrete frequencies. Then, the $(m-1)$-dimensional vector $\tilde{\mathbf{Y}}(\omega)$ is scalarly multiplied by $(m-1)$-dimensional complex vector $\mathbf{B}(\omega)$ which is the transfer function of the $(m-1)$-dimensional filter $\tilde{\eta}(\omega) = \tilde{\mathbf{Y}}^H(\omega)\mathbf{B}(\omega)$, where "$H$" indicates an Hermitean transposed matrix. Finally, an inverse discrete Fourier transform, with its real part denoted as $\eta(t)$, is obtained from the one-dimensional series of complex numbers $\tilde{\eta}(\omega)$.

The multidimensional filter $\mathbf{B}(\omega)$ is chosen so that the squared modulus of the coherence between the series $x(t)$ and $\eta(t)$

$$\rho^2_{x\eta}(\omega) = |S_{xY}\mathbf{B}|^2/(S_{xx}\mathbf{B}^H S_{YY}\mathbf{B}) \quad (2)$$

is maximal at each frequency ω. Here, S_{xx} is the power spectrum of the scalar series $x(t)$, the $(m-1)$-dimensional complex row-vector S_{xY} is the cross-spectrum between $x(t)$ and $\mathbf{Y}(t)$, and S_{YY} is the spectral, $(m-1) \times (m-1)$-matrix of the time series $\mathbf{Y}(t)$. The maximization problem (2) is solved by the Lagrange method of multipliers, with vector $\mathbf{B}$ being normalized, $\mathbf{B}^H\mathbf{B} = 1$. After simple manipulations, we find that the sought-for vector filter $\mathbf{B}(\omega)$ should be the eigenvector of the $(m-1) \times (m-1)$-matrix

$$S^{-1}_{YY}S_{Yx}S^{-1}_{xx}S_{xY}, \quad (3)$$

which corresponds to its maximum eigenvalue. The latter is exactly the maximum squared coherence (2) [9]. Once the filter $\mathbf{B}(\omega)$ is calculated, the sought-for series $\eta(t)$ can be found. Since $x(t)$ was chosen to be the first component $Z_1(t)$, this series is denoted by $\eta^{(1)}(t)$. The series $\eta^{(1)}(t)$ is the first canonical component of the multidimensional series $\mathbf{Y}(t) = (Z_2(t), \ldots, Z_m(t))^T$ with respect to $x(t) = Z_1(t)$ [9]. For shortness and convenience, we call it the canonical component of the series $Z_1(t)$. Thus, applying the filtering to all series and taking the series $Z_1(t)$ as a "standard" one, which should be approached as close as possible (in the sense of closeness measure (2)), we obtain a new series $\eta^{(1)}(t)$. Being free from the noise signals specific only to $Z_1(t)$ and absent in other series $Z_i(t)$, the series $\eta^{(1)}(t)$ retains the signals common to all series. Repeatedly applying this procedure to other components, the original time series $\mathbf{Z}(t) = (Z_1(t), \ldots, Z_m(t))^T$ is transformed to the series of canonical components $\mathbf{G}(t) = (\eta^{(1)}(t), \ldots, \eta^{(m)}(t))^T$, each $\eta^{(i)}$ being free from local and individual noise signals characteristic only of Z_i.

