Problem of Assimilation the Data of Meteorological and Aerosol Observations

EKATERINA G. KLIMOVA and GDALIY S. RIVIN

Institute of Computational Technologies, Siberian Branch of the Russian Academy of Sciences, Pr. Akademika Lavrentyeva 6, Novosibirsk 630090 (Russia)

E-mail: klimova@ict.nsc.ru

Abstract

A brief review of works on the development of the system of mathematical modeling of processes in the atmosphere and on the assimilation of the data of meteorological observations is given. In addition, the works on the applicability of the Kalman filter theory to the problem of the assimilation of observation data are reviewed.

INTRODUCTION

Mathematical modeling of atmospheric processes, as well as the solution of the problem of transport and transformation of a passive air pollutants are usually based on solving the corresponding Cauchy problem for hydrothermodynamic model of atmosphere, or for transport and diffusion of a passive minor ingredient; hence, it is necessary to assign initial data in the grid points. Recovery of the values of meteorological fields in a definite region on the basis of the observation data in a given irregular ensemble of points is usually called numerical (objective) analysis of meteorological data. At present, in the leading meteorological centres of the world, the numerical analysis of meteorological data is performed with the help of three-dimensional multivariant optimal interpolation method. The optimal interpolation method is a statistical one; it involves minimization of the analysis error averaged over the ensemble [1]. The development of the methods of numerical analysis and the appearance of the asynchronous meteorological information, *i. e.* the information that arrives at time moments out of the major synoptic observation terms, have led to the formulation of the problem of recovery of spatial-temporal distribution of meteorological fields. It is generally accepted that the change of meteorological fields with time is described by the mathematical model of the atmosphere, while the problem of the recovery of spatial-temporal distribution of meteorological fields is usually referred to as the problem of adoption of meteorological data with the help of mathematical model.

Two major types of the adoption systems exist: continuous and discrete ones. In continuous systems, the assimilation of observation data (analysis step) occurs at the time moment corresponding to the observation moment; in the discrete systems, it occurs at time moments that correspond to the main synoptic observation terms. Let us list the major kinds of meteorological information transmitted over the system of global telecommunication. The information in the SYNOP code contains the data of surface observations obtained every 3 h, starting from 0:00 Greenwich, information in the TEMP code contains the data of aerological observations up to the height of 30–40 km, obtained every 6 h. The satellite observations SATEM and SATOB, as well as the airborne observations AIREP are made asynchronously, *i. e.* at arbitrary time moments. The data in the GRIB code contains results of forecasts and analyses by the National Centre Environmental Forecast of the USA (NCEP USA), European Centre of medium-term forecasts of weather, and the Hydrometeorological Centre of Russia.

Generally accepted system of the data adoption is a cycle consisting of forecast – analysis – initialization. In this system, the forecast according to the mathematical model of the atmosphere is taken as the first guess for the numerical analysis, *i. e.* not the absolute values of meteorological elements are interpolated into the grid points but their deviations from a definite field, which is called the first guess. At the stage of initialization, the field obtained as a result of the analysis is projected onto the slow manifold that corresponds to the given model of the atmosphere, in order to decrease undesirable influence of inertial gravitational waves. The adoption system is a complicated technological line including a set of software.

In the recent years, these adoption systems based on the cycle forecast – analysis – initialization are mastered using the general variation formulation of the problem and the conjugated model in order to minimize the corresponding functional, on the basis of the known results on the theory of optimal control [2–5]. At the same time, some authors consider the formulation of the problem of data adoption in terms of estimation theory involving the algorithm of Kalman filter [3, 6, 7]. It should be noted that the problems of optimal control and optimal filtration are linked with each other through the principle of duality; in addition, there are algorithms of simultaneous estimation and control of the process [3, 7].

In application to the problem of the adoption of meteorological observation data, both approaches have some shortcomings, along with advantages. For example, the variation approach either does not take into account random errors of observation and forecast, or they are assigned independent of time.

The adoption algorithm based on Kalman filter is a natural generalization of the adoption systems which are forecast-analysis cycles [3]. Actually, this cycle is supplemented with a procedure for the calculation of covariations of errors for the fields to be estimated. At the same time, this algorithm requires substantial computer resources; in its full implementation, it cannot be realized at present even with a supercomputer because the order of covariation matrices for modern global models is tens thousands. In addition, under definite conditions, Kalman filter can be divergent with time. Nevertheless, the applicability of the theory of Kalman filter to the problem of data adoption is being investigated by many authors.

