Spatial extrapolation of meteorological fields in mesoscale range based on the mixed four-dimensional dynamic-stochastic model and the Kalman filtering apparatus

V.S. Komarov, A.V. Lavrinenko, N.Ya. Lomakina, Yu.B. Popov, A.I. Popova, and S.N. Il'in

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

Received April 27, 2004

A technique and algorithm of mesometeorological field spatial extrapolation based on the mixed four-dimensional dynamic-stochastic model and the Kalman filtering apparatus are considered. The statistical estimates of the quality of the proposed algorithm as applied to spatial extrapolation of the mesoscale wind and temperature fields to a territory uncovered by observations are discussed.

Introduction

Basic problems of current mesometeorology include estimation and forecasting of the atmospheric state over the territory uncovered by observations from the measurements obtained in adjacent regions. For a long time these problems were solved within the framework of objective analysis of meteorological fields based on the method of optimal interpolation.[1,2]

In recent years, in connection with the increasing amount and types of meteorological information, the traditional procedure of objective analysis is often displaced by the procedure of four-dimensional assimilation of data. The latter is a combination of two traditionally different tasks: objective analysis and forecast of meteorological fields. The prognostic model usually incorporates a set of hydrothermodynamic equations (hydrodynamic model). A disadvantage of this approach is that the prognostic model of hydrodynamic type is used only as a time extrapolant, and the forecasting procedure does not provide for refinement of the model parameters at the next time step.

Another approach to solution of the problem of spatial extrapolation of meteorological fields in the mesoscale region is proposed in this paper, which is based on the use of the Kalman filtering algorithm and the mixed four-dimensional dynamic-stochastic model. The model describes the meteorological field variation simultaneously in time and space without invoking voluminous operational information about the atmosphere state. At the same time, application of the Kalman filter allows a real time estimation and step-by-step correction of the prognostic model parameters as the measurements income from a local network of aerological stations.

This paper continues our earlier works,[3,4] in which the linear few-parameter dynamic-stochastic model based on the first-order differential equations was laid in the foundation of the approach.

The dynamic-stochastic approach is used to decrease the size of covariance matrices of estimation and forecasting errors due to simplification of the prognostic model of a meteorological field behavior in space and time, that is, at decreasing the state vector dimension. The same issues are also considered in Refs. 5–7. For the prognostic hydrodynamic model, it is proposed to pre-calculate covariance matrices with the use of approximating functions. In this case, the results of calculation of the corresponding matrices depend only on the initial data. From the mathematical point of view, such a solution is not adequate to the classical optimal Kalman filtering.

1. Formulation and method of solution of the problem

The problem of spatial extrapolation of a centered meteorological field ξ' in the mesoscale range consists in estimating its value at the point with rectangular coordinates (x_0, y_0, z_0) from measurements at the points with coordinates (x_i, y_i, z_i) (i = 1, 2, 3, ..., n) and constructing some mathematical model describing the field variation in space and time. As was already mentioned, for this purpose we use a mixed four-dimensional dynamic-stochastic model of the form

$$\xi'_{i,h}(k) = \sum_{j=1}^{K} a_j \xi'_{i,h}(k-j) + \sum_{m=1}^{M} b_m \xi'_{i,m}(k) + \sum_{s=1}^{S} \frac{c_s \xi'_{s,h}(k)}{\rho_{is}}, \quad (1)$$

where $k = 0, 1, 2, \ldots$ is the discrete time with the discretization interval Δt, ($t_k = k\Delta t$); K is the order of time lag, which determines the depth of autoregression; $\grave{I}$ is the number of altitude levels taking part in formation of the estimate of the field ξ'; S is the number of observation sites; a_j, b_m, and c_s are the unknown parameters to be estimated, which determine the time, height, and space dependence between the field measurements at different instants of the discrete time k at different levels and at different points of a mesoscale region, respectively; $\rho_{is} = \rho_0 / (\rho_0 - R_{is})$ is the normalizing coefficient, which determines the mutual arrangement of observation points on the plane within the mesoscale region and reflects the presence of spatial correlation between them (here ρ_0 is the spatial correlation length and $R_{is} = \sqrt{(x_i - x_s)^2 + (y_i - y_s)^2}$ is the distance between the points i and s, in km).