The next step is naturally the construction of the scalar series $\zeta(t)$ from the multidimensional series $\mathbf{G}(t)$, providing maximum information on the general behavior of the components $\eta^{(i)}$. For Gaussian time series, this is equivalent to multidimensional linear filtering that, when applied to the series $\mathbf{G}(t)$, would yield the scalar series $\zeta(t)$ whose power is maximal at each discrete frequency ω [9]. Like the method described above for the calculation of the canonical components, the calculation of $\zeta(t)$ may be represented as the following sequence of operations. Let $\tilde{\mathbf{G}}(\omega)$ be the m-dimensional discrete Fourier transform of $\mathbf{G}(t)$, and let $\mathbf{A}(\omega)$ be an as yet unknown frequency vector filter whose output is the 1D signal $\tilde{\zeta}(\omega) = \tilde{\mathbf{G}}^H(\omega)\mathbf{A}(\omega)$. The filter $\mathbf{A}(\omega)$ is chosen under condition that the power spectrum of the series ζ, $S_{\zeta\zeta} = \mathbf{A}^H S_{GG}\mathbf{A}$, where S_{GG} is the spectral $m \times m$-matrix of the series $\mathbf{G}(t)$, is maximal at each frequency ω. With the normalization $\mathbf{A}^H\mathbf{A} = 1$, this problem is readily solved: $\mathbf{A}(\omega)$ should be the eigenvector of the spectral matrix S_{GG}, corresponding to the maximum eigenvalue of this matrix. This solution is known as the first principal component of the series $\mathbf{G}(t)$, with the maximum eigenvalue of the matrix S_{GG} being the power spectrum of the series $\zeta(t)$, which is the real part of the inverse Fourier transform of the complex signal $\tilde{\zeta}(\omega)$ [9]. We call series $\zeta(t)$ the *aggregated signal* of the original time series $\mathbf{Z}(t)$. Thus, the aggregated signal is the first principal component of the multidimensional series of canonical components. Can $\zeta(t)$ immediately be defined as the principal component of the original series $\mathbf{Z}(t)$? The point is that, if some of series $Z_i(t)$ includes an intense, narrow-band noise signal, e.g., of anthropogenic origin, the construction of the general signal by projecting the multidimensional Fourier transform of the original series onto the principal eigenvector of the spectral matrix incompletely removes the noise signal from the general one (because such an approach tends to retain the maximum power of the general signal within each frequency band). On the other hand, this effect is efficiently eliminated by the two-step procedure described above (first, the construction of the multidimensional series of canonical components principally free from local noise and then the projection of their multidimensional Fourier transform onto the principal eigenvector of the spectral matrix).

Below, some computational aspects of the aggregated signal construction are described in more detail. First, since the time series analyzed are commonly dominated by low frequencies (the low-frequency power greatly exceeds that of higher-frequency components), we employ incremental series, rather than $Z_i(t)$, in order to remove the dominating effect of low-frequency components:

$$\mathbf{z}(t) = \mathbf{Z}(t+1) - \mathbf{Z}(t), \quad t = 1, \ldots, N-1. \tag{4}$$

Further, since the algorithm is designed to process physically diverse or scale-varying information, the sampling means s_i and dispersions σ_i are calculated for each scalar component $z_i(t)$ of the incremental series,

$$\begin{aligned} s_i &= \sum_{t=1}^{N-1} z_i(t)/(N-1), \\ \sigma_i^2 &= \sum_{t=1}^{N-1} (z_i(t) - s_i)^2/(N-2) \end{aligned} \tag{5}$$

and each component is normalized to the unit dispersion,

$$z_i(t) := (z_i(t) - s_i)/\sigma_i, \quad i = 1, \ldots, m. \tag{6}$$

Because of operation (6) and the construction technique, the resulting time series of canonical components $\eta^{(i)}(t)$ and aggregated signal $\zeta(t)$ provide formal, mathematical constraints and, being physically dimensionless, are interesting only from the standpoint of their behavior. Moreover, because the signals $\eta^{(i)}(t)$ and $\zeta(t)$ are also represented by incremental series due to transition (4), the visual comparison of their behavior with the original series $Z_i(t)$ requires them to be integrated, starting from an arbitrary initial value, e.g., zero (each member of the integrated series is equal to the sum of the preceding members of the incremental series). The series obtained upon integrating the signals $\eta^{(i)}(t)$ and $\zeta(t)$ are denoted by $Y^{(i)}(t)$ and $Z_0(t)$, respectively, and will also be called the canonical components of the original series $Z_i(t)$ and aggregated signal of the multidimensional time series $\mathbf{Z}(t)$. Note also that $\eta^{(i)}(t)$ and $\zeta(t)$ are determined by projecting the multidimensional Fourier transforms onto the eigenvectors of corresponding matrices. Because the eigenvectors are determined up to sign, so are $\eta^{(i)}(t)$ and $\zeta(t)$. This ambiguity may induce one to turn one of the realizations $Y^{(i)}(t)$ or $Z_0(t)$ upside down if the behavior of the "overturned" realization appears to be more natural with respect to the original data. This can simply be done by the change of sign at the corresponding series $\eta^{(i)}(t)$ or $\zeta(t)$ before the integration.