There are several approaches to solving the problem linked with large order of covariation matrices calculated in the Kalman filter algorithm that are intensively developing at present. One of these approaches is an attempt to describe analytically the behaviour of covariation errors of forecast. For example, in [8] the authors obtained an analytical equation in partial derivatives for the covariation function locally in the vicinity of a definite point under consideration. The ensemble Kalman filter is widely used. In it Monte-Carlo method is used to calculate covariations of forecast errors [9, 10]. The third approach consists in using simplified prognostic models to calculate the matrices of forecast error covariations. This aproach was proposed in [6]. The algorithm of Kalman filter in which simplified models are used to calculate covariation matrices is called suboptimal [3, 8]. Generalization of this approach in combination with the ensemble method is considered in [11].

As we have already mentioned above, the application of Kalman filtration theory to the problem of adoption of meteorological data is a very complicated task which is difficult for realization in modern prognostic problems because of the high order of the arising covariation matrices. However, this task has some advantages: the dynamics of the atmosphere is described by specific prognostic equations with well investigated properties. In addition, the known analytical equations have been obtained within turbulence theory for the structural functions of the fields of temperature and wind on the basis of Navie–Stokes equations and heat penetration. It should be noted that at present the optimal Kalman filtration procedure is applied also to the problem linked with the adoption of aerosol data. Similar difficulties arise in connection with solving the problem of forecasting covariation matrices of very high order. As a rule, various simplifications are involved, similar to the case of meteorological problems [12, 13].

The present article is a brief review of our works dealing with the development of the system of mathematical modeling of processes in the atmosphere and adoption of the data of meteorological observations, as well as the works on investigation of the applicability of Kalman filter theory to the problem of adoption of the data of meteorological and aerosol observations.

SYSTEM OF MATHEMATICAL MODELING OF PROCESSES IN THE ATMOSPHERE

As we have already mentioned above, from the viewpoint of the formulation of the problem linked with the analysis of the data of meteorological observations, two approaches are most popular in the world: variation approach based on optimal control procedure, and probabilistic approach involving the methods of estimation theory [14].

When developing the system of modeling, we took into account the possibility of its generalization with the help of modern data adoption methods, in particular those based on estimation theory.

The major components of this system are as follows:

– governing programme;
– meteorological data adoption system;
– model of transport of air pollutants;
– visualization unit.

In turn, the meteorological data adoption system consists of the following units:

– numerical (objective) analysis of observation data;
– initialization;
– a model of the atmosphere;
– interpretation of results.

It should be noted that the adoption system, along with the entire modeling system, is developed on the basis of the unit substitutability concept. All the units mentioned above are united into a complex with the help of governing programme. Then the description of the major units of the meteorological data adoption system and of the complex as a whole comes.

1. Analysis of observation data.

The basis for the development of a scheme for numerical analysis of observation data was optimal interpolation method proposed by L. S. Gandin. This method is an application of the statistical estimation theory to the problem of meteorological data analysis. To perform analysis, the forecast fields and the data on covariation functions of the errors of observation and forecast are involved as an additional *a priori* information. Generally, the optimal interpolation procedure can be represented as:

$$x_a = x_{fg} + K(y_0 - Mx_{fg})$$

$$K = PM^T(MPM^T + O)^{-1}$$

$$y_0 = Mx_{fg} + \varepsilon_0$$

where x_a is the n-vector of analyzed fields in the nodes of regular grid, x_{fg} is n-vector of the first approximation (forecast being taken as the latter), y_0 is m-vector of observations, K is a ($m\times m$) matrix of interpolation weights, P is a ($n\times n$) matrix of forecast error covariations, O is a ($m\times m$) matrix of observation error covariations, M is a ($n\times m$) matrix for determining the first approximation values at observation sites, ε_0 is a random m-vector of observation errors.

The analysis scheme is based on the box version of optimal interpolation method. Since the scheme is intended for real data, it contains all the components necessary for the work with real data: primary processing, box sort, data management, assignment of statistical characteristics, the very interpolation into the nodes of regular grid, and result estimation procedure. The scheme is developed in such a manner that one could use it to perform different versions of analysis: two-dimensional, three-dimensional, multivariant (in the case of multivariant analysis, the data on several meteorological elements is involved for the analysis of one element). This algorithm is described in detail in [15].

