Space-based monitoring of anomalies on the Earth's surface by adaptive point change algorithm

K.T. Protasov

Institute of Atmospheric Optics, Siberian Branch of the Russian Academy of Sciences, Tomsk

Received November 30, 1999

The problem of constructing an empirical model of the video data dynamics caused by time behavior of ecosystems is solved based on the linear representation in sliding Karhunen-Loeve basis of phenological variations in the ensemble of images. To detect unexpected anomalies recorded against the background of natural variations of radio brightness of the Earth's surface the adaptive Bayes' rule of the hypothesis testing is synthesized and the examples of algorithm operation presented.

Introduction

Appearance of the relatively inexpensive stations to receive video information from low-orbiting satellites of NOAA, "Meteor," SPOT, and "Resurs" series and similar to them allows one to formulate and solve the problems of resource and climate-ecological monitoring within a specific region using GIS (geographic information systems) technologies. The development of spaceborne instrumentation, increase in the number of spectral channels, the passage to hyper-spectral and stereo surveys, and the increased resolution of radiation detectors allow one to solve not only qualitative problems of nature management, but also the quantitative ones. The class of a national economy and ecological problems solved using the information acquired from the space-based platforms is now very wide, the detailed reviews in the collection "Science and engineering results"[1–4] are an evidence of that.

The opinion exists that the described methods to solve that or another problem can be used in the very broad scales and the period of laboratory studies is coming to end. However, a more detailed knowledge on the monitoring problems shows that at present only the simplest sub-problems of the most complex problem of space information analysis have been solved. Even a repetition of already obtained results requires their adaptation to specific conditions of a region, the high culture of data analysis, and powerful software and hardware of a GIS. One of such unsolved problems is the absence of a description model of the seasonal variations of landscape brightness characteristics. Indeed, in estimating the state of various ecosystems from observed video data one needs, first, a model of the dynamics of seasonal variations in the optical field of radio-brightness of the landscape formations, against the background of which the ecological events occur. Since the theoretical foundations to solve that complex problem are not developed, and the created phenological maps of seasonal variations of natural complexes indicate only the statistically average pattern of the landscape components dynamics, the necessity arises of developing empirical models of the dynamic video data caused by the behavior of ecosystems in time.

In this connection, let us consider an approach to construction of the linear model for describing the phenological time variations of an ensemble of images and an adaptive algorithm to detect the "anomalous outbursts" recorded against the background of natural variations of the Earth's surface (ES). Approbation of the algorithm has been carried out for the half-natural data obtained with the AVHRR device of the NOAA satellite and by simulating the anomalies.

1. Linear model of ensemble of video data in the Karhunen-Loeve basis and adaptive reconstruction of the probability descriptions of situations

At present an extensive material containing the data on dynamics of behavior of landscape optical characteristics has been accumulated by the stations of receiving of the satellite information. As independently functioning landscape units we choose the fragments which are selected in images by the cluster analysis algorithm as the texture homogeneous parts of image. The result of video data automatic classification is segmentation of images into the texture homogeneous regions with the similar optical characteristics.[1–4] Having enlarged the initial video information in such a natural way we will describe a behavior in time not of every separate pixel of an image but a cluster as a whole, as an optical image of landscape fragments. Let t denotes the time or integer number of the session of receiving the next space image, then the recorded collection of video data coordinated by scales for the same terrain part at the moment t has the form

$$\xi_0(u, \nu), \ldots, \xi_{t-\tau}(u, \nu), \ldots, \xi_t(u, \nu),$$

where $\xi(u, \nu) \equiv \xi(i, j)_{l \times m}$ is the fragment of a continuous image or its digitized matrix model; $(u, \nu) \in R^2$ are the values of pixel coordinates with the brightness $\xi(u, \nu)$ ($\xi(i, j)$ are the coordinates in the

0235-6880/00/04 317-09 $02.00

digitized version of notations, respectively, $i = 1, ..., l$, $j = 1, ..., m$; $l \times m$ is the number of samples (elements) of a fragment); τ is the time interval during which the landscape optical image is practically stationary (τ is the quasi-stationarity interval). A possible version of the observations of a fragment stored on an input medium of a more complex form, for example, when it corresponds to the certain cluster d selected beforehand, $d = 1, ..., Q$, where Q is the total number of clusters. We will select a collection of cluster points by the unity function $\mathbf{1}_d(u, \nu)$ of those values (u, ν) of the pixels, which form the cluster $d \in Q$ ($\mathbf{1}_d(u, \nu)$ is the indicator of the cluster domain). For clarity of notations used in the expressions, we will use continuous models of data representation assuming, however, that the latter are digitized, and integration can be replaced by summation within the admissible accuracy. We will describe the dynamics of the behavior in time of the generalized fragment by the following linear model:

$$\xi_t(u, \nu) = \mu_{t\tau}(u, \nu) + \sum_{i=1}^{k} X^i \, \phi_i^{t\tau}(u, \nu) + \eta(u, \nu), \quad (1)$$

where $\mu_{t\tau}(u, \nu)$ is the expectation; $\{\phi_i^{t\tau}(u, \nu)\}_1^k$ is the orthonormal Karhunen-Loeve (KL) basis that is estimated over the interval of quasi-stationarity τ; $\eta(u, \nu)$ is the δ-correlated (white) noise with the variance σ^2. The technique of obtaining such a basis from the sample data and the iterative algorithm to find successively "the most important" basis functions are expounded in the Appendix.[8–12] Note that the basis functions are defined in the set $\mathbf{1}_d(u, \nu)$ of those values of pixels $\{(u, \nu)\}$, which form a fragment or a cluster $d \in \{1, ..., Q\}$.

