Some Current Problems in Numerical Weather Prediction

Article in Izvestiya Atmospheric and Oceanic Physics · May 2005

Mikhail Tolstykh
Russian Academy of Sciences

Izvestiya, Atmospheric and Oceanic Physics, Vol. 41, No. 3, 2005, pp. 285–295. From Izvestiya AN. Fizika Atmosfery i Okeana, Vol. 41, No. 3, 2005, pp. 315–327.

Some Current Problems in Numerical Weather Prediction[1]

M. A. Tolstykh* and A. V. Frolov**

** Institute of Numerical Mathematics, Russian Academy of Sciences, ul. Gubkina 8, Moscow, 119991 Russia*

e-mail: tolstykh@inm.ras.ru

*** Russian Federal Service of Hydrometeorology and Environmental Monitoring, Novovagan'kovskii per. 12, Moscow, 123995 Russia*

e-mail: afrolov@mecom.ru

Received June 8, 2004

Abstract—The problem of improving the accuracy of numerical weather prediction is considered. The review of the current state of two directions related to the solution of this problem is presented: increasing the accuracy of numerical solution of the atmospheric hydrothermodynamics equations and improving the evaluation of the initial state of the atmosphere. To increase the accuracy of solutions to the equations, it is necessary to apply efficient numerical methods and to switch to the nonhydrostatic equations. The improvement of the evaluation of the atmosphere's initial state is related to the development of the observation system and assimilation methods for observational data. Variational methods of data assimilation are considered in detail.

1. INTRODUCTION

Improvement in weather forecasting is an important problem of considerable practical value. The first attempts of numerical weather prediction were made in 1916 [1]; however, it was not until 1940 that I.A. Kibel developed the first prognostic model by using expansions of the equations of motion for a baroclinic atmosphere in terms of a small parameter. He introduced the quasi-geostrophic approximation with mathematical correctness and calculated the first numerical forecast of the pressure field for Eurasia with the help of an arithmometer [2]. The works of Kibel occupy a special place in the history of numerical weather prediction. He was the first to develop a quasi-geostrophic prognostic model [3], one of the first who called for a return to the full quasi-static Navier–Stokes equations in the mid-1950s [3, 4], and, finally, one of the first who realized the limitations of the quasi-static approximation while being interested in mesometeorological problems in 1970 [5]. The domestic school of numerical weather prediction was created by Kibel, and this school achieved considerable success. The history and detailed bibliography of hydrodynamic weather prediction in the Soviet Union is given in [6].

Any modern numerical weather-prediction system consists of a prognostic model responsible for reproducing the dynamics of the atmospheric global circulation and a system assimilating real observational data, which is used to evaluate the current state of the atmosphere. The research on improving the quality of numerical weather prediction is currently being carried out in the following principal directions.

[1] This article was submitted by the authors in English.

(a) Improving the accuracy of numerical solution of the atmospheric hydrothermodynamics equations. Here, it is possible to note two subtasks: the application of efficient numerical methods and transition to the nonhydrostatic equations. Numerical weather prediction with a high spatial resolution requires huge computing resources; in particular, the operational implementation imposes a restriction on the admissible elapsed time of the model run. The spatial resolution of the model can be increased by using both efficient numerical methods to the solution of the atmospheric hydrothermodynamics equations and modern computers with parallel architecture. The hydrostatic approximation commonly used in weather-forecasting models becomes inaccurate if the horizontal resolution is increased to approximately 10 km, especially near mountains. Nowadays, it is possible to run regional models with this resolution with the help of high-performance computers. Such a resolution will become possible for global models in the immediate future. Thus, transition to the full system of nonhydrostatic equations of atmospheric hydrothermodynamics becomes important, as was expected by Kibel in [5].

(b) Taking into account subgrid-scale physical processes. Many dangerous meteorologic events, whose prediction is especially important from the practical point of view, are closely connected to mesoscale meteorologic processes. Examples of such processes are squalls and showers. The simulation of mesoscale highly nonlinear processes requires both a high spatial resolution of the prognostic model, and a correct account of nonadiabatic subgrid-scale processes. Traditionally, models take into account radiation in the shortwave and longwave parts of the spectrum, turbulent exchange in the planetary boundary layer and in the free atmosphere, large-scale condensation, deep and shallow convection, and friction generated by vertically propagating gravity waves. A detailed description of convection processes is used in modern models, in particular, the momentum modification due to convection is included [7]. Studies on parametrizing microphysical processes (coagulation, autoconversion) during cloud formation and precipitation, three-dimensional parametrization of the planetary boundary layer, and the influence of the inhomogeneous underlying surface (vegetation, lake) and small atmospheric species (ozone, aerosols etc.) have been carried out. The development of new parametrizations often requires an appropriate development of dynamical cores of atmospheric models. For example, a correct account of ozone is impossible without introducing its advection equation into the model. Similarly, a detailed consideration of precipitation processes requires that the advection equations for the specific water content of liquid and crystalline phases be introduced into the model.