The algorithm description is completed by the method for estimating the spectral matrices $S_{ZZ}(\omega)$ and $S_{GG}(\omega)$. The complex row (S_{xY}) and column (S_{Yx}) vectors from (3) obey the obvious relation $S_{Yx} = S_{xY}^H$ and are blocks of the matrix $S_{ZZ}(\omega)$. The spectral matrices can be estimated by the nonparametrical or parametrical methods. The former consists of the application of the discrete Fourier transformation to each scalar sample and the calculation of the periodogram matrix subsequently averaged at each discrete frequency ω over adjacent frequencies within a window [6, 9, 10]. The parametrical method of the vector autoregression model [6–8, 10] is preferable at small N (in particular, at the moving time window estimation of spectral matrices) [2–4]. Since the aggregated signal is constructed from either the whole sample or its part of a considerable length, the nonparametrical method, due to its higher structural stability, is preferable in this case. Below, this method is applied. The numerical algorithm is described in detail in [1].

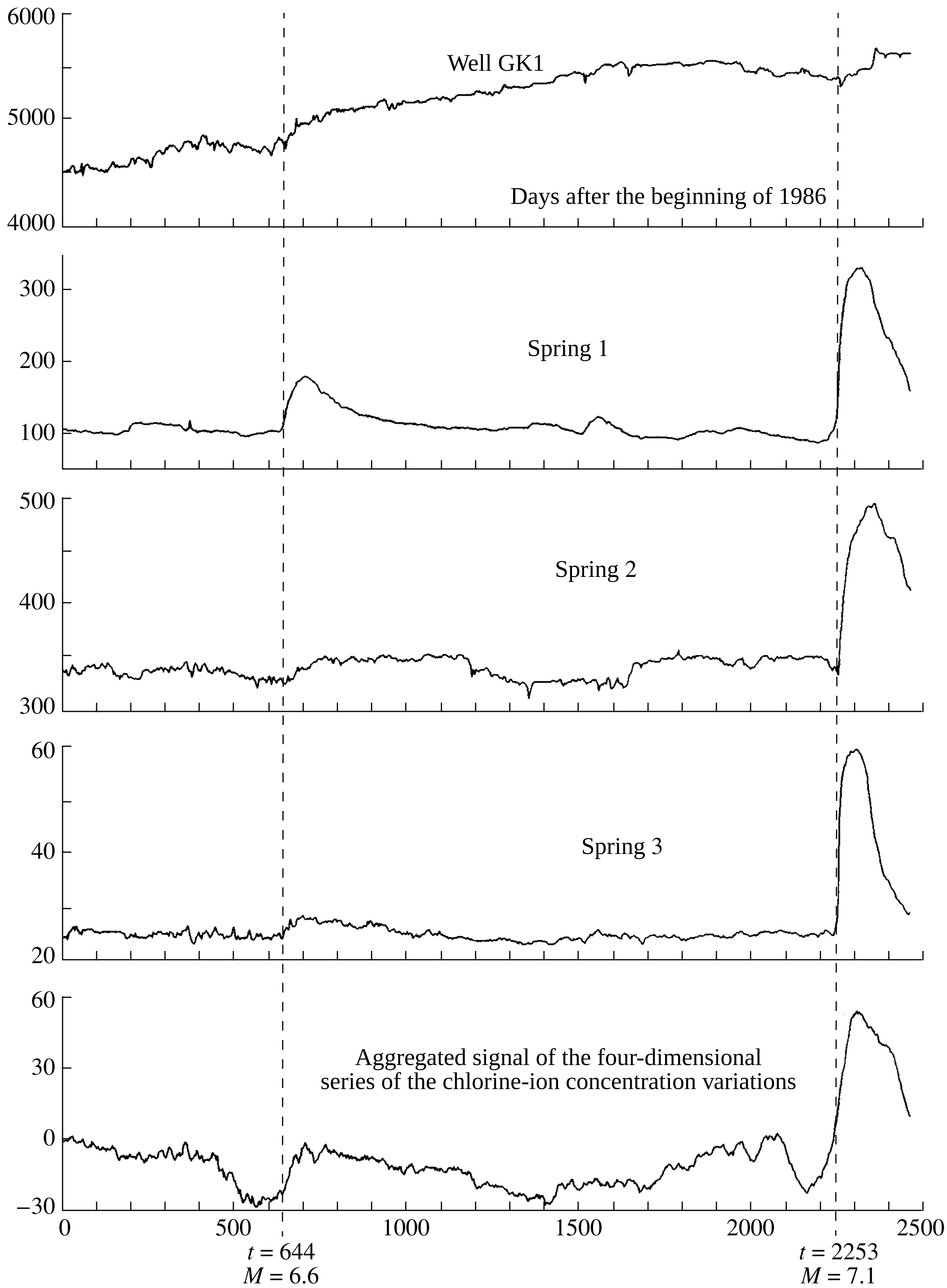
NUMERICAL EXAMPLE

Lyubushin [2] and Lyubushin *et al*. [3, 4] gave an exhaustive analysis of the maximum eigenvalue evolution diagrams of the spectral matrix for the time series of the variations in hydrogeochemical characteristics. The data provided by the Experimental Seismological Expedition, Institute of Volcanology (ESE IV), Far East Division, Russian Academy of Sciences, were obtained from the system of wells and overflowing springs of the Petropavlovsk, Kamchatka research site during the observation period from the beginning of 1986 through 1992; the measurements were made once every three days. The bursts in the maximum eigenvalue were compared with the seismic regime for various combinations of time series.

The table presents the data on the seismicity in the vicinity of the research site, the linear size of the observation network being about 50 km. The epicentral distances are given for the Pinachevo station chosen as the "center" of the site.

The figure shows the plots of the original time series of Cl^- concentrations from four wells and springs; the vertical broken lines indicate two stronger earthquakes with magnitudes of 6.6 and 7.1 (see table). The measurements were made every three days, from January 3, 1986 to the end of 1992, the length of the processed time series amounted to 821 observations.

As is seen from the figure, three realizations of the original data clearly exhibit the postseismic effects of two strong earthquakes. The uppermost plot shows marked concentration drops preceding the earthquakes,



Plots of the original time series of the Cl^- concentrations, observed in the groundwater from the wells and springs of the Petropavlovsk, Kamchatka, geodynamic research site (Pinachevo station) 1986–1992, and the aggregated signal constructed from these series. The vertical broken lines mark the moments of the two strongest earthquakes in the vicinity of the site.