The analysis for the Siberian region is performed with fifteen standard isobaric surfaces (from 1000 to 10 mbar) of the fields of the height of isobaric surface and wind velocity. The field of pressure at the sea level is analyzed separately. The data of land and aerological observations is used as the initial information; the fields of both the 12-hour forecasts of the National Meteorological Centre

(NMC) of the USA and the 12-hour forecasts according to the regional model of short-term weather forecast described below are used as a first guess. The quality of analysis was estimated both by comparing with the analysis fields of the NMC of the USA and by estimating the forecast results using the mathematical model of the atmosphere.

2. Regional model of short-term weather forecast.

A baroclinic adiabatic model of the atmosphere in the isobaric system of coordinates is used in the modeling system (pressure is considered as the vertical coordinate). An important feature of the numerical algorithm of the realization of this model is the application of G. I. Marchuk's idea concerning splitting the system of prognostic equations over physical processes [16]. The description of previous versions of this model, solution procedure and estimation of the prognostic properties of the newest modification applied in the system are reported in [17].

3. Non-linear initialization for the regional model of the atmosphere.

The initialization procedure is used to decrease undesirable effect of inertial gravitational waves on the solution of the system of equations describing the model of the atmosphere. For this purpose, the fields obtained as a result of the analysis are projected on the slow manifold. A detailed description of the algorithm of non-linear initialization used in the system is presented in [18].

The system of adoption of meteorological data for the Siberian region, described in [19], is the basic technological line "forecast – analysis – initialization" for the development and investigation of new data adoption methods in quasi-operative regime and the related calculations of background fields for the problems of environmental protection. This system is a component of a complex system of modeling atmospheric processes. The numerical experiments performed with the adoption system indicate the applicability of this system for the adequate estimation of meteorological fields over the data of operative observations of meteorological elements.

The results of the operation of adoption system are the values of the meteorological fields of the height of isobaric surface, temperature and wind velocity in the nodes of regular grid. The height fields assigned at the input and obtained as the output of the adoption system are shown in Fig. 1: the fields of isolines of the height of isobaric surface of 500 bmbar, obtained with the help of the adoption system, for 0:00 AM of the mean Greenwich time (MGT) on March 30, 1991 (start of adoption), and 0:00 AM MGT on April 01, 1991 (end of adoption).

PROBLEMS OF NUMERICAL REALIZATION

The optimal interpolation method was chosen as an analysis method in the adoption system. The system was further developed and mastered using the optimal Kalman filtration theory. The algorithm of Kalman filter in the linear case is described in [7].

One of the major problems arising in the application of Kalman filter algorithm to the problem of the adoption of observation data is the high order of covariation matrices. A series of works is dealing with the investigation of this problem [20–23]. Simplified models for calculating covariations of forecast errors are described in these works.

The simplifications were based on two assumptions. The first one involves the following consideration: errors of the forecast of the coefficients of decomposition over eigen functions of the vertical operator L_p of the model (vertical normal modes) do not correlate with each other. The vertical operator of the prognostic model under consideration is a finite-difference analogue of the operator $-\frac{\partial}{\partial p}\frac{p^2}{m^2}\frac{\partial}{\partial p}$ where $m^2 = \frac{R^2 T^*(\gamma_a - \gamma_0)}{\partial f^2}$, T^* is the standart temperature at the sea level. This assumption is based on the known fact that the vertical normal modes are close to the empirical orthogonal fanctions [1, 24]. If during adoption the estimation according to Kalman filter is made for the coefficients of decomposition into vertical normal modes independently of each other, then, since we assume that the errors of vertical normal modes do not correlate with