Let us call the model (1) Gaussian-like since for the fixed vector of parameters $\mathbf{X} = \mathbf{x}$ the observations are described accurate to the Gaussian noise with the distribution[11] $N\left\{0, \dfrac{\delta(u - r)\,\delta(\nu - s)}{\sigma^2}\right\}$. Sufficiently broad class of situations can be interpreted using the model (1), if we assume that the parameters $\mathbf{X} \in R^k$ are also random and are distributed with the probability density function $\omega(\boldsymbol{x})$. In this case the unconditional functional of the probability density distribution of observations $\xi(u, \nu)$ can be obtained by integration of the Gaussian density functional over the distribution $\omega(\boldsymbol{x})$, namely

$$f[\xi(u, \nu)] = \int_{R^k} f[\eta(u, \nu)/\boldsymbol{x}]\, \omega(\boldsymbol{x})\, d\boldsymbol{x} =$$

$$= C \int_{R^k} \exp\left\{-\frac{1}{2\sigma^2} \iint_{\mathbf{1}_d} \left[\mathring{\xi}(u,\nu) - \boldsymbol{x}^{\mathrm{T}}\phi(u,\nu)\right]^2 du\,d\nu\right\} \omega(\boldsymbol{x})\, d\boldsymbol{x}, \quad (2)$$

where C is the constant connected with the normalization of the functional $f[.]$; $\mathring{\xi}(u, \nu)$ are the centered observations; $d\boldsymbol{x} = dx^1 \times ... \times dx^k$, and the index τ is omitted; T is the transposition sign. To calculate $f[\xi(u, \nu)]$ we use the ideas of the adaptive Bayes' approach[5] based on the integration in Eq. (2) using the Laplacian approximation, then

$$f[\xi(u, \nu)] \cong C \exp\left\{-\frac{1}{2\sigma^2}\varepsilon_k^2\right\} \frac{\sigma^{-k}}{(\sqrt{2\pi})^2} \times$$

$$\times \int_{R^k} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{x} - \mathbf{x}^*)^{\mathrm{T}}(\boldsymbol{x} - \mathbf{x}^*)\right\} \omega(\boldsymbol{x})\, d\boldsymbol{x}, \quad (3)$$

$$\varepsilon_k^2 = \iint_{\mathbf{1}_d} \left[\mathring{\xi}(u, \nu) - \mathbf{x}^{*\mathrm{T}}\, \phi(u, \nu)\right]^2 du\, d\nu, \quad (4)$$

$$\mathbf{x}^* = \iint_{\mathbf{1}_d} \mathring{\xi}(u, \nu)\, \phi(u, \nu)\, du\, d\nu, \quad (5)$$

where $\mathbf{x}^*$ is the maximum likelihood (ML) estimate of the unknown vector of parameters $\mathbf{x}$, the components of this vector are the coefficients of expansion of the observed centered fragment $\mathring{\xi}(u, \nu)$ of the image in the basis $\phi(u, \nu) = \{\phi_i(u, \nu)\}_{i=1}^k$. Under conditions of weak effect of $\omega(\boldsymbol{x})$ on the value of the integral in Eq. (3) it is natural to change the *a priori* distribution of parameters $\omega(\boldsymbol{x})$, when it is unknown, for the distribution $\tilde{\omega}(\boldsymbol{x})$ of the maximum likelihood estimates of this vector, and to use, as an estimate of the integral in Eq. (3), its sample mean over the sample $\tilde{\mathbf{x}}_1, ..., \tilde{\mathbf{x}}_\tau$. This sample of ML estimates $\tilde{\mathbf{x}}_1, ..., \tilde{\mathbf{x}}_\tau$ is obtained by the expansion (5) of the data of training material $\mathring{\xi}_1(u, \nu), ..., \mathring{\xi}_\tau(u, \nu)$ over the quasi-stationarity interval τ. With the allowance for this fact the expression for the probability density functional of observations has the form

$$f[\xi(u, \nu)] = C \exp\left\{-\frac{1}{2\sigma^2}\varepsilon_k^2\right\} \frac{\sigma^{-k}}{\left(\sqrt{2\pi}\right)^k \cdot \tau} \times$$

$$\times \sum_{i=1}^{\tau} \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{x}^* - \tilde{\mathbf{x}}_i)^{\mathrm{T}}(\mathbf{x}^* - \tilde{\mathbf{x}}_i)\right\}. \quad (6)$$

Knowledge of conditional density functions $f[\xi(u, \nu)]$ determined for each class $d \in \{1, ..., Q\}$ allows one to construct the adaptive Bayes decision rules of recognition of situations in dynamics of movement of the stationarity interval τ within the ensemble of observations

$$U = \arg \max_{i=1,...,Q} P_i\, f_i[\xi(u, \nu)], \quad (7)$$

where P_i are the *a priori* probabilities of the appearance of classes; U is the solution, $U \in \{1, ..., Q\}$. The decision rule, in this case, has the following sense: the component $\exp\{-\varepsilon_k^2/(2\sigma^2)\}$ estimates the degree to which the observed sample $\xi(u, \nu)$ belongs to the linear subspace $\{\phi_i(u, \nu)\}_1^k$, which well describes the given class, and the component

$\sum_{i=1}^{\tau} \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{x}^* - \tilde{\mathbf{x}}_i)^{\mathrm{T}}(\mathbf{x}^* - \tilde{\mathbf{x}}_i)\right\}$ estimates the degree to which the projection of the sample $\xi(u, \nu)$ on the subspace $\{\phi_i(u, \nu)\}_1^k$ belongs to another points of the training material which are disposed in this subspace. Note that practical implementation of the algorithm of constructing the estimating Bayes' rule is based on the use of conditional density functions $f(\xi(i, j))$, but not the density functionals because the image $\xi(u, \nu)$ is assumed to be digitized with a high accuracy and is represented in a discrete form. To solve the classification problems by the rule (7) and to search the array of elements of the image that unexpectedly changed their statistical characteristics in connection with the discord of the process, it is necessary to set such a search window of minimum dimensions, within which the decision on the process discord is being taken. The whole space of image is analyzed by scanning the image field with an elementary fragment and the recognition problem is solved accurate to the space fragment.

Let us illustrate the operation of the algorithm to construct the KL basis with the following simple example. When the observed ensemble of images of the same part of the Earth's underlying surface (flood-land of the Ob River and Novosibirsk Reservoir) is recorded by the NOAA satellite (AS) with the AVHRR device within the sliding temporal interval of $\tau = 12$ successive observations in the first spectral range. From these images, the sliding KL basis was reconstructed by the method considered in Appendix.[8–12] The approximation significance of each of the basis functions is characterized by the eigenvalue λ, the spectrum of eigenvalues obtained is presented in Fig. 1*a*. The quality of the approximation (4) by one or another set of basis functions is illustrated in Fig. 1*b*.

The plots show that the small number (for example, five) of basis functions quite well describes the observed statistics of the images. The estimate of mathematical expectation and five the most significant functions of the KL basis are presented in Fig. 2 (where *a* is the mathematical expectation, *b*–*f* are the 1st–5th basis functions, respectively). The model (1) can further be used not only for filtration and compact description of the phenological variations of the portraits of landscape formations, but for the detection of anomalous phenomena on the Earth's surface from the video data acquired from space by setting or estimating the confidence intervals for the norm and "pathology."