According to Refs. 1 and 8, in the planetary boundary layer (at $h < 1.5$ km), the spatial correlation length ρ_0 is 2000 km (for temperature) and 750 km (for orthogonal components of wind velocity), while for the free atmosphere it is, respectively, 2500 and 1000 km.

It follows from Eq. (1) that the ξ' value at any point of the region at the chosen height level parametrically depends on the field values in the previous instants to the depth K (connected with the temporal correlation length), on its values at all height levels up to the level M at the instant k (time of observation), and on all measured values of the filed at the chosen level h at other observation points.

Since Eq. (1) uses the centered field ξ' as the initial one, let us dwell on the procedure of centering before going on to consideration of the technique of its estimation at some spatial point with the coordinates (x_0, y_0). For this purpose, we represent ξ as a sum of regular $\bar{\xi}$ and fluctuating ξ' components, that is, $\xi = \bar{\xi} + \xi'$.

To estimate the regular component of $\bar{\xi}$ at the ith points of the given mesoscale region, i.e., at the measurement points, and at the hth level, we use the region-averaged value of this field determined from

$$\bar{\xi}_n^{(h)} = \frac{1}{S}\sum_{i=1}^{S}\xi_i^{(h)}, \qquad (2)$$

where $\xi_i^{(h)}$ is the measured value of the meteorological field at the ith station and the hth height level.

At the same time, to estimate the regular component of $\bar{\xi}$ at the point of extrapolation (x_0, y_0) at the given height level h, we use the weighted mean value calculated from the measurements at three the closest (to the point of extrapolation) stations by the equation

$$\bar{\xi}_0^{(h)} = \sum_{i=1}^{3} q_i \xi_i^{(h)} \bigg/ \sum_{i=1}^{3} q_i , \qquad (3)$$

where $\xi_i^{(h)}$ is the measured value of ξ at the ith point (or station) and at the hth height level; $q_i = 1 - \left(R_{i0} \Big/ \sum_{j=1}^{3} R_{j0}\right)$ is the weighting coefficient (here R_{i0} is the distance from the ith station to the point of extrapolation (x_0, y_0), x and y are the rectangular coordinates of the station).

To find the values of the centered field ξ_i' at the measurement points at every step of estimation of ξ_i (hereinafter the superscript h is omitted for simplicity), we can use the following equation:

$$\xi_i' = \xi_i - \bar{\xi}_i , \qquad (4)$$

while to estimate ξ at the point of extrapolation (x_0, y_0) with the chosen algorithm, we can use the equation

$$\xi_0 = \xi_0' + \bar{\xi}_0 . \qquad (5)$$

Consider now the technique of spatial extrapolation based on the use of the Kalman filter and the model (1).

In the case of application of the four-dimensional difference dynamic-stochastic model, the problem of ξ' estimation at some point with the coordinates (x_0, y_0) breaks into two stages. At the first stage, the values of ξ' obtained at the measurement points are used to estimate the model coefficients a_j, b_m, and $\tilde{n}_s$, which are constant in the meaning of average because the field is assumed homogeneous and isotropic for the given mesoscale region. At the second stage, the determined coefficients are used in the mathematical model (1) to reconstruct the values of the centered field at the given spatial point and at different height levels.

According to Ref. 9, to estimate the unknown parameters of the model (1), i.e., a_j, b_m, and $\tilde{n}_s$, it is necessary to set a system of difference equations in the matrix form:

$$\mathbf{X}(k+1) = \mathbf{F}(k) \cdot \mathbf{X}(k) + \mathbf{\Omega}(k), \qquad (6)$$

where $\mathbf{X}(k+1) = |a_1(k+1),\ a_2(k+1), \ldots, a_K(k+1); b_1(k+1), b_2(k+1), \ldots, b_M(k+1); c_1(k+1), c_2(k+1), \ldots, c_S(k+1)|^{\mathrm{T}} = |X_1(k+1),\ X_2(k+1), \ldots, X_K(k+1), \ldots, X_{K+M}(k+1), \ldots, X_{K+M+S}(k+1)|^{\mathrm{T}}$ is a column $(n \times 1) = ((K + M + S) \times 1)$ vector, including the unknown variables of the dynamic system state (state vector); Ò denotes transposition; $\mathbf{F}(k)$ is the $(n \times n)$ transition matrix for the discrete system; $\mathbf{\Omega}(k) = |\omega_1, \omega_2, \ldots, \omega_n|^{\mathrm{T}}$ is a column vector of random perturbations of the system (vector of state noise).