(c) Improving the evaluation of the initial state of the atmosphere. Ground-based observational data (high-altitude radiosounding of the atmosphere; meteorological observations on dry land, ships, drifting and stationary buoys, and planes) are distributed extremely nonuniformly over the globe; they have various measurement errors, and, in general, they are adhered to uniform synoptic periods of observations.

The group of geostationary and polar-orbital meteorological Earth satellites ensures continuous and global measurements of many quantities, such as the intensity, polarization, and angular and spectral characteristics of the radiation reflected and emitted by the atmosphere and the underlying surface, and other physical quantities. To derive the values of meteorological quantities from this information, it is necessary either to solve the inverse problem by involving *a priori* information, or to construct algorithms allowing a direct assimilation of radiation measurements with the help of an atmospheric prognostic model.

To enumerate the most important problems in this area, it is necessary to mention the task of combined assimilation for ground-based and satellite observational data, as well as the problem of optimal "weighing" for the contributions of observational data and model forecasts used as a first-guess field in evaluating the initial state of the atmosphere.

(d) Account of random properties of atmospheric processes. Rather recently, some meteorological centers have started to apply various ensemble weather-prediction systems. An ensemble prediction system (EPS) consists of the same weather-prediction model (as a rule, with a somewhat lower spatial resolution) and the system to generate an ensemble of initial data. This system allows a quantitative evaluation of the deterministic forecast reliability, in particular, to calculate the probability of a particular unfavorable weather event.

We note that all four directions are interrelated. For example, it is unnecessary to have a parametric description for deep-convection processes when the model horizontal resolution is about 2 km; at the same time, it is necessary to describe processes in the planetary boundary layer as three-dimensional processes in this case.

It is known that the increase of the model horizontal resolution is able to increase the forecast quality only insignificantly if it is not supported by an appropriate improvement of initial data. In turn, the implementation of modern data-assimilation schemes on the basis of a variational approach or of generalized Kalman filters requires multiple integration of the atmospheric model, which imposes rigid requirements on its computational efficiency.

In this paper, only the first and third problems are considered because of its limited volume. A review of modern ensemble prediction systems is given in [8], a detailed review of methods for parametrizing subgrid-scale processes and paths of their development in connection with the transition to the nonhydrostatic equations is presented in [9], and a review of parametrizations for hydrostatic models is published in [10].

2. INCREASE OF ACCURACY FOR NUMERICAL SOLUTION OF THE ATMOSPHERIC HYDROTHERMODYNAMICS EQUATIONS

2.1. Computational Efficiency of a Weather Prediction Model

A modern forecast model should adequately describe synoptic-scale atmospheric processes with periods from several hours to several days, especially cyclogenesis and frontogenesis processes, and the part of mesoscale processes with characteristic periods from tens of minutes to several hours. The complexity of the weather-prediction problem is determined by baroclinic and barotropic instabilities of the atmosphere. The characteristic time of error doubling in forecasting synoptic processes is 2–3 days. This yields an estimate for the predictability limit of about 1–2 weeks [11]. The forecast accuracy is actually the accuracy of predicting the trajectory of the model atmosphere in the phase space with a dimension of 10^7 or higher. This is the main difference between hydrodynamic weather prediction and climate modeling, where the principal factor is the description of the statistics of atmospheric processes. From times of Kibel, the Russian scientific school is traditionally strong in developing efficient numerical methods for weather-prediction problems [12]. The splitting method developed by G.I. Marchuk in the 1960s for meteorological problems then resulted in the development of the semi-implicit method by A. Rober [13], which allowed the time step to be increased by a factor of 5 in comparison with explicit time integration schemes. This method is still used in both most hydrostatic models and many nonhydrostatic atmospheric models.

The application of efficient numerical methods is necessary to implement a high-resolution forecast model under restrictions imposed by the operational technology and computing resources. The efficiency is understood here as the time necessary for one processor to integrate the system of model equations for one hour of model time at a given error of atmospheric-circulation reproduction. It is also desirable that the algorithms used would possess additional properties arising from the properties of original equations (for example, the positivity of moisture during advection). In addition, the methods used should have an intrinsic parallelism allowing an efficient implementation on modern parallel computers.

In the late 1980s, numerical weather-prediction models used second-order finite-difference methods or spectral Eulerian methods for solving the atmosphere-hydrothermodynamics equations. We will call the "spectral Eulerian method" the discretization of the full hydrothermodynamics equations in longitude and latitude using expansion series in spherical harmonics for all variables in combination with finite-difference discretization in time and vertical coordinate.