2. Detection and selection of EUS anomalies by the adaptive change point algorithm

Let us consider the problem of detecting unexpected anomalies in the following formal statement[6,7]: let us assume that the optical image of the same part of the underlying surface $\xi_t(x, y) = \xi_t(i, j)_{l\times m}$ is recorded periodically at the times $t_1, t_2, \ldots, t_r, \ldots, t_n$. At the moment t_n, the quasi-stationary state of the fragment of observation is described by the following model:

$$\xi_n(u, \nu) = \mu_n(u, \nu) + \boldsymbol{X}_n^{\mathrm{T}} \phi_n(u, \nu) + \eta_n(u, \nu), \quad (8)$$

where $\eta_n(u, \nu)$ are independent realizations of the uncorrelated noise. Let us assume that the observed fragments of images can change their properties in the following way:

$$\mu_n(u, \nu) + \boldsymbol{X}_n^{\mathrm{T}} \phi_n(u, \nu) = $$

$$= \begin{cases} \mu_0(u, \nu) + \boldsymbol{X}^{\mathrm{T}} \phi(u, \nu), \; n \le r + 1, & \text{situation } A_0, \\ \mu_1(u, \nu) + \boldsymbol{X}^{\mathrm{T}} \varphi(u, \nu), \; n \ge r, & \text{situation } A_1, \end{cases} \quad (9)$$

where $\varphi(u, \nu)$ is the new orthonormal KL basis.

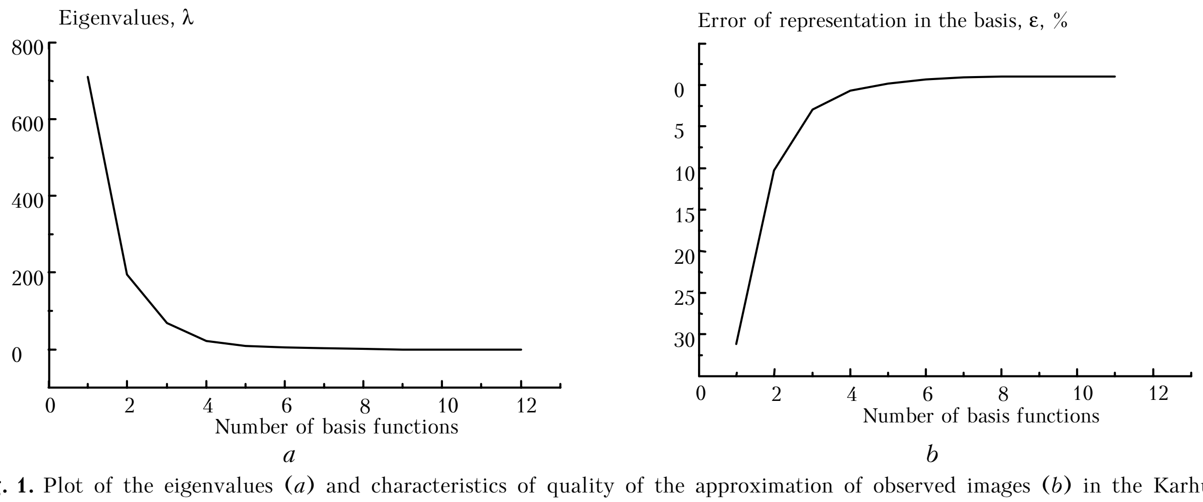


Fig. 1. Plot of the eigenvalues (*a*) and characteristics of quality of the approximation of observed images (*b*) in the Karhunen-Loeve basis.