If we assume that the considered meteorological field is isotropic and stationary, and at the given time interval the unknown parameters $\mathbf{X}(k)$ to be estimated are, on the average, unchanged, then

$$\mathbf{X}(k + 1) = \mathbf{X}(k), \quad (7)$$

or

$$\mathbf{X}(k + 1) = \mathbf{F}(k)\ \mathbf{X}(k). \quad (8)$$

Thus, in our case the transition matrix $\mathbf{F}(k)$ corresponds to the $(n \times n)$ unit matrix $\mathbf{I}$:

$$\mathbf{F}(k) = \mathbf{I} = \begin{vmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \end{vmatrix}. \quad (9)$$

The mathematical model of measurements, which are used in the Kalman filtering algorithm to estimate the system state, is generally described by the additive mixture of the useful message and the measurement error:

$$\mathbf{Y}'(k) = \xi'(k) = \mathbf{H}(k) \cdot \mathbf{X}(k) + \mathbf{E}(k)\ , \quad (10)$$

where $\mathbf{Y}'(k)$ is the $(s \times 1)$ vector of actual measurements; $\mathbf{H}(k)$ is the $(s \times n)$ matrix of observations, determining the functional relation between the true values of the state variables and the actual measurement; $\mathbf{Å}(k)$ is the vector of measurement errors (measurement noise).

The vector of actual measurements $\mathbf{Y}'(k)$ at the time k includes the whole set of the measured values of the meteorological field of interest at S stations and at all height levels. This vector can be written in the form

$$\mathbf{Y}'(k) = |Y'_{11}(k),\ Y'_{12}(k),\ Y'_{13}(k),\ \ldots,\ Y'_{1,M}(k),$$
$$Y'_{21}(k),\ Y'_{22}(k), \ldots, Y'_{2,M}(k), \ldots Y'_{S,M}(k)|^{\grave{O}}, \quad (11)$$

where the first element in the subscripts is the number of the observation point (station), while the second one is the number of the height level included in the consideration. Thus, the vector of measurements includes a sequential series of vertical profiles measured from all observation points involved in consideration, and $s = S \times M$.

Specify the matrix of observations $\mathbf{H}(k)$. When comparing the equations for the basis function (1) and the mathematical model (10), we can see that the elements of $\mathbf{H}(k)$ are the measurements of ξ' at the observation points at all heights at this and previous instants (to the depth K). The matrix $\mathbf{H}(k)$ has a three-block structure. The first $(n \times K)$ block includes the values of the meteorological parameter at the previous (to the depth K) instants. As the discrete time k changes and new data come from the measurement stations, the elements of $\mathbf{H}(k)$ move step-by-step inside the first block from the left to the right, thus forming the moving window of the autoregression process with the effective width K.

The second $n \times (M + 1)$ part of $\mathbf{H}(k)$ is formed by the blocks of angular matrices, whose upper parts consists of zeros, while the lower ones include the field values measured at neighboring levels. The third $(n \times S)$ part of $\mathbf{H}(k)$ is filled with the field values measured at all observation points at the current instant k.

Once all the elements entering into Eqs. (6) and (10) are determined, the estimation problem is solved with the aid of the linear Kalman filter, which provides for estimation of the state vector elements with minimal rms errors.

The estimation equations have the following form[9]:

$$\hat{\mathbf{X}}(k+1) = \hat{\mathbf{X}}(k+1|k) +$$
$$+\mathbf{G}(k+1) \cdot [\mathbf{Y}(k+1) - \mathbf{H}(k+1) \cdot \hat{\mathbf{X}}(k+1|k)], \quad (12)$$

where

$$\hat{\mathbf{X}}(k+1) = |\ \hat{X}_1(k+1), \hat{X}_2(k+1), \ldots,\ \hat{X}_n(k+1)\ |^{\mathrm{T}}$$

is the estimate of the state vector at the instant $(k + 1)$, $\hat{\mathbf{X}}(k+1|k)$ is the vector of estimates predicted for the instant $(k + 1)$ from the data at the step k; $\mathbf{G}(k + 1)$ is the $(n \times s)$ matrix of weighting coefficients.

