Dynamic probabilistic method for four-dimensional analysis of meteorological elements fields

A.V. Protasov

Institute of Computational Mathematics and Mathematical Geophysics, Siberian Branch of the Russian Academy of Sciences, Novosibirsk

Received March 3, 2007

A method of reconstruction of meteorological element fields in nodes of spatial-temporal regular net from data of observational stations is considered. The method is based on interpretation of the sought fields as expansion into finite series over functions of natural orthogonal basis, which are calculated over the ensemble of spatial-temporal realizations computed by the dynamic probabilistic method at a preassigned actual statistic structure of the meteoelement fields. In fact, the method of fast assimilation of observational data is proposed.

The problem of analysis and interpretation of actual information is one of the most important problems, which appears when constructing mathematic models for physical processes, weather forecast, general circulation in the atmosphere and ocean, climate theory, as well as when studying and estimating the human activity effects on the environment. One of the points of this problem is in working out of methods of "compressing" the information and separation of its most informative part as a sum of a finite Fourier series with a small number of terms.

In this paper, a method of four-dimensional analysis of data is proposed based on climatic ensemble of possible realizations of corresponding multidimensional hydrometeorological fields for some chosen time interval and given region in the form[1,2]:

$$\{\xi^i_{(n)}, i = 1,2,\ldots\}, \tag{1}$$

where

$$\xi^i_{(n)} = [\mathbf{U}^i(\mathbf{X}_j,t_k), T^i(\mathbf{X}_j,t_k), H^i(\mathbf{X}_j,t_k),\ldots]^{\mathrm{T}}$$

is the realization vector of fields of the speed, temperature, geopotential, and others in spatial-temporal points $(\mathbf{X}_j, t_k)$ of the considered net region of n dimensionality; the index T defines the transposition operation.

To construct this ensemble, the dynamical probabilistic method is used, which is described in detail in Refs. 1–3. A feature of such an approach is the use in the limits of a uniform model of actual measurements, statistical modeling, and numerical model of atmospheric hydrothermodynamics. The basic connecting component in this case is variational assimilation of information by the hydrodynamic model.

The data of the NCEP/NCAR temperature field reanalysis, 1948–2005, were used as the actual ones at 10 standard levels for winter season with a time discreteness of 6 h and a horizontal one of $2.5\times2.5°$. The selection was conducted from a given North hemisphere local region Ω of $10\times10°$ size, centered at a point with coordinates 60.56° N and 77.7° E. The problem was considered in the coordinate system x, y, p in the region, the bottom base of which was a rectangle on a plane tangential at this central point. When constructing the net region, a resolution of 24×20 along x and y with steps $\Delta x = 23.85$ km and $\Delta y = 58.74$ km, respectively, was chosen.

The initial stage in constructing the climatic ensemble (1) is the building of the corresponding ensemble of realizations

$$\{\xi^i_{(n)_c}, i = 1,2,\ldots\} \tag{2}$$

with the use of the following statistical modeling method.[1]

Let R be a multidimensional correlation matrix. In our case, the correlation matrix R is calculated by the reanalysis data for the above region. The matrix spectral decomposition is presented as

$$R = W\Lambda W^{\mathrm{T}}, \tag{3}$$

where W is the matrix of eigenvectors of the correlation matrix R; Λ is the diagonal matrix of corresponding eigenvalues. Note that representation (3) is just its decomposition into so called main factors, and Equation (3) is the corresponding problem of determination of the main factors. The further step is determination of the R square root in the form $R^{1/2} = W\Lambda^{1/2}W^{\mathrm{T}}$, where $\Lambda^{1/2}$ is the diagonal matrix with square roots of corresponding matrix eigenvalues on the diagonal.

Then we can determine the random vector

$$\xi^{(i)}_{(n)_c} = D_\xi R^{1/2}\psi^{(i)}(x_j,y_j,p_j,t_j)^{\mathrm{T}} + \bar{\xi}(x_j,y_j,p_j,t_j),$$
$$i = 1,2,\ldots, \tag{4}$$

where $\psi^{(i)}(x_j,y_j,p_j,t_j)^{\mathrm{T}}$, $i = 1,2,\ldots$ is the Gaussian random vector with a unit dispersion and zero mean;

0235-6880/07/06 505-04 $02.00

D_ξ is the diagonal matrix of dispersions; $\bar{\xi}(x_j, y_j, p_j, t_j)$ is the corresponding vector of means. It is easy to see that the correlation matrix of the random vector $\xi^{(i)}_{(n)_c}$ exactly coincides with R.