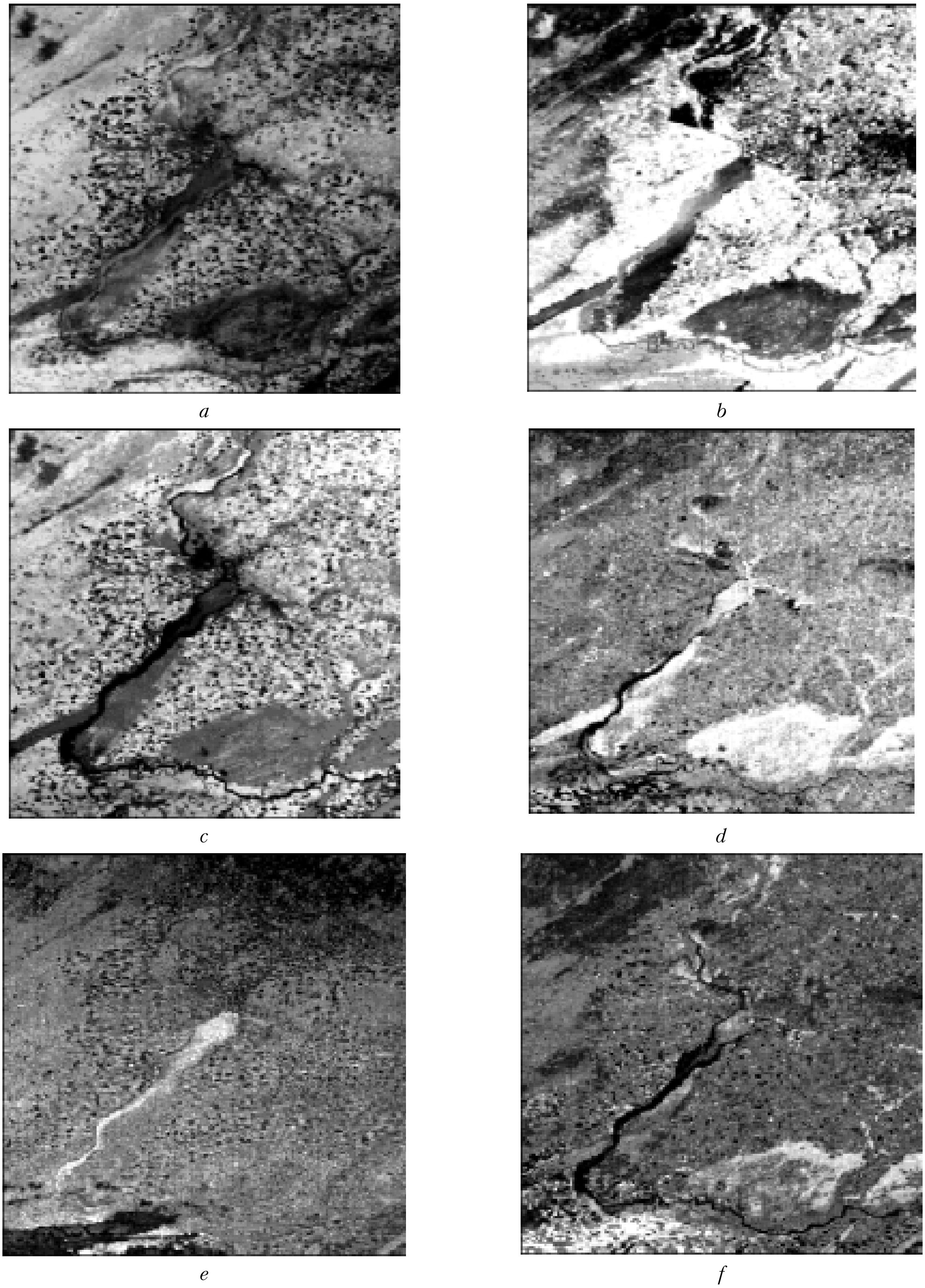


Fig. 2. The expectation (*a*) and five KL basis functions (*b* − *f*) obtained from the statistics of satellite images.

In other words, in the point changing process both the mathematical expectation and the correlation properties of the images, which are described by the set of basis functions, change. Detection of changes in the properties of the process is equivalent to the H_1 hypothesis making (the change points of the process are observed) when this hypothesis is tested against the alternative H_0 (the change points are absent, $r > n$).

For testing the hypotheses H_1 and H_0 we write the decision rule of the likelihood ratio allowing for independence of the observations

$$L(r, \mu_1) = \frac{\prod_{l=1}^{r-1} f_0[\xi_l(u, v)] \prod_{l=r}^{n} f_1[\xi_l(u, v)]}{\prod_{l=1}^{n} f_0[\xi_l(u, v)]} = \prod_{l=r}^{n} \frac{f_1[\xi_l(u, v)]}{f_0[\xi_l(u, v)]} \cong$$

$$\cong \prod_{l=r}^{n} \frac{\exp\left\{-\frac{1}{2\sigma^2} \iint (\mathring{\xi}_l(u,v) - \mathbf{y}_l^{*\mathrm{T}} \boldsymbol{\varphi}(u, v))^2 \, du dv\right\} \times}{\exp\left\{-\frac{1}{2\sigma^2} \iint (\mathring{\xi}_l(u,v) - \mathbf{x}_l^{*\mathrm{T}} \boldsymbol{\phi}(u, v))^2 \, du dv\right\} \times}$$

$$\frac{\times \sum_{j=r, j\neq l}^{n} \exp\left\{-\frac{1}{2\sigma^2} (\mathbf{y}_l^* - \tilde{\mathbf{y}}_j)^{\mathrm{T}} (\mathbf{y}_l^* - \tilde{\mathbf{y}}_j)\right\}}{\times \sum_{j=r, j\neq l}^{n} \exp\left\{-\frac{1}{2\sigma^2} (\mathbf{x}_l^* - \tilde{\mathbf{x}}_j)^{\mathrm{T}} (\mathbf{x}_l^* - \tilde{\mathbf{x}}_j)\right\}} \underset{H_0}{\overset{H_1}{\gtrless}} 1, \quad (10)$$

where the centering is carried out relative to the corresponding expectations, and the expansion coefficients (their maximum likelihood estimates) $\mathbf{y}$ and $\mathbf{x}$ are calculated relative to the corresponding KL bases $\boldsymbol{\varphi}(u, v)$ and $\boldsymbol{\phi}(u, v)$ by the formula (5). For the correct operation of the algorithm, the stage of preliminary training is needed. This stage consists in finding the basis $\boldsymbol{\phi}(u, v)$ and ML estimates $\tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_\tau$ from the quasi-stationary sample $\xi_1(u, v), \ldots, \xi_\tau(u, v)$ $\tau < r$ of preliminary observations when it is known for sure that the state of the nature is characterized by the situation A_0. We use the obtained decision rule to detect the change points by moving from n backward and taking $r = n - 2$ we calculate

a) $\mu_1(u, v) \cong \frac{1}{n - r + 1} \sum_{l=r}^{n} \xi_l(u, v)$ (this estimate makes it unnecessary to solve the complex problem on optimizing the likelihood functional with respect to $\mu_1(u, v)$);

b) we center the realizations $\xi_r(u, v), \ldots, \xi_n(u, v)$ relative to $\mu_1(u, v)$ and $\mu_0(u, v)$;

c) we construct a new basis $\boldsymbol{\varphi}(u, v)$ by the method described in Appendix[8–12] and obtain the ML estimates $\tilde{\mathbf{y}}_r, \ldots, \tilde{\mathbf{y}}_n$ in this basis and $\tilde{\mathbf{x}}_r, \ldots, \tilde{\mathbf{x}}_n$ in the basis $\boldsymbol{\phi}(u, v)$ obtained before;

d) moving (decreasing) r we find $(\tilde{r}, \tilde{\mu}_1) = \arg \max_{\{r\}} \max_{\{\mu_1(u,v)\}} L(r, \mu_1)$;

e) finally, we test the hypotheses H_1 and H_0 for $\tilde{r}, \tilde{\mu}_1, L(\tilde{r}, \tilde{\mu}_1) \underset{H_0}{\overset{H_1}{\gtrless}} 1$ which have been sought.

In a similar way the algorithm with moving along a sample forward is constructed.