against a background of the general linear trend. Similar minimums are recognizable from the other three plots of the original data, but they are strongly masked there by fluctuations. The lowermost plot represents the aggregated signal of the four-dimensional time series. The bay-like precursor anomalies are most clearly expressed in this plot, and moreover, they have a "classical" pattern: the process first reaches a minimum

Main evidence on the earthquakes of 1986–1992 (after the ESE IV data [3, 4]) that produced changes in the behavior of the observation springs and wells at the Petropavlovsk research site

No.	Date	Number of days after the beginning of 1986	Distance to the Pinachevo station	Magnitude	Depth, km
1	Jun. 17, 1986	168	160	5.0	40
2	Oct. 6, 1987	644	130	6.6	34
3	Sep. 15, 1989	1354	105	4.9	44
4	Mar. 1, 1990	1521	120	5.8	24
5	Dec. 19, 1990	1814	155	6.1	24
6	Apr. 8, 1991	1924	170	4.7	139
7	Mar. 2, 1992	2253	115	7.1	40

within the bay and then begins to increase, the earthquake occurring during the phase of its increase. Thus, the signal aggregation resulted in a considerable increase in the signal/noise ratio and made the precursor signals more contrasting.

DISCUSSION AND CONCLUSIONS

The algorithm proposed for the construction of a scalar (aggregated) signal accumulates, to the largest extent, the tendencies common to all processes recorded by the low-frequency geophysical monitoring systems and removes signals specific to only one process (individual and local noise).

The experience of the aggregated signal application to the processing of the hydrogeochemical observations in a seismically active region shows that this method considerably enhances the signal/noise ratio in the detection of precursor effects of strong earthquakes.

The application of the algorithm offered to the search for strong earthquake precursors is most intriguing. In fact, the example illustrated in figure is relevant to this problem. The method suggested in [1–2] for the precursor detection is based on the very general principles of the statistical physics of phase transitions and catastrophe theory. It is the synchronization and collective behavior of measured characteristics that are relevant to the problems of monitoring and preparation of an earthquake or other natural catastrophe. In this respect, certain methodological recommendations can be suggested which result from the most general regularities of the system behavior; we mean the regularities that draw a system nearer to a bifurcation, or catastrophe. An increase in the fluctuation correlation radius in the bifurcation vicinity indicates that the system tends to be self-consistent throughout its volume, thereby preparing for the collective transition to a new state [11, § 4.4]. In the statistical physics of fluids, such behavior is known as "critical opalescence" or abnormal dispersion, and is considered a universal signal of the approaching catastrophe [12, Section 9]. However, this idea becomes applicable only if it is implemented as a complex of numerical algorithms and programs that enable the broadband analysis of multidimensional time series whose scalar components are measurements of physically diverse geophysical values at spatially different stations of a monitoring network. This complex includes the aggregation signal construction proposed in this paper.

