Estimation of the potential accuracy of the algorithm for meteorological data extrapolation by the state space method

Yu.B. Popov and A.I. Popova

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

Received November 9, 2005

An analytical formula is derived for variance of the extrapolation error allowing a computation of the spatial prognosis error of some meteoparameter at any initial conditions and any measurement interval. For the analysis, the filtering algorithm is chosen, which uses a spatial polynomial of the second order as the observational model.

Earlier, we have proposed the technique for synthesis of algorithms of spatial extrapolation of meteorological parameters on the basis of the Kalman filtering.[1,2] Accuracy of all algorithms was studied using actual aerologic data. The final estimation error depended on conditions, size, and characteristics of the particular measurement sample. There appears a rightful question about potential capabilities of the algorithms and temporal behavior of extrapolation errors as the data enter. In this paper we consider the analytical formula for the variance of the extrapolation error, which allows one to calculate the error of spatial forecast of a meteorological parameter for any measurement period and different initial conditions. The filtering algorithm was selected for the analysis, using the spatial polynomial of the second order as an observation model.

Statement of the problem

Let the meteorological parameter $\xi_i(t)$ at the ith point of the given plane at the moment t be determined by the second order polynomial

$$\xi_i(t) = X_1(t) + X_2(t)x_i + X_3(t)y_i + X_4(t)x_i y_i + X_5(t)x_i^2 + X_6(t)y_i^2, \quad (1)$$

where x_i and y_i are the Cartesian coordinates of the measurement stations, $X_1(t) - X_6(t)$ are unknown coefficients determining the meteorological parameter value at each moment for any point within the limits of a mesoscale testing area. Unknown coefficients of the polynomial are estimated by means of the Kalman filter discrete variant. The state vector at the discrete moments t_k has the form

$$\mathbf{X}(k) = |X_1(k), X_2(k), X_3(k), X_4(k), X_5(k), X_6(k)|^{\mathrm{T}}. \quad (2)$$

The symbol T means here a transposition. The dynamics of variation of the vector components (2) can be described by a system of difference equations of the form ($l = 1, ..., 6$)

$$X_l(k+1) = X_l(k) + \omega_l(k), \quad (3)$$

where

$$\mathbf{\Omega} = \left| \omega_1(k)\ \omega_2(k)\ \omega_3(k)\ \omega_4(k)\ \omega_5(k)\ \omega_6(k) \right|^{\mathrm{T}}$$

are the random perturbations of the system (generating noises of states). The measurement model is presented in the form of an additive mixture of the true value of the meteorological parameter $\xi_i(k)$ and the measurement error $\varepsilon_i(k)$:

$$Y_i(k) = \xi_i(k) + \varepsilon_i(k). \quad (4)$$

In terms of variables of the state (2), the measurement model (4) can be rewritten as[2]

$$Y_i(k) = X_1(k) + X_2(k)x_i + X_3(k)y_i + X_4(k)x_i y_i + X_5(k)x_i^2 + X_6(k)y_i^2 + \varepsilon_i(k). \quad (5)$$

Equations (3) and (5) completely determine the structure of the linear Kalman filter providing the estimation of the polynomial (1) coefficients with a minimal rms error. The algorithm of spatial extrapolation into the preset point j of the testing area has the form

$$\hat{Y}_j(k) = \hat{X}_1(k) + \hat{X}_2(k)x_j + \hat{X}_3(k)y_j + \hat{X}_4(k)x_j y_j + \hat{X}_5(k)x_j^2 + \hat{X}_6(k)y_j^2, \quad (6)$$

where $\hat{X}_i(k)$, $\hat{Y}_j(k)$ are the estimate of the state vector and the extrapolated value of the meteorological parameter at the moment k; x_j, y_j are coordinates of the extrapolation point. The variance of the error in estimating (spatial extrapolation) of the meteorological variable $Y_j(k)$ at each step of the