Note that in the general case the realizations from the ensemble (2) are determined on some set of irregular spatial-temporal points. This is fully specified by measurements used in calculating R. Besides, the ensemble (2) includes, as a rule, not all fields necessary for numerical models of atmosphere dynamics, and the components (temperature, speed, pressure, etc.) of the ensemble (2) are not consistent with the corresponding numerical model.

Thus, there appears a problem of the best approximation of the ensemble (2) by the corresponding ensemble of realizations, in which every realization would conform to the hydrothermodynamics numerical model and corresponding statistic properties would maximally close.

To do this, we used the variation assimilation method.[3] It should be noted that the variation assimilation is to be applied within the whole considered time interval and within the limits of predictability of the model. Because of nonlinearity of the initial numerical model of the atmospheric hydrodynamics and, as a consequence, nonuniqueness of solution of the problem of minimum of the considered quality functional, as well as by virtue of the fact that in general case the prove of the convergence and uniqueness of the obtained solutions is absent, it is necessary very accurately introduce additional minimizable fuctionals into the model, such as the model quality or model measurement functionals, or others. Therefore, in each particular case, additional theoretic or numerical investigations of the efficiency of such introductions of functionals or equations determining corresponding links between the components under consideration are necessary.

Besides, to construct the climatic ensemble of realizations, it is important to consider the problem of variational assimilation on the whole time interval, because the use of the so called sequential step-by-step assimilation does not provide the necessary smoothness of solutions and the corresponding trend for further use, for example, the obtained field in the prediction mode.

Thus, to build the finite climatic ensemble of realizations (1) we apply the variational assimilation. To do this, for each realization from ensemble (2) the variation assimilation problem is solved with the use of the mathematical model of atmosphere hydrothermodynamics,[1,2] that is resulted in an ensemble of new realizations, differing from the initial one by the accuracy of assimilation problem solution and complying with properties of mathematical model.

The dynamic probabilistic model and some its characteristics are described in detail in Ref. 3. To solve the problem, the iteration method of gradient descent based on the Lagrange method and solving direct and conjugate problems were used. In numerical calculations the ensemble (2) was presented only by temperature field realizations. However, the ensemble (1) already contains all fields of meteorological elements in accordance with the used model.[1,2] So, the model both is the spatial-temporal interpolant and allows reproducing the absent fields of meteoelements. Analysis of the ensemble (1) statistic structure shows that it can be used as climatic one when solving applied problems including admixture transfer in the atmosphere, or when studying processes of emission into the atmosphere.

In this paper we use the obtained climatic ensemble (1) to solve the problem of four-dimensional analysis of atmospheric hydrometeorological data. One of the algorithms for such application[4] is based on representation of the sought hydrometeorologic field in the form of the corresponding series over natural orthogonal functions, calculated with actual data only for geopotential's field in winter period at a rather limited sample in hand.

Since the ensemble (1) already contains statistically independent spatial-temporal realizations, including a full set of hydrometeorologic components (temperature, geopotential, speed of wind), mutually complied relative to the numerical model of the atmospheric dynamics, then the use of this method in the four-dimensional analysis and assimilation of the corresponding actual data on the whole is natural. It is clear that the statistical significance of the obtained results is fully determined by the ensemble (1).

This approach has a series of advantages. First, the basis of natural orthogonal functions, built with sufficiently large sample, has necessary properties of statistic structure of meteoelement fields, which is of importance at a rare net of stations. Second, the number of basic functions is supposed to be comparatively small, which allows constructing the efficient algorithm. In addition, it follows from methods of constructing of the natural orthogonal basis that every its function has statistically complied components, therefore, the result of reconstruction by this method has the same degree of agreement that the basic functions.

To calculate basic functions of the natural orthogonal basis (basic factors)

$$\{\varphi_i\}(i = \overline{1, m}) \tag{5}$$

over the ensemble (1), one of modifications of the algorithm[5] for generalized covariation matrix R_a of the ensemble (1) was used. Logarithms of eigenvalues characterizing the information content of the calculated basic functions, are presented in Fig. 1, which shows that at 1615 realizations in the ensemble (1) 50 basic functions are sufficient to describe the considered meteorological fields with a good accuracy ($m = 50$).