Taking into account that the realization of such algorithms for detection of change points is too cumbersome we consider a more simple version connected with that in the model (9) only the mathematical expectation changes at the change points, while the correlation of the process remains unchanged and is described by the set of the KL basis functions $\boldsymbol{\phi}(u, v)$. One more simplification is connected with that the averaging in the expression for the density functional (2) is carried out using improper distribution $\omega(\boldsymbol{x})$, which is set on the extending data medium as a constant A inessential for the decision rule

$$\int_{R^k} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{x} - \mathbf{x}^*)^{\mathrm{T}} (\boldsymbol{x} - \mathbf{x}^*)\right\} \omega(\boldsymbol{x}) \, d\boldsymbol{x} \cong$$

$$\cong \lim_{A\to\infty} \frac{1}{A} \int_{A} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{x} - \mathbf{x}^*)^{\mathrm{T}} (\boldsymbol{x} - \mathbf{x}^*)\right\} d\boldsymbol{x} \cong C .$$

In this case, the general form of the expression for the density functional has the following form:

$$f[\xi(u,v)] \cong$$

$$\cong C \exp\left\{-\frac{1}{2\sigma^2} \iint (\xi(u,v) - \mu(u,v) - \mathbf{x}^* \boldsymbol{\varphi}(u,v))^2\right\} du dv,$$

where $\mathbf{x}^*$ is the maximum likelihood estimate determined from the expression (5). The logarithm of likelihood ratio in this case is expressed as

$$\Lambda_n(r, \mu_1) =$$

$$\sum_{l=r}^{n} \frac{1}{2\sigma^2} \left\{ \iint_{\mathbf{1}_d} [\xi_l(u,v) - \mu_0(u,v) - \mathbf{x}_l^{*\mathrm{T}} \boldsymbol{\phi}(u,v)]^2 \, du dv - \right.$$

$$\left. - \iint_{\mathbf{1}_d} [\xi_l(u,v) - \mu_1(u,v) - \mathbf{x}_l^{*\mathrm{T}} \boldsymbol{\phi}(u,v)]^2 \, du dv \right\} =$$

$$= \iint_{\mathbf{1}_d} \left\{ \frac{\mu_1(u, v) - \mu_0(u, v)}{\sigma^2} \sum_{l=r}^{n} \left[\xi_l(u,v) - \mathbf{x}_l^{*\mathrm{T}} \boldsymbol{\phi}(u,v) - \right.\right.$$

$$\left.\left. - \mu_0(u,v) - \frac{\mu_1(u, v) - \mu_0(u, v)}{2} \right] du dv \right\} \underset{H_0}{\overset{H_1}{\gtrless}} 0. \quad (11)$$

Let $\Delta(u, v) = \mu_1(u, v) - \mu_0(u, v)$ then

$$\frac{\delta\Lambda_n(r, \Delta)}{\delta\Delta(u, v)} \cong \frac{1}{n - r + 1} \times$$

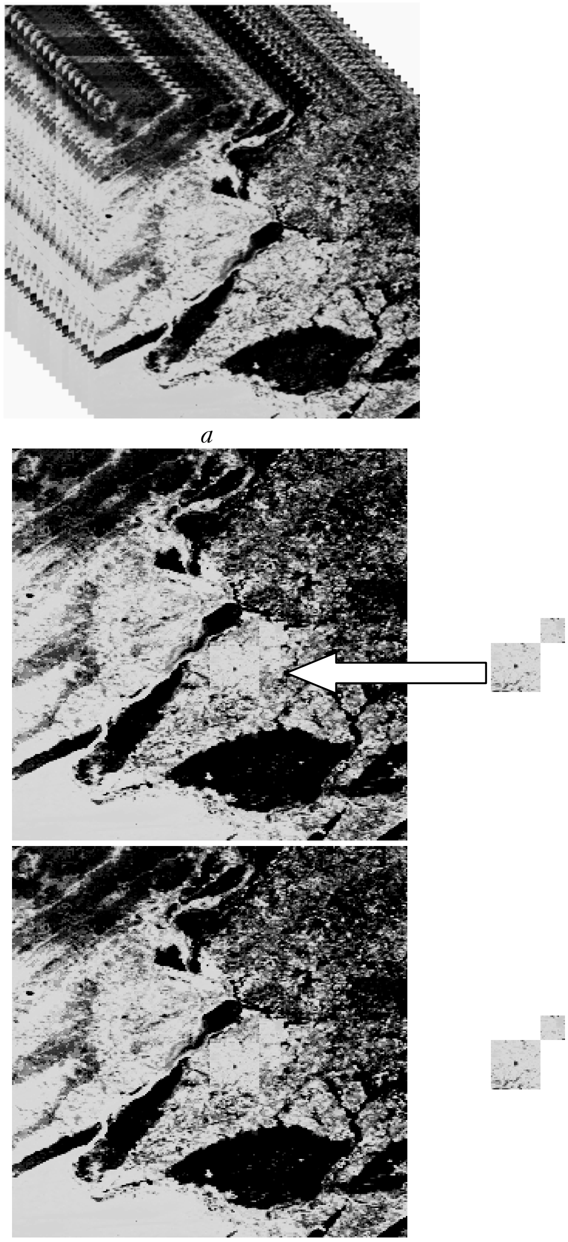
$$\times \sum_{l=r}^{n} [\xi_l(u,v) - \mu_0(u,v) - \mathbf{x}_l^{*\mathrm{T}} \boldsymbol{\phi}(u,v) - \Delta(u,v)] = 0 ,$$

$$\tilde{\mu}_1(u, v) = \frac{1}{n - r + 1} \sum_{l=r}^{n} [\xi_l(u,v) - \mathbf{x}_l^{*\mathrm{T}} \boldsymbol{\phi}(u, v)], \quad (12)$$

$$\tilde{\mu}_1(u, v) = \frac{1}{n - r + 1} \sum_{l=r}^{n} \xi_l(u,v),$$

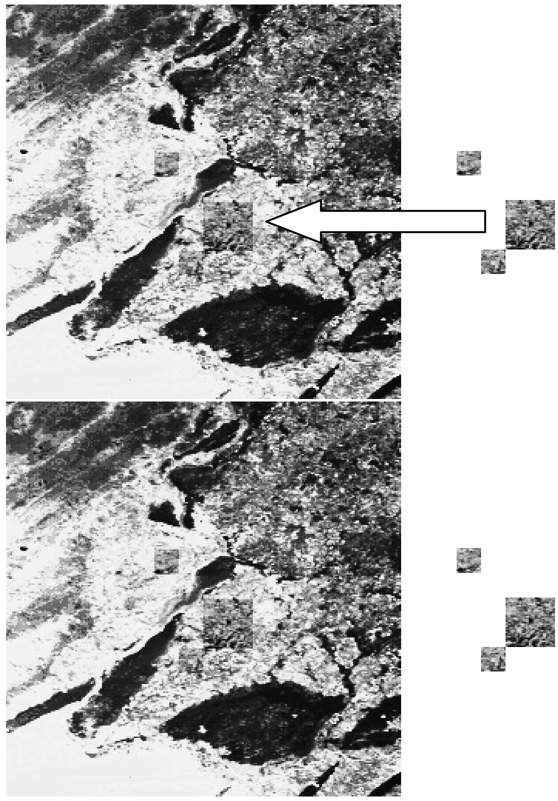
since

$$\frac{1}{n - r + 1} \sum_{l=r}^{n} \mathbf{x}_l^{*\mathrm{T}} \boldsymbol{\phi}(u, v) \cong 0.$$



a

b



c

Fig. 3. Ensemble of "homogeneous" images (*a*); a nonstandard increase of brightness detected and isolated by the algorithm (*b*); a decrease of brightness detected and isolated by the algorithm (*c*).