0235-6880/06/04 311-03 $02.00

forecast averaged over the ensemble of realizations is determined by the formula

$$D_Y(k) = \mathrm{E}\left\{\left[Y_j(k) - \hat{Y}_j(k)\right]^2\right\}, \qquad (7)$$

where E is the operator of mathematical expectation, $Y_j(k)$ is the true value of the forecasted variable; $\hat{Y}_j(k)$ is the estimated value of the forecasted variable. Substituting Eq. (6) to Eq. (7), one can obtain the dependence of the variance of the error in estimating $Y_j(k)$ on the errors of estimating the coefficients of the polynomials $X_i(k)$:

$$D_Y(t) = D_{11}(t) + x_j^2 D_{22}(t) + y_j^2 D_{33}(t) + x_j^2 y_j^2 D_{44}(t) + x_j^4 D_{55}(t) + y_j^4 D_{66}(t) + 2x_j D_{12}(t) + 2y_j D_{13}(t) + x_j y_j D_{14}(t) + 2x_j^2 D_{15}(t) + 2y_j^2 D_{16}(t) + 2x_j y_j D_{23}(t) + 2x_j^2 y_j D_{24}(t) + 2x_j^3 D_{25}(t) + 2x_j y_j^2 D_{26}(t) + 2x_j y_j^2 D_{34}(t) + 2x_j^2 y_j D_{35}(t) + 2y_j^3 D_{36}(t) + 2x_j^3 y_j D_{45}(t) + 2x_j y_j^3 D_{46}(t) + 2x_j^2 y_j^2 D_{56}(t), \qquad (8)$$

where $D_{lm}(k)$ are the elements of the covariation matrix of the errors in estimating the coefficients $\mathbf{D}(t_k)$ (2).

Technique for solving the problem

The covariation matrix of errors in estimating the elements of the state vector for the linear Kalman filter can be calculated *a priori* by means of the matrix Riccati differential equation[3,4]:

$$\frac{d\mathbf{D}}{dt} = \mathbf{F}(t)\mathbf{D}(t) + \mathbf{D}(t)\mathbf{F}^{\mathrm{T}}(t) + \mathbf{G}(t)\mathbf{R}_{\Omega}(t)\mathbf{G}^{\mathrm{T}}(t) - \mathbf{D}(t)\mathbf{H}^{\mathrm{T}}(t)\mathbf{R}_{\varepsilon}^{-1}(t)\mathbf{H}(t)\mathbf{D}(t), \qquad (9)$$

where $\mathbf{F}(t) = 0$ is the transitional matrix of states, $\mathbf{G}(t)$ is the unit transitional matrix of state noises ($\mathbf{diag}\ \mathbf{G}(t) = |1\ 1\ 1\ 1\ 1\ 1|$); $\mathbf{R}_{\Omega}(t) = 0$ is the covariation matrix of state noises under condition that the noises are absent (3);

$$\mathbf{H} = \begin{vmatrix} 1 & x_{11} & y_{11} & x_{11}\cdot y_{11} & (x_{11})^2 & (y_{11})^2 \\ 1 & x_{22} & y_{22} & x_{22}\cdot y_{22} & (x_{22})^2 & (y_{22})^2 \\ 1 & x_{33} & y_{33} & x_{33}\cdot y_{33} & (x_{33})^2 & (y_{33})^2 \\ 1 & x_{44} & y_{44} & x_{44}\cdot y_{44} & (x_{44})^2 & (y_{44})^2 \\ 1 & x_{55} & y_{55} & x_{55}\cdot y_{55} & (x_{55})^2 & (y_{55})^2 \\ 1 & x_{66} & y_{66} & x_{66}\cdot y_{66} & (x_{66})^2 & (y_{66})^2 \end{vmatrix}$$

is the transitional matrix of the measurements, $\mathbf{R}_{\varepsilon}(t)$ is the covariation matrix of the observation noises ($\mathbf{diag}\ \mathbf{R}_{\varepsilon}(t) = |\sigma_{\varepsilon}^2\ \sigma_{\varepsilon}^2\ \sigma_{\varepsilon}^2\ \sigma_{\varepsilon}^2|$).

To solve Eq. (9), it is necessary to set the initial value of the matrix of the error variances in estimating $\mathbf{D}(0)$, which is equal to the matrix of variances of the estimated process $\mathbf{X}(t)$ at the moment $t = 0$. The value of this matrix at synthesis of the algorithms is set based on *a priori* data. One of the ways of analytical solution of Eq. (9) described in detail in Ref. 4 is used in this paper. The principle of solution is based on replacing Eq. (9) by the system of linear differential equations with the matrix $2n\times 2n$ (where n is the dimension of the state vector). The closed solution of the Riccati equation in the matrix form is presented below:

$$\mathbf{D}(t) = \left[\mathbf{C}_{11}(t)\mathbf{D}(0) + \mathbf{C}_{12}(t)\right] \times \left[\mathbf{C}_{21}(t)\mathbf{D}(0) + \mathbf{C}_{22}(t)\right]^{-1}. \qquad (10)$$