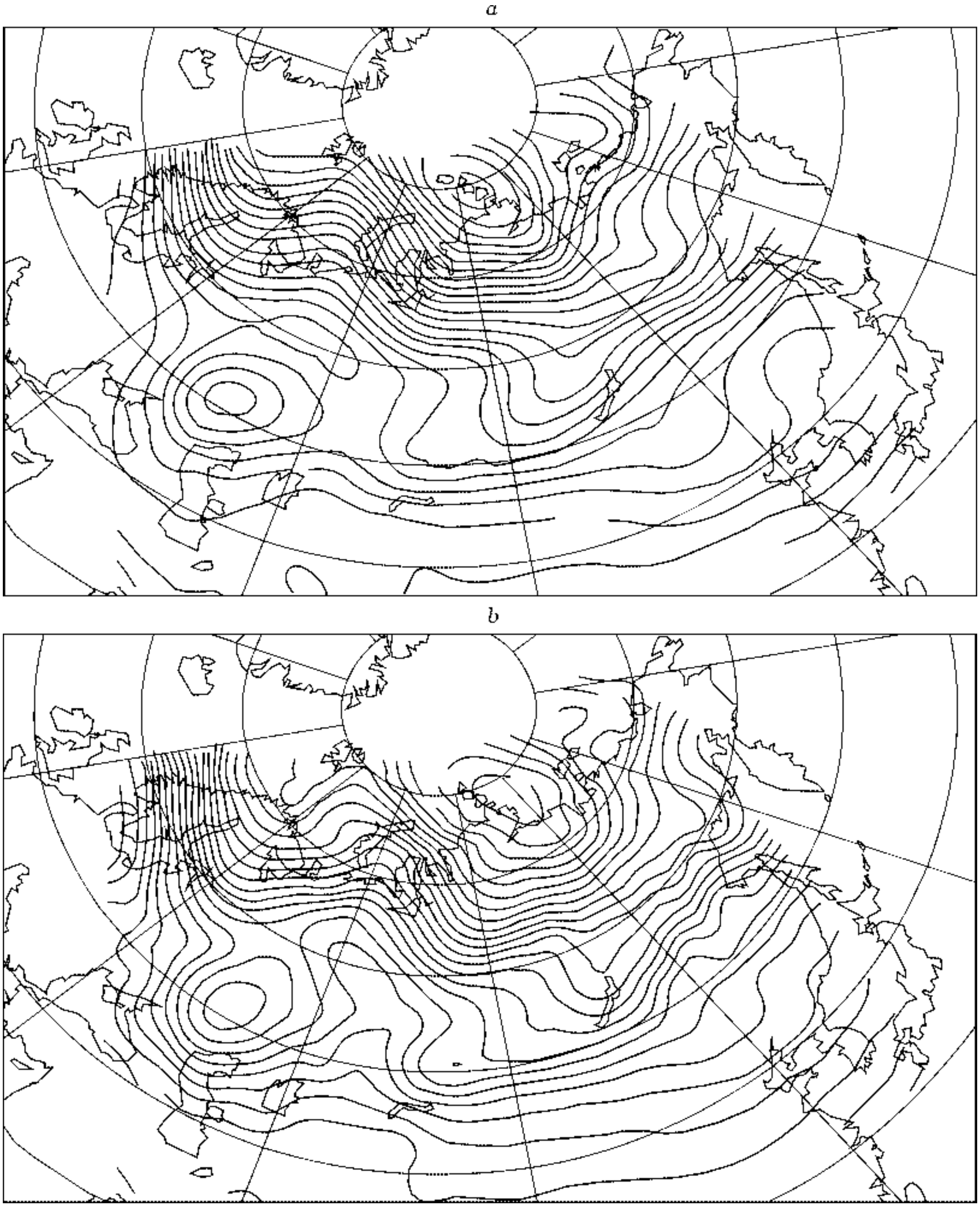

Fig. 1. Height of isobaric surface of 500 mbar for 0:00 AM MGT on March 30, 1991 (*a*) and for 0:00 AM MGT on April 1, 1991 (*b*).

each other, the covariation matrix P_k^f of forecast errors will be of diagonal-block structure. This provides a substantial decrease in the consumption of computer resources. The idea to perform the analysis of observation data for the coefficients of the decomposition over eigen functions of the L_p operator independently of each other was authored by G. I Marchuk; it is based on the fact that these coefficients have substantially different horizontal scales [1].

The second assumption is that the major contribution into the forecast error is made by its large-scale part; so, one can assume that the errors of the forecast of geopotential and wind fields are linked with each other by the geostrophicity equation; in other words, if we represent the solution of the initial system of prognostic equations as the transport steps over the trajectories of particles and adaptation of meteorological fields, only the transport step can be used to describe the behaviour of fore-