To illustrate the algorithm operation the anomalous distortions were artificially introduced into the newly observed images. These were in the form of an increase in the brightness level (two times exceeding the background) and a decrease of the brightness level in the system of fragments with 10×10 pixels. Then these distortions were detected by the decision rule (11) at the steps t_{13}, t_{14}, and t_{15} and the position of a fixed anomaly was localized. Figure 3*a* presents the ensemble of observations on the quasi-stationarity interval, and Figures 3*b* and *c* present the anomalies detected correctly by the algorithm (they are shown outside the frame of analyzed fragments).

Conclusion

An advantage of the proposed approach is that simple generalization of the KL representation[8–12] allows one to construct linear models for vector images with the components obtained in several spectral ranges. It is preferable to take the probability model of the sequence of observations (8) and (9) in constructing the decision rule to detect the change points using, for example, the Johnson parameterization.

Acknowledgments

This work was carried out with the partial financial support of the Russian Foundation of Basic Research (Grant No. 98–0503194).

Appendix
Linear models of multidimensional fields in the Karhunen-Loeve basis

Processing of multidimensional experimental information widely uses the linear models of the data representation in orthogonal bases. Among such bases, the Karhunen-Loeve basis (which is known in the publications in meteorology, hydrology, oceanology, and atmospheric and ocean physics as the basis of empirical orthogonal functions) is preferable. In this case, the approximating series has the least number of components and preserves the high accuracy of the data approximation.[8–12]

The tendencies existing toward the increase of the dimensionality of recorded information (in particular, due to appearance of multi-zonal aerospace survey) and the problems of combined data processing which arise in this case make the problem of synthesis of linear models of multidimensional observations described by the random vector fields to be very urgent. In this connection, we consider below a sufficiently general problem to represent vector fields of vector argument and to find the corresponding Karhunen-Loeve basis based on the experimental data and propose the iterative algorithm of seeking approximate solution to this problem. We will assume that the random vector field (for simplicity and without loss of generality, it is centered) $\xi(\mathbf{u}) = (\xi^1(\mathbf{u}), \ldots, \xi^s(\mathbf{u}))^{\mathrm{T}}$ of the vector argument $\mathbf{u} = (u^1, \ldots, u^\nu)^{\mathrm{T}}$ (s and ν are, respectively, the dimension of the function $\xi(.)$ and the dimension of the argument $\mathbf{u}$, T is the transposition sign) is set in the domain $D = \{\mathbf{u}\colon u_a^i \le u^i \le u_b^i,\ i = 1, \ldots, \nu\}$ by the array of N samples $\xi_1(\mathbf{u}), \ldots, \xi_N(\mathbf{u})$.

Let us present the vector field in following way (it is not unique, generally speaking[11]):

$$\xi(\mathbf{u}) = \lim_{k\to\infty} \sum_{i=k}^{n} X^i\, \phi_i(\mathbf{u}), \tag{A.1}$$

where the limit is interpreted in the sense of convergence by norm in the space of realizations of a random vector field; $\{\phi_i(\mathbf{u})\}_k$ are the vector basis functions of the vector argument. Random coefficients $\{X^i\}_k$ are determined from the condition of root-mean-square deviation minimum

$$\varepsilon_k^2 = M \left\| \left(\xi(\mathbf{u}) - \sum_{i=1}^{k} X^i\, \phi_i(\mathbf{u})\right) \right\|^2, \tag{A.2}$$

where M is the sign of the mathematical expectation operator; $\|\cdot\|$ is the Euclidean norm in the space of observations. If we impose the orthonormality conditions on the basis functions $\{\phi_i(\mathbf{u})\}_k$

$$(\phi_i, \phi_j) = \int_D \phi_i^{\mathrm{T}}(\mathbf{u})\, \phi_j(\mathbf{u}) \mathrm{d}\mathbf{u} = \delta_{ij}, \tag{A.3}$$

where δ_{ij} $(i, j = 1, \ldots, k)$ is the Kronecker symbol, $\mathrm{d}\mathbf{u} = \mathrm{d}u^1 \times \ldots \times \mathrm{d}u^\nu$, and $(\cdot,\cdot)$ is the sign of scalar product, then the representation coefficients $\{X^i\}_k$ minimizing (A.2) will take the form

$$X^i = (\xi, \phi_i) = \int_D \xi^{\mathrm{T}}(\mathbf{u})\, \phi_i(\mathbf{u}) \mathrm{d}\mathbf{u}, \quad i = 1, \ldots, k. \tag{A.4}$$

The existence of the limit in (A.1) and complete orthonormal sequence of the basis functions $\{\phi_i(\mathbf{u})\}_\infty$ is provided by considering only such processes $\xi(\mathbf{u})$, which satisfy the following condition: for every fixed set of the values of vector components $\mathbf{u} \in D$ $M\,[\xi^{\mathrm{T}}(\mathbf{u})\, \xi(\mathbf{u})] < \infty$. It is natural also to find the basis functions $\{\phi_i(\mathbf{u})\}_k$ from the conditions of minimum of the root-mean-square criterion of quality (A.2) of the vector field approximation $\xi(\mathbf{u})$ by a segment of the set (A.1) which contains k terms. Solution of this variational problem on seeking the extremum that is conditional in the sense of constraints (A.3), included in the functional (A.2) using the Lagrangian factors, causes the following homogeneous Fredholm equation of the second kind

$$\int_D M[\xi(\mathbf{u})\, \xi^{\mathrm{T}}(\mathbf{v})]\, \phi(\mathbf{v})\, \mathrm{d}\mathbf{v} = \lambda\phi(\mathbf{u}), \tag{A.5}$$

where λ is the Lagrangian factor, and the indices of basis functions and λ are omitted since all the equations are equivalent. The unknown basis $\{\phi_i(\mathbf{u})\}_k$ corresponding to the largest k eigenvalues $\{\lambda_i\}_k$ is found by the solution of the equation (A.5), however, in the general case it is not a simple problem. Having at a disposal the collection of N samples $\xi_1(\mathbf{u}), \ldots, \xi_N(\mathbf{u})$, which characterize "sufficiently brim-full" the general collection of all realizations caused by the random field $\xi(\mathbf{u})$ it is natural to use the following sample estimate of the correlation function:

$$M[\xi(\mathbf{u})\, \xi^{\mathrm{T}}(\mathbf{v})] \cong \frac{1}{N} \sum_{j=1}^{N} \xi_j(\mathbf{u})\, \xi_j^{\mathrm{T}}(\mathbf{v}). \tag{A.6}$$