The following formula is used to determine the block matrix $\mathbf{C}(t)$

$$\mathbf{C}(t) = \begin{vmatrix} \mathbf{C}_{11}(t) & \mathbf{C}_{12}(t) \\ \mathbf{C}_{21}(t) & \mathbf{C}_{22}(t) \end{vmatrix} = \mathcal{L}^{-1}\{[s\mathbf{I} - \mathbf{A}]^{-1}\}, \qquad (11)$$

where $\mathcal{L}^{-1}$ is the operation of the inverse Laplace transform; $\mathbf{I}$ is the unit matrix $2n\times 2n$, s is the Laplace transform parameter

$$\mathbf{A} = \begin{vmatrix} \mathbf{F} & \mathbf{G}\mathbf{R}_{\Omega}\mathbf{G}^{\mathrm{T}} \\ \mathbf{H}^{\mathrm{T}}\mathbf{R}_{\varepsilon}^{-1}\mathbf{H} & -\mathbf{F}^{\mathrm{T}} \end{vmatrix}$$

is the auxiliary block matrix.

The block matrix $\mathbf{A}$ for the models (3) and (5) has the following form:

$$\mathbf{A} = \begin{vmatrix} 0 & 0 \\ \mathbf{H}^{\mathrm{T}}\cdot\mathbf{R}_{\varepsilon}^{-1}\cdot\mathbf{H} & 0 \end{vmatrix}. \qquad (12)$$

Results

The rms error in extrapolation $\sigma_Y = \sqrt{D_Y(t)}$ was investigated for three configurations of the testing area (Fig. 1*a*) including the stations: (1) Bologoe – Sukhinichi – Smolensk – Ryazan' – Moscow; (2) Smolensk – Sukhinichi – Kursk – Ryazan' – Moscow; (3) Sukhinichi – Kursk – Ryazan' – Nizhnii Novgorod – Moscow.

The initial covariation matrix of errors was set in the diagonal form: $\mathbf{D}(0) = \sigma^2\mathbf{I}$, where $\mathbf{I}$ is the unit matrix of the dimension $(n\times n)$. The investigation period was 5 days with the rate of arrival of the measurements equal to 12 hours.[1] Results of investigations of the initial conditions $\mathbf{D}(0)$ effect on σ_Y (at $\sigma = 1$, 2, 3 and $\sigma_{\varepsilon} = 1$) are shown in Fig. 1*b* for configuration 1. Results of studying the σ_Y variation for different stations and testing area configurations (at $\sigma = 2$ and $\sigma_{\varepsilon} = 1$) are presented in Fig. 1*c*. The plots of σ_Y variations for all considered configurations with the point of extrapolation at the station Moscow are shown in Fig. 1*d*. The performed study of the effect of the σ_{ε} measurement error on σ_Y has shown its insignificance in the value range of $\sigma_{\varepsilon} = 0.5$; 1; 1.5; 2 and $\sigma = 1$; 1.5; 2. All calculations used the software complex Mathcad 2001.