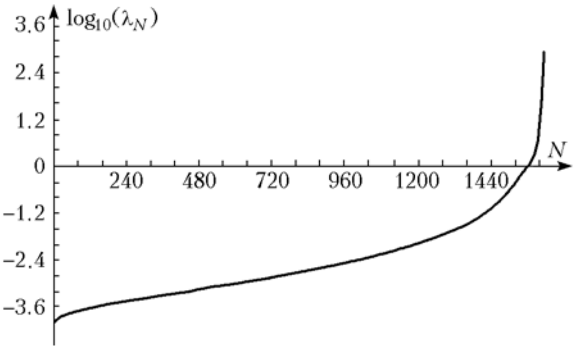

Fig. 1. Logarithms of eigenvalues of the correlation matrix calculated over the realization ensemble (1).

Thus, following Ref. 4, consider a subspace of $\tilde{R}_m$ vectors of the real vector space R_N, which components are values of the meteocomponent fields in nodes of the regular spatial-temporal net region $\Omega^{ht} \subset \Omega$. Let vector-functions (5) be the basis of $\tilde{R}_m$. Then any vector $\varphi \in \tilde{R}_m$ can be presented in the form of the Fourier series

$$\varphi = \Phi\mathbf{a}, \qquad (6)$$

where Φ is the matrix of $N \times m$ dimensionality composed of the basis vectors $\{\varphi_i\} (i = \overline{1, m})$; $\mathbf{a} = (a_1, \ldots, a_m)^{\mathrm{T}}$ is the vector of Fourier coefficients.

Let in the considered region Ω an irregular net Θ be given, in nods of which the measurement data of the meteoelement fields under study are known. Consider the subspace G of the Euclidian space determined on Θ, and take the values of the studied fields of one or several meteoelements (like in $\tilde{R}_m$) as the vector components. Introduce a scalar product in this subspace

$$(\varphi, \psi)_M = (M\varphi, \psi),$$

where $\varphi, \psi \in G$; symbol (,) denotes the scalar product in the Euclidian space; M is the positively determined symmetric matrix, the choice of which is determined by the investigation goals, physical dimensions of the vector components, and *a priory* information on the structure of the considered fields. In this case, the scalar product is a net analog of the corresponding scalar product determining the full energy integral in the used hydrothermodynamics model in solving the variation assimilation problem.

Reconstruction of meteoelement fields in nodes of Ω^{ht} from their values measured on the irregular station net in this case is reduced to finding the vector of decomposition coefficients in formula (6) in such a way that the interpolated values of $\varphi \in \tilde{R}_m$ less deviate from the corresponding measured values in Θ nodes.

Let ψ_{meas} be the vector composed of values in nodes of the irregular net region $\Theta \subset \Omega$, and φ is a vector from $\tilde{R}_m$, which should be constructed by the given vector ψ_{meas}. Denote through $\psi_{\text{meas}} = L\varphi$ the vector φ image in the subspace G, obtained with the help of the linear operator L of interpolation from the regular net to irregular one. Since $\varphi \in R_m$ can be presented in the form (6), then $\psi = L\Phi\mathbf{a}$. Consider the functional characterizing the degree of deviation of ψ_{meas} in points of irregular net of stations from values of the vector-function $\varphi \in R_m$, which are interpolated to the irregular net:

$$J = (\psi_{\text{meas}} - L\Phi\mathbf{a}, \psi_{\text{meas}} - L\Phi\mathbf{a})_M. \qquad (7)$$

From the condition of the J functional extremum we obtain a linear inhomogeneous algebraic system of equations for determination of the coefficients a_i, $i = \overline{1, m}$,

$$(L\Phi)^{\mathrm{T}} M L\Phi\mathbf{a} = (L\Phi)^{\mathrm{T}} M\psi. \qquad (8)$$

The system can be rewritten in the form

$$B\mathbf{a} = \mathbf{f}, \qquad (9)$$

where $B = (L\Phi)^{\mathrm{T}} M L\Phi$ is a symmetrical nonnegatively defined matrix; $\mathbf{f} = (L\Phi)^{\mathrm{T}} M\psi_{\text{meas}}$ is the vector of the system (8) right part.

Note that system (9) in some cases of mutual arrangement of nods of the irregular net of stations can be ill-conditioned. Therefore, the following algorithm is used to solve it.

Matrix B is presented as

$$B = W_B \Lambda_B W_B{}^{\mathrm{T}}, \qquad (10)$$

where Λ_B is the diagonal matrix of eigenvalues; W_B is the orthogonal matrix of the transform, the columns of which are eigenvectors of B.