Note that $\hat{\mathbf{X}}(k+1)$ and $\hat{\mathbf{X}}(k+1|k)$ are $(n \times 1)$ vectors, and the matrix equation for calculation of the vector of forecast is

$$\hat{\mathbf{X}}(k+1|k) = \mathbf{F}(k) \cdot \hat{\mathbf{X}}(k). \quad (13)$$

The weighting coefficients in the Kalman filter are calculated using the standard equations for calculation of covariance of estimation errors.[9]

To start the filtering algorithm (12) under the conditions of lacking *a priori* information, the initial values of the coefficients should be taken zero: $a_i = 0$, $b_i = 0$, $c_i = 0$. Thus, $\hat{\mathbf{X}}(0) = 0$. Other initial conditions connected with estimation of elements of the state noise and observation matrices are set based on the known values of standard deviations and errors of radiosonde data.

Thus, in the course of processing of measurements at the instants k, the estimate of the state vector is forming

$$\hat{\mathbf{X}}(k) = \left|\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_K, \hat{b}_1, \hat{b}_2, \ldots, \hat{b}_M, \hat{c}_1, \hat{c}_2, \ldots, \hat{c}_S,\right|, \quad (14)$$

with the use of Eqs. (12) and the given initial conditions. This estimate is used for direct reconstruction of the meteorological field ξ at the point of extrapolation based on the equation

$$\hat{\Xi}_0(k) = \bar{\Xi}(k) + \hat{\mathbf{Y}}'_0(k) = \bar{\Xi} + \hat{\mathbf{H}}(k) \times \hat{\mathbf{X}}(k), \quad (15)$$

where $\bar{\Xi}$ and $\hat{\Xi}_0(k)$ are the vectors of weighted-mean and reconstructed values of meteorological parameters at the point (x_0, y_0); $\hat{\mathbf{X}}(k)$ is the obtained estimate of the state vector at the instant k; $\hat{\mathbf{Y}}'_0(k)$ is the vector of estimates of the field fluctuation component at the point of extrapolation; $\hat{\mathbf{H}}(k)$ is the $(M \times n)$ transition matrix for reconstruction of the same field at the point (x_0, y_0).

2. Results of studies of the Kalman filtering algorithm

The above algorithm was subjected to the quality assessment as applied to the problem of spatial extrapolation of mesoscale temperature and wind fields.

Since the spatial extrapolation is considered here as applied to prediction of an industrial pollution cloud propagation, we took the layer-average values in some height range $h_k - h_0$ (instead of wind and temperature measurements at individual levels), where $h_0 = 0$ corresponds to the ground level, and h_k is the height of the top boundary of the studied kth atmospheric layer. The layer-average (or simply average) values of temperature and zonal and meridional wind components were calculated as

$$<\xi>_{h_k-h_0} = \sum_{i=1}^{k}\left\{\left(\frac{\xi_{h_{i-1}}+\xi_{h_i}}{2}\right)\left(\frac{h_i-h_{i-1}}{h_k}\right)\right\}, \qquad (16)$$

where $<\cdot>$ denotes averaging of observations in some atmospheric layer, and ξ is the measured value of the meteorological parameter at different atmospheric levels.

To assess the quality of the Kalman filtering algorithm, we used the archive of data of two-year (2000–2001) two-time (0 and 12 GMT) radiosonde observations at eight aerological stations: Moscow (55°45′ N, 37°57′ E), Smolensk (54°45′ N, 32°04′ E), Bologoe (57°54′ N, 34°03′ E), Vologda (59°19′ N, 39°55′ E), Nizhnii Novgorod (56°16′ N, 44°00′ E), Ryazan (54°38′ N, 39°42′ E), Sukhinichi (54°06′ N, 35°21′ E), and Kursk (51°46′ N, 36°10′ E), forming a typical mesoscale region (Fig. 1). All the observations of temperature and wind presented for winter and summer on standard isobaric surfaces and singular-point levels were reduced, using linear interpolation, to the uniform system of the following geometric heights: 0 (ground level), 0.2, 0.4, 0.8, 1.2, 1.6, 2.0, 3.0, 4.0, 5.0, 6.0, and 8.0 km. This system of geometric heights allows describing almost all the troposphere and, especially, the boundary layer with high vertical resolution.

To assess the accuracy of the Kalman filtering algorithm (the results are shown in Fig. 2), station Smolensk, spaced by 225 km from the closest neighboring station Sukhinichi, was used as a control point (to which the spatial extrapolation was carried out).