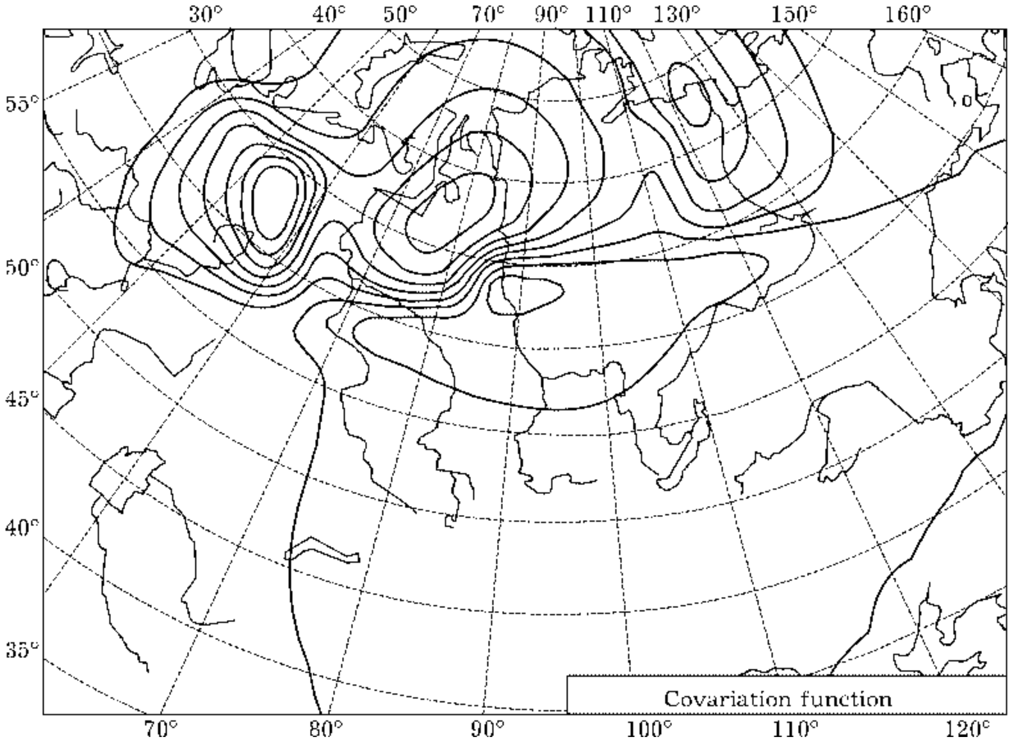

Fig. 2. The 6 h forecast of covariations of the errors of height fields and wind velocity components *U*.

cast errors, while the adaptation step may be omitted.

In order to meet the requirement that the errors of vertical normal modes should be noncorrelating at all the subsequent time steps, it is considered in all the simplified models that wind velocity fields in the operator of transport over trajectories are independent of the vertical coordinate p (this means that the background flux is close to barotropic one); at this condition, one can calculate covariations of forecast errors separately for each vertical mode.

In addition to simplified models that involve quasi-geostrophic assumption, a model for calculating the matrices of covariations of forecast errors in suboptimal Kalman filter is considered in [22]; it is based on full equations of hydrothermodynamics. An algorithm to be realized is proposed for calculating covariations of errors of height and wind velocity with the help of splitting over physical processes. The appearance of the covariation functions to be calculated is shown in Fig. 2 demonstrating the result of the calculation of the joint covariation of the errors of height and wind velocity fields for 6 h (for the central point of the region).

To estimate the quality of the simplified models, numerical experiments were made on the basis of Monte-Carlo method [21, 23]. With the help of these experiments it is demonstrated that the proposed simplified models give better results than the inertial forecast does, both for the forecast of matrices and for the use of the calculated matrices in the analysis. It is demonstrated that the proposed algorithms allow to calculate covariation matrices and can be used in the procedure of adopting the data of meteorological observations. In the results of numerical experiments reported in [21] one can see that small correlation of errors of vertical modes does not increase with time; the most important modes are the first 3–5 ones. Because of this, it seems promising to develop a 3-mode model to describe covariation matrices.

Since the algorithm involving the splitting over physical processes requires much more computer time, we are likely to prefer the calculation of covariations of forecast errors using the quasi-geostrophic assumption. It is demonstrated in [22] that this assumption is quite reasonable, in particular in the case when covariations of the forecast errors are homo-

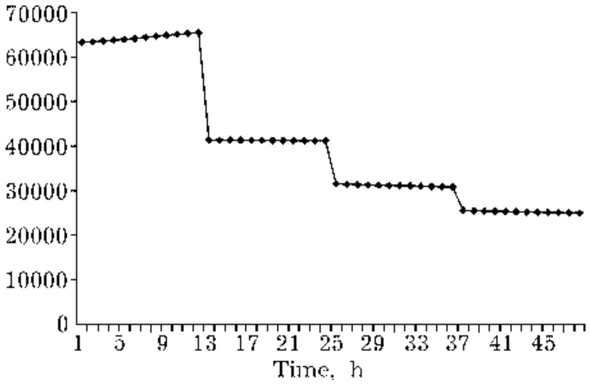

Fig. 3. Spur of forecast error covariation matrix (temporal behaviour).