In this case the problem (A.5) is essentially simplified (the case of the degenerate kernel in (A.5)[12]). Really, substituting (A.6) into the integral equation (A.5) we obtain

$$\int_D \sum_{j=1}^{N} \xi_j(\mathbf{u})\, \xi_j^{\mathrm{T}}(\mathbf{v})\, \tilde{\phi}(\mathbf{v})\, \mathrm{d}\mathbf{v} = \Delta\tilde{\phi}(\mathbf{u}), \tag{A.7}$$

where $\Delta = N\hat{\lambda}$ and $\hat{\lambda}$ and $\tilde{\phi}(\mathbf{u})$ are the estimates of the corresponding λ and $\phi(\mathbf{u})$. Let us introduce the following designations:

$$\int_D \xi_j^{\mathrm{T}}(\mathbf{v})\, \tilde{\phi}(\mathbf{v})\, \mathrm{d}\mathbf{v} = c^j, \quad j = 1, \ldots, N,$$

then from (A.7) we obtain expression for the basis functions

$$\tilde{\phi}(\mathbf{u}) = \frac{1}{\Delta} \sum_{j=1}^{N} c^j \xi_j(\mathbf{u}), \qquad (A.8)$$

where the coefficients $\{c^j\}_N$ are not yet determined. Substituting the parameterized expression (A.8) of the basis function $\tilde{\phi}(\mathbf{u})$ into the equation (A.7) we obtain the equality

$$\frac{1}{\Delta} \sum_{j=1}^{N} \xi_j(\mathbf{u}) \sum_{i=1}^{N} c^j \int_D \xi_j^{\mathrm{T}}(\mathbf{v})\, \xi_i(\mathbf{v})\, \mathrm{d}\mathbf{v} = \sum_{j=1}^{N} c^j \xi_j(\mathbf{u}). \quad (A.9)$$

Let us calculate, in this expression, the scalar product (ξ_j, ξ_i) over the realizations of the random field $\xi_1(\mathbf{u})$, ..., $\xi_N(\mathbf{u})$ with a_{ji} to denote it, then (A.9) takes the form

$$\sum_{j=1}^{N} \xi_j(\mathbf{u}) \left\{ \frac{1}{\Delta} \sum_{i=1}^{N} c^i a_{ji} - c^j \right\} = 0. \qquad (A.10)$$

Because of the linear independence of realizations of the random field in the probability sense and because of the property of linearly independent, with the scalar product, elements of space, the equality (A.10) holds under the condition

$$\frac{1}{\Delta} \sum_{i=1}^{N} c^i a_{ji} - c^j = 0, \quad j = 1, \ldots, N. \qquad (A.11)$$

Using the matrix representation the expression (A.11) reduces to the following form:

$$(a_{ji})\mathbf{c} = \mathbf{c}\Delta, \qquad (A.12)$$

where $\mathbf{c} = (c^1, \ldots, c^N)^{\mathrm{T}}$; (a_{ji}) is the Gram matrix $N \times N$; $\Delta = (\Delta^i \delta_{ij})$.

Thus, when the structure of basis functions as a linear combination of realizations of a random process was determined, the coefficients of these linear combinations were obtained by solving the complete eigenvalue problem for the positive definite Gram matrix (a_{ji}) of the order N, and it is already the problem that can practically be solved using the numerical methods of algebra. It is not difficult to test, by substituting the expression (A.8) into the equation (A.3) that to normalize the functions $\{\tilde{\phi}_i(\mathbf{u})\}_k$ allowing for the obtained values $\{c^i\}_N$ it is necessary to replace Δ by $\sqrt{\Delta}$ in Eq. (A.9) for basis functions.

Constructing linear models (A.1) using the Karhunen-Loeve basis, which is optimal in the root-mean-square sense, requires solving the equations (A.12), as a rule, by numerical methods. However, difficulties of practical realization hamper broad use of this basis, since in this case it is necessary to solve the complete eigenvalue problem for the positive definite matrices when their order exceeds 10^2. This normally makes one to refuse from using direct methods to solve the problem of finding the Karhunen-Loeve basis (A.12) and to construct iterative algorithms. These algorithms, while reducing the number of operations in obtaining approximate results cause the optimal solution only asymptotically. Another advantage of the iterative algorithms is that they allow one to find, in the first instance, “the most important” basis functions the number of which can be small.

One of the approaches to overcome this difficulty at the expense of refusing from the optimality in the root-mean-square sense is the algorithm enabling one to construct an adapted basis,[10] which uses the idea of orthogonalization of a sequence of linear-independent functions under the condition that the choice of the next function satisfies certain criterion. In this case, the uniform approximation of the process by its linear variety with small dimensionality is carried out. The iteration character of the procedure to construct this basis allows the algorithm for transformation of spaces with large dimensionality up to 10^5 to be used.

Further, we propose an iteration algorithm for constructing a basis adapted in the root-mean-square sense (ARMSS basis). In this case, the choice of the next basis function is based on the minimization of the root-mean-square criterion of quality (A.2).