Analysis of Fig. 2 indicates that the proposed algorithm based on the Kalman filtering and the four-dimensional dynamic-stochastic model gives rather good results as applied to the spatial extrapolation of the layer-average values of temperature and orthogonal wind components to the distance up to 225 km. Actually, at this distance regardless of the season and the atmospheric layer, the rms errors of extrapolation vary within 0.7–1.6°Ñ (for the mean temperature) and 1.0–2.2 m/s (for zonal and meridional components of the mean wind).

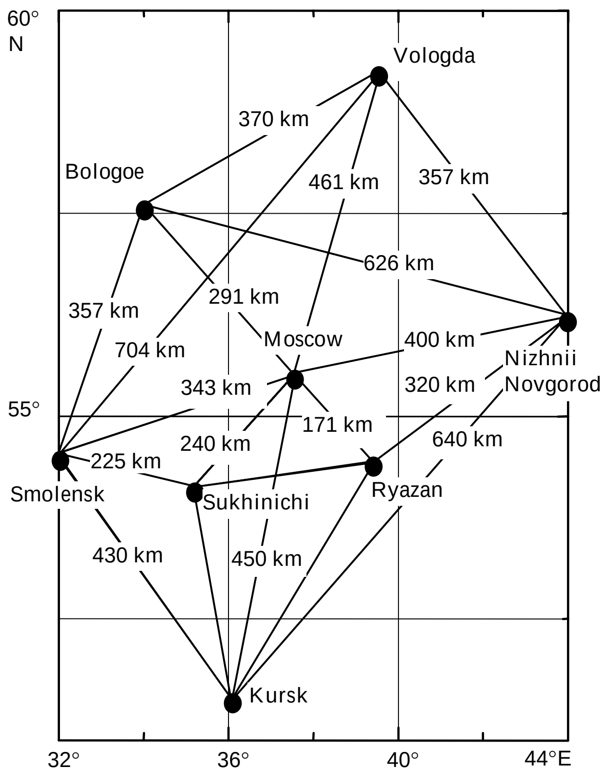

Fig. 1. Map of mesoscale region.

In addition, this algorithm allows extrapolation of the wind field in the atmospheric boundary layer, i.e., 0–1.6 km layer (where, according to Ref. 10, the principal transport of industrial pollutants takes place) with a rather high accuracy (rms error about 1.0–2.0 m/s), which is close to the accuracy of wind radiosounding, being 0.7–2.0 m/s [Ref. 11].

In conclusion, it should be emphasized that the method considered can be improved. Toward this end, it is necessary to develop an adaptive algorithm, which would allow estimating the correlation length of a chosen meteorological parameter at every observation point and fitting, in the proper way, the coefficients c_s in the model (1). An individual subject of study is also the influence of parameters determining the temporal and spatial dependences of the meteorological field at the point of forecast on the values of this field at the observation points (the first and second terms in Eq. (1)).