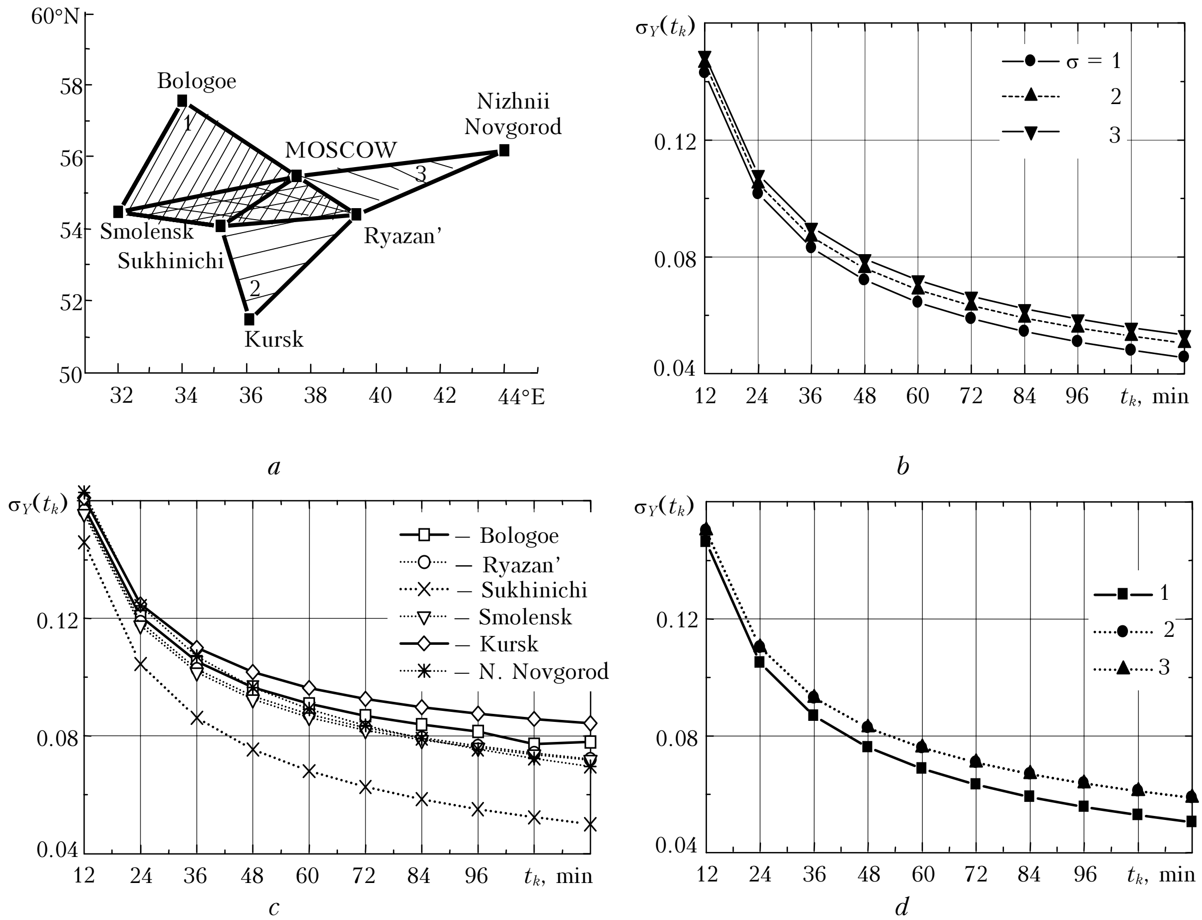

Fig. 1. The diagram of configurations of the testing area (*a*), as well as the plots of the temporal behavior of rms errors in estimating the meteorological parameter: (*b*) for the station Moscow (testing area 1) at $\sigma = 1, 2, 3$ and $\sigma_\varepsilon = 1$; (*c*) for the stations Bologoe, Ryazan', Sukhinichi, Smolensk (testing area 1); for the station Kursk (testing area 2), for the station Nizhnii Novgorod (testing area 3); (*d*) for the station Moscow at $\sigma_\varepsilon = 1$ and $\sigma = 2$ for three configurations of the testing area (1, 2, 3).

Conclusions

1) The obtained results allow one to estimate the potential accuracy of extrapolation at different values of measurement errors.

2) The value of the initial covariation matrix σ mostly affects the filter convergence rate, the minimum σ_Y is reached at $\sigma = 1.0$.

3) The study of the effect of the error in measuring σ_ε on σ_Y did not reveal its significance in the value range of $\sigma_\varepsilon = 0.5$; 1; 1.5; 2 and $\sigma = 1$; 1.5; 2.

4) The minimum error σ_Y is obtained for the stations Moscow and Sukhinichi situated inside the selected testing areas, that well agrees with the data from Ref. 5.

References

1. V.S. Komarov, Yu.B. Popov, V.A. Kurakov, V.V. Kurushev, and A.I. Popova, Atmos. Oceanic Opt. **15**, No. 11, 915–919 (2002).

2. V.S. Komarov, Yu.B. Popov, and A.I. Popova, Atmos. Oceanic Opt. **17**, No. 7, 500–504 (2004).

3. N.D. Assimakis, D.G. Lainiotis, and F.L. Sanida, Nonlinear Analysis Theory, Methods & Applications **30**, No. 4, 2409–2420 (1997).

4. S.V. Pervachev, *Radioautomatics* (Radio i Svyaz', Moscow, 1982), 295 pp.

5. L.S. Gandin and R.L. Kagan, *Statistical Methods for Interpretation of Meteorological Data* (Gidrometeoizdat, Leningrad, 1976), 359 pp.