Let us approximate the realizations of the initial description $\xi(\mathbf{u})$ by the elements of a linear subspace G_k (containing all linear combinations) which is set by the orthonormal basis $\{\varphi_i(\mathbf{u})\}_k$ in the following way. As the next basis function $\varphi_{s_j}(\mathbf{u})$ ($j = 1, \ldots, k$) we take that function from the s_j orthonormal functions $\varphi_{s_j}(\mathbf{u})$ obtained by the process of Gram-Schmidt orthogonalization[12] of the sample functions $\xi_1(\mathbf{u})$, ..., $\xi_N(\mathbf{u})$

$$\varphi_{s_j}(\mathbf{u}) = \frac{\varphi^*_{s_j}(\mathbf{u})}{2\varphi^*_{s_j}(\mathbf{u})2},$$

$$\varphi^*_{s_j}(\mathbf{u}) = \xi_{s_j}(\mathbf{u}) - \sum_{i=1}^{j-1} (\xi_{s_j}, \varphi_i)\, \varphi_i(\mathbf{u}), \qquad (A.13)$$

when

$$\lambda_j = M\,[(\xi, \varphi_j)^2] = \max_{[s_j]} M[(\xi, \varphi_{s_j})^2], \quad (A.14)$$

$$s_j = 1, \ldots, N; \quad j = 1, \ldots, k; \quad k \le N,$$

where the sample estimate associated with (A.6) is used as the mathematical expectation. The process of seeking the basis functions is finished at the k-step once the preset accuracy ε_k^2 of the approximation of random vector field by the linear combination of k basis elements from G_k is achieved. In this case, following the theorem of construction of elements $\xi(\mathbf{u})$ of the Hilbert space at G_k we have

$$\xi(\mathbf{u}) \cong \sum_{j=1}^{k} X^i \varphi_j(\mathbf{u}), \qquad (A.15)$$

where $\{X^i\}_k$ is the collection of random numbers determined by the formula

$$X^j = (\xi, \varphi_j), \quad j = 1, \ldots, k. \qquad (A.16)$$

The accuracy of approximation in the expression (A.15) is determined (practically, the estimates $M[\cdot]$ are used on the same sample) by the following way:

$$\varepsilon_k^2 = M\,[(\xi, \xi)] - \sum_{j=1}^{k} \lambda_j, \qquad (A.17)$$

where λ_j are ordered according to the magnitude decrease ($\lambda_1 \ge \lambda_2 \ge \ldots \ge \lambda_k$) because of the construction reasons, and maximally, as far as it is permitted by the ensemble $\{\xi_j(\mathbf{u})\}_N$, "exhausts" the root-mean-square error ε_k^2 of the approximation of the ensemble $\{\xi(\mathbf{u})\}$ by the linear variety from G_k. As is known, the optimal, in the root-mean-square sense, basis $\{\phi_j(\mathbf{u})\}_k$ with the corresponding spectrum of eigenvalues $\{\lambda_j\}_k$ of the Karhunen-Loeve expansion is found by the optimization of the criterion (A.17) over $\{\phi_j(\mathbf{u})\}$ allowing for the orthonormality property of the latter. The same basis $\{\phi_j(\mathbf{u})\}_k$ is caused by the problem of a successive maximization of the positive definite quadratic form (A.14) over the spheres of unit radii in the subspaces, which are orthogonal to the functions $\{\phi_j(\mathbf{u})\}_k$ obtained by the procedure (A.13), $j = 1, \ldots, k$, $\phi_0 \equiv 0$. The latter circumstance points that the ARMSS basis $\{\varphi_j(\mathbf{u})\}_k$ obtained by the algorithm (A.13) and (A.14) asymptotically, with the increase of N, becomes the Karhunen-Loeve basis under certain assumptions on the ensemble $\{\xi(\mathbf{u})\}$. Really, the considered algorithm is, in fact, based on the stochastic principles of search of extremum[12] with the peculiarity only that the "test" functions in the considered case are the elements of a sample. Therefore, for the convergence of the search procedure it is necessary that the function of sample value distribution be positive along the "direction" of unknown solutions. If this fact is not established *a priori*, then we can judge on the quality of the obtained solution by the magnitude of estimate of the approximation error (A.17):

$$\tilde{\varepsilon}_k^2 = \frac{1}{N} \sum_{j=1}^{N} \left[(\xi_j, \xi_j) - \sum_{i=1}^{k} (\xi_j, \varphi_i)^2 \right].$$

Note that the algorithms obtained for continuous fields are correct for the fields set by separate samples on a discrete, regular or stochastic (but fixed) network of observations. However, in these cases, the integration is replaced by the corresponding summation over the set of points, where the realizations of field are recorded. It is simple to perform some other modifications of the algorithms which are connected with the representation of fields over the domain D with the variable boundaries, sliding boundaries or to separate out from the linear model (A.1) the temporal variable in the following way:

$$\xi(\mathbf{u}, t) = \sum_{i=1}^{k} X^i(t)\, \phi_i(\mathbf{u}).$$

References

1. *Science and Engineering Results.* Issue: *Investigation of Earth from Space.* Vol. 1. *Physical Foundations, Methods, and Means of Investigation of the Earth from Space* (All-Union Institute of Science and Engineering Information, Moscow, 1987), 196 pp.

2. *Science and Engineering Results.* Issue: *Investigation of Earth from Space* Vol. 2. *Processing and Application of Aerospace Information on the Earth* (All-Union Institute of Science and Engineering Information, Moscow, 1987), 172 pp.

3. *Science and Engineering Results.* Issue: *Investigation of Earth from Space* Vol. 3. *Geographic Information Systems, Remote Sensing, and Their Practical Application* (All-Union Institute of Science and Engineering Information, Moscow, 1989), 212 pp.

4. *Science and Engineering Results.* Issue: *Investigation of Earth from Space.* Vol. 1. *Spaceborne Systems, Instrumentation, and Devices to Solve Problems of Nature Management and Ecological Control* (All-Union Institute of Science and Engineering Information, Moscow, 1991), 144 pp.

5. V.G. Repin and G.P. Tartakovckii, *Statistical Synthesis at an a Priori Uncertainty and Adaptation of the Information Systems* (Sovetskoe Radio, Moscow, 1977), 432 pp.

6. *Detection of Change of Properties of Signals and Dynamic Systems,* M. Bassvil, A. Vilski, A. Banvenist, et. al. [Russian translation] (Mir, Moscow, 1977), 432 pp.

7. K.T. Protasov, "*Detection of anomalies of the Earth's underlying surface in space photos by the point change algorithm for GIS,*" in: *Geographic Information Science. Theory and Application,* issue 1 (Tomsk State University Publishing House, Tomsk, 1998, 415).

8. K. Fukunaga, *Introduction in Statistical Theory of Pattern Recognition* (Nauka, Moscow, 1979), 368 pp.

9. J.T. Tou and R.C. Gonzales, *Pattern Recognition Principles* (Addison-Wesley Publishing Company, London, 1974), 411 pp.

10. Yu.I. Neyimark, ed., *Pattern Recognition and Medical Diagnostics* (Nauka, Moscow, 1972), 328 pp.

11. H.L. van Trees, *Detection, Estimation, and Modulation Theory.* P. I. (John Wiley and Sons, New York, 1968), 744 pp.

12. K.T. Protasov, Atm. Opt. **3**, No. 1, 42–46 (1990).