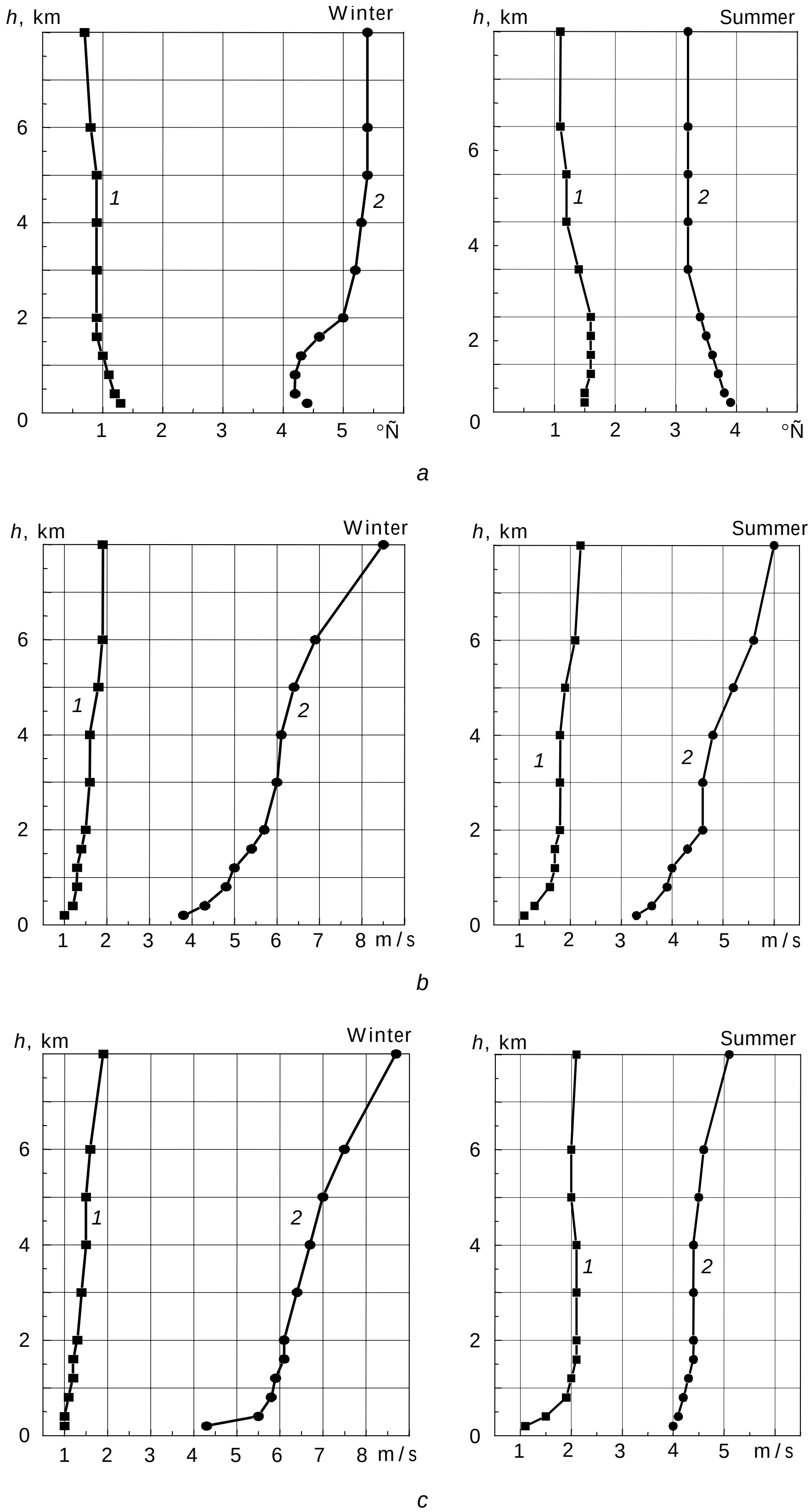

Fig. 2. Vertical profiles of rms errors of extrapolation of the layer-average values of temperature (*a*), zonal (*b*) and meridional (*c*) wind components to the distance 225 km with the use of the Kalman filtering algorithm and the four-dimensional dynamic-stochastic model (*1*) along with the corresponding standard deviations (*2*).

References

1. L.S. Gandin and R.L. Kagan, *Statistical Methods of Interpretation of Meteorological Data* (Gidrometeoizdat, Leningrad, 1976), 359 pp.
2. V.A. Gordin, *Mathematical Problems of Hydrometeorological Weather Forecasting* (Gidrometeoizdat, Leningrad, 1987), 264 pp.
3. V.S. Komarov and Yu.B. Popov, Atmos. Oceanic Opt. **14**, No. 4, 230–234 (2001).
4. V.S. Komarov, S.N. Il'in, B.P. Kuznetsov, Yu.B. Popov, and A.I. Popova, Atmos. Oceanic Opt. **15**, Nos. 5–6, 433–436 (2002).
5. D.P. Dee, Quart. J. Roy. Meteorol. Soc. **117**, 365–384 (1991).
6. E.G. Klimova, Meteorol. Gidrol., No. 11, 55–65 (1997).
7. E.G. Klimova, Meteorol. Gidrol., No. 10, 24–33 (2001).
8. S. Panchev, *Random Functions and Turbulence* (Gidrometeoizdat, Leningrad, 1967), 447 pp.
9. A.P. Sage and J.L. Melsa, *Estimation Theory with Applications to Communications and Control* (McGraw-Hill, New York, 1971).
10. A.M. Vladimirov, Yu.I. Lyakhin, L.T. Matveev, and V.G. Orlov, *Environmental Protection* (Gidrometeoizdat, Leningrad, 1991), 423 pp.
11. *Guide to Meteorological Instrument and Observing Practices* (WMO, Paris, 1984), 130 pp.