Of course, the construction of the aggregated signal does not completely exhaust the precursor search problem. Similar to the algorithms previously suggested by the author, this one should be considered a promising direction for the attempts to solve the prediction problem of a strong event in a given region. In particular, both the avalanche-unstable fracturing and dilatancy diffusion models [13] involve a general increase in the behavior of geophysical fields at stage II (appearance of long- and medium-term precursors) characterized by the collective orientation of cracks and occurrence of the dilatancy strengthening zone. The synchronization effects should decrease at stage III immediately preceding an earthquake occurrence and characterized by the destruction of bridges between cracks and their dehydration (the stage of occurrence of short-term precursors including the pulsed ones). Therefore, the algorithms offered in this paper and in [1, 2] should be applied only to the recognition of long- and medium-term precursors. However, the earthquake preparation model offered in [14] does not imply the existence of a general signal in the monitoring observations covering the preparation area.

ACKNOWLEDGMENTS

This work was supported by the International Association for the Promotion of Cooperation with Scientists from the Independent States of the Former Soviet Union, grant no. 94-0232.

REFERENCES

1. Lyubushin, A.A., Jr., Analysis of the Multiple Time Series of Geophysical Monitoring, *Fiz. Zemli*, 1993, no. 3, pp. 103–108.
2. Lyubushin, A.A., Jr., Classification of the States of Low-Frequency Geophysical Monitoring Systems, *Fiz. Zemli*, 1994, no. 7/8, pp. 135–141.
3. Lyubushin, A.A., Jr., Kopylova, G.N., and Khatkevich, Yu.M., Multivariate Approach to the Data Processing of the Hydrogeological Observations at the Petropavlovsk, Kamchatka, Research Site: Earthquake Precursor Study, *Vulkanol. Seismol.*, 1996, no. 1, pp. 79–97.
4. Lyubushin, A.A., Jr., Kopylova, G.N., and Khatkevich, Yu.M., Spectral Matrix Analysis of the Hydrogeological Observations at the Petropavlovsk, Kamchatka, Research Site: Comparison with the Seismic Regime, *Fiz. Zemli*, 1997, no. 6, pp. 79–89.

5. Lyubushin, A.A., Jr. and Latynina, L.A., Meteorological Noise Suppression in Strainmeter Observations, *Fiz. Zemli*, 1993, no. 3, pp. 98–102.
6. Dadgeon, D. and Mersereau, R., *Multidimensional Digital Signal Processing*, Englewood Cliffs: Prentice-Hall, 1984. Translated under the title *Tsifrovaya obrabotka mnogomernykh signalov*, Moscow: Mir, 1988.
7. Kushnir, A.F. and Lapshin, V.I., *Parametricheskie metody analiza mnogomernykh vremennykh ryadov* (Parametrical Approach to Multidimensional Time Series), Moscow: IFZ AN SSSR, 1986.
8. Kushnir, A.F. and Mostovoi, S.V., *Statisticheskii analiz geofizicheskikh polei* (Statistical Analysis of Geophysical Fields), Kiev: Naukova Dumka, 1990.
9. Brillinger, D., *Time Series. Data Analysis and Theory*, New York: Holt, Rinehart and Winston, 1975. Translated under the title *Vremennye ryady*, Moscow: Mir, 1980.
10. Marple, S., Jr., *Digital Spectral Analysis with Applications*, Prentice Hall, 1986. Translated under the title *Tsifrovoi spektral'nyi analiz i ego prilozheniya*, Moscow: Mir, 1990.
11. Nikolis, G. and Prigogine, I., *Exploring Complexity*, New York: Freeman, 1989. Translated under the title *Poznanie slozhnogo*, Moscow: Mir, 1990.
12. Gilmore, R., *Catastrophe Theory for Scientists and Engineers*, New York: Wiley, 1981. Translated under the title *Prikladnaya teoriya katastrof*, vols. 1 and 2, Moscow: Mir, 1984.
13. Sobolev, G.A., *Osnovy prognoza zemletryasenii* (Fundamentals of Earthquake Prediction), Moscow: Nauka, 1993.
14. Dobrovol'skii, I.P., *Mekhanika podgotovki tektonicheskogo zemletryaseniya* (Mechanics of Tectonic Earthquake Preparation), Moscow: IFZ AN SSSR, 1984.