geneous and isotropic. In this work, differential equations are deduced to describe local covariations of forecast errors in the uniform isotropic case. It is demonstrated that the prognostic model of the transport of substance over particle trajectories can be used in this case. Such a simplification does not provide a substantial decrease in the order of matrices to be calculated; thus, in order to make it possible to calculate covariations, we assume non-correlated errors of the vertical normal modes of the model. In doing this, covariations are calculated with the help of transport equations with wind averaged over the vertical direction. Numerical experiments performed under this requirement demonstrate that the covariation matrices obtained in calculations should not be strongly distorted by the assumptions made.

Experiments on the adoption of the modeled meteorological data were made on the basis of algorithms proposed in [21–23]. Calculations over 48 h were performed, involving adoption of the data on geopotential field to be modeled every 12 h. The observation data was assigned in the nodes of regular grid in 143 points distributed uniformly over the integration region. Results of these experiments are shown in Fig. 3 depicting the behaviour of theoretical error of the adoption algorithm based on Kalman filter, namely, the spur of the matrix of forecast error covariations. One can see that the error value decreases every 12 h, when the adoption of observation data occurs. In the future we plan to perform numerical experiments on the adoption of real data on the basis of algorithms of the suboptimal Kalman filter, and to apply the results obtained to the adoption of aerosol data. It should be stressed that the problem of the recovery of spatial and temporal distribution of the aerosol data requires knowing the distribution of meteorological data, such as wind, temperature, humidity; thus, these problems are linked with each other.

Acknowledgements

The work has been supported by the Russian Foundation for Basic Research (Grant of the leading scientific schools of Russia, 2000, No. 00–015–98543), and the Siberian Branch of the RAS (Integration Project No. 64 "Aerosol of Siberia").

REFERENCES

1 G. I. Marchuk, Chislennye metody v prognoze pogody, Gidrometeoizdat, Leningrad, 1967.
2 V. V. Peneko, Metody chislennogo modelirovaniya atmosfernykh protsessov, Gidrometeoizdat, Leningrad, 1981.
3 M. Ghil, P. Malanotte-Rizzolli, Advances in Geophysics, Vol. 33, Acad. Press, 1991, p. 141.
4 F.-X. Le Dimet, O. Talagrand, *Tellus*, 38A (1986) 97.
5 A. C. Lorenc, *Quarterly J. of the Royal Meteorol. Soc.*, 114 (1988) 205.
6 D. P. Dee, *Ibid.*, 117 (1991) 365.
7 A. H. Jazwinski, Stochastic Processes and Filtering Theory, Acad. Press, New York, 1970.
8 R. Todlimg, S. Cohn, *Monthly Weather Review*, 122 (1994) 2530.
9 H. L. Mitchell, P. L. Houtekamer, *Ibid.,* 128 (2000) 416.
10 P. L. Houtekamer, H. L. Mitchell, *Ibid.,* 129 (2001) 123.
11 P. F. J. Lermusiaux, A. R. Robinson, *Ibid.,* 127 (1999) 1385.
12 R. Menard, S. E. Conh, L. P. Chang, P. M. Lyster, *Ibid.,* 128 (2000) 2654.
13 R. Menard, S. E. Conh, L. P. Chang, P. M. Lyster, *Ibid,* 128 (2000) 2672.
14 G. S. Rivin, *Optika atmosfery i okeana*, 6 (1996) 780.
15 E. G. Klimova, G. S. Rivin, *Meteorologiya i gidrologiya*, 3 (1992) 16.
16 G. I. Marchuk, Metody vychislitel'noy matematiki, Nauka, Moscow, 1989.
17 G. S. Rivin, S. B. Medvedev, *Meteorologiya i gidrologiya*, 5 (1995) 13.
18 S. B. Medvedev, *Ibid.*, 6 (1998) 35.
19 E. G. Klimova, G. S. Rivin, *Ibid.,* 12 (1996) 19.
20 E. G. Klimova, *Ibid.*, 11 (1997) 55.
21 E. G. Klimova, *Ibid.*, 6 (2000) 18.
22 E. G. Klimova, *Ibid.*, 10 (2001) 24.
23 E. G. Klimova, *Ibid.*, 11 (2001) 11.
24 A. S. Monin, Prognoz pogody kak zadacha fiziki, Nauka, Moscow, 1969.