Then, taking into account relation (10), the solution of the system (9) is obtained by the formula

$$\mathbf{a} = W_B \Lambda_B^{+} W_B{}^{\mathrm{T}} \mathbf{f},$$

where

$$\Lambda_B^{+} = \mathrm{diag}\{\lambda_i^{+}\}, \; (i = \overline{1, m})$$

is the diagonal matrix built by analogy with the pseudo-inverse matrix, namely,

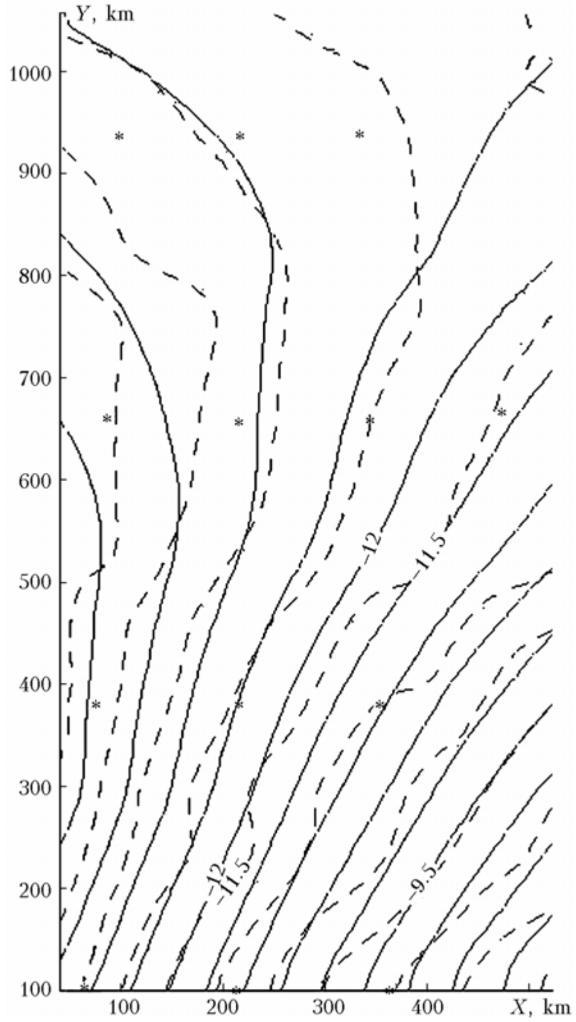

Fig. 2. Temperature field lines at a level of 500 mbar and $t = 0$, obtained after the variational assimilation of data (solid lines), given in the point defined by *, as well as the similar lines resulted from the four-dimensional analysis (dash lines) over basic factors.

$$\lambda_i^{+} = \begin{cases} 1/\lambda_i, & \text{at } \lambda_i > \varepsilon \\ 0, & \text{at } \lambda_i \le \varepsilon, \end{cases}$$

ε is some sufficiently small number.

Finally, the field φ can be retrieved using the obtained coefficient vector **a** in the regular net region Ω^{ht} by formula (6).

To illustrate the efficiency of the above method, the temperature fields in the region Θ for moments $t = 0$ and 6 h were modeled by the formula (4) at 10 standard levels. These data were input for solving the problem of variational assimilation with the help of numerical model and for four-dimensional analysis by formulas (6)–(9). Figure 2 presents the comparative calculation results for a level of 500 mbar and $t = 0$ h, which show a sufficiently good qualitative coincidence of the corresponding line fields. Maximal difference between values of these fields and data is 0.93°.

Acknowledgements

This work was supported by Russian Foundation for Basic Research (Grant No. 05-05-98000-r_ob'_a.)

References

1. A.V. Protasov, in: *International Sci. Congress "GEO-Siberia-2006,"* Novosibirsk (2006), Vol. 3, Part 2, pp. 112–118.
2. A.V. Protasov, in: *5 Int. Symp. "Mathematical Modeling of Dynamic Processes in Atmosphere, Ocean, and Solid Earth,"* Novosibirsk (2006), pp. 50–55.
3. V.A. Ogorodnikov and A.V. Protasov, Russ. J. Numer. Anal. Math. Modelling VSP **12**, No. 5, 461–479 (1997).
4. A.V. Protasov and V.V. Chekurova, Meteorol. Gydrol., No. 1, 105–109 (1983).
5. V.V. Penenko and A.V. Protasov, Izv. Akad. Nauk SSSR. Fiz. Atm. i Oceana **14**, No. 12, 1361–1365 (1978).