The advantages of finite-difference methods are their locality and the quadratic growth of operation count when the resolution is increased. However, a finite- difference atmospheric model formulated on a regular latitude–longitude grid has several drawbacks apart from the phase error. Owing to the convergence of meridians, this grid has a significant nonuniformity of resolution in longitude and latitude near the poles (for the resolution 10 km in latitude, the mesh size in longitude near the pole would be about 150 m). This drawback leads to a severe limitation on the CFL number in Eulerian models, to difficulties in using parallel iterative algorithms, and to unjustified expenses on calculations at "wasted" grid points (about 25% of total computation cost).

In comparison with traditional second-order finite-difference approximations, there is no phase error and nonlinear instability in the spectral method. Other advantages of the spectral method are the uniformity of its resolution on the sphere, the opportunity to analytically solve the elliptical equations on the sphere, and the opportunity to use a reduced grid (i.e., a grid with decreasing number of points in the longitude as the poles are approached) for calculating the tendencies of prognostic variables due to subgrid-scale processes. The principal disadvantage of the spectral method is the cubic growth of the number of operations with increasing resolution. In addition, the method is essentially nonlocal and there can be problems of a stable calculation of basis functions (associated Legendre functions) for spectral resolutions of more than 1000 spherical harmonics. The task to provide the load balance of all processors while implementing this method on parallel computers is highly difficult due to a commonly used triangular truncation of the series.

Therefore, since the late 1980s, there have been works devoted to the research and application of the numerical methods of discretization in the horizontal plane in atmospheric modeling. These methods have been successfully used in other areas of computational fluid dynamics, such as total variation diminishing (TVD) schemes, and Eulerian high-order finite-difference and semi-Lagrangian methods. The double Fourier series [16], finite-element method on icosahedral grids [17], spectral-element method [14], and pseudospectral method [15] were also considered.

2.1.1. Semi-Lagrangian Method

Recently, the majority of spectral and finite-difference numerical weather-prediction models have applied the grid semi-Lagrangian method to describe advection [18]. The predecessor of this method—the backward characteristics method—has been well known in computational fluid dynamics and meteorology since the early 1960s. However, the backward characteristics method and its modifications (for example, the grid-characteristic method [19]) were usually applied when the CFL numbers were less than unity and the order of approximation was no greater than two. The essence of the semi-Lagrangian method lies in the discretization of the advection equation along trajectories. The arrival point of trajectories are always points of the grid, while the departure points most often do not coincide with grid points. The values of a variable at departure points are obtained with interpolation by using the values at the surrounding points of the grid.

The modern semi-Lagrangian schemes mostly use a cubic interpolation to determine the values of a function at the departure point of the trajectory. Thus, the semi-Lagrangian method has an approximation error $O((\Delta x)^4/\Delta t)$ (at a constant wind speed). This method eliminates the restriction on the time step by the CFL condition, especially strict near the poles because of the convergence of meridians. An increase in the time step (by a factor of 3 to 5) makes it possible to speed up the forecast at a given resolution of the model or to increase the horizontal resolution at a given elapsed time of the forecast. Such an increase in the time step does not violate the approximation because the restriction on the CFL number in the atmosphere mainly manifests itself in calculations of rather smooth jet currents in the upper troposphere.

The semi-Lagrangian method gives a considerably smaller phase error in the solution as compared to Eulerian second-order finite-difference schemes and makes it possible to avoid the Gibbs phenomenon as opposed to the spectral method. At the same time, formally, there is no property of norm preservation of an advected quantity in the semi-Lagrangian method, which is necessary theoretically for a long-term (over some decades or more) integration of the model. The practice has shown that this does not deteriorate the quality of the solution if a model with a resolution of about one degree or higher is used for medium-range weather forecasting. Now the semi-Lagrangian approach is widely used in weather prediction models. Certainly this method is inapplicable to the problems requiring strict conservation, e.g., the problems having jump solutions (shock waves). We note that the semi-Lagrangian advection on the grid can be combined with a spectral algorithm for the solution of the Helmholtz equation arising in the semi-implicit time-integration scheme.

2.1.2. Brief Description of Modern Hydrostatic Operational Weather-Prediction Models

The leading prognostic centers currently use numerical hydrodynamic models with a horizontal resolution of 40–100 km for a global medium-range forecast. The overwhelming majority of the models (exceptions are the American, Russian, and Japanese models) use the semi-Lagrangian representation of the advection and a semi-implicit time integration scheme (mostly a two-time-level scheme). To represent the nonadvective terms of the hydrothermodynamics equations, one applies either a spectral method (European Center for Medium-Range Weather Forecasts (ECMWF), Météo-France, USA National Center for Environmental Prediction (NCEP), Russian Hydrometeorological Research Center, Japanese Meteorological Agency), or a finite-difference method (national meteorological centers of England and Germany), or a finite-element method (National Meteorological Center of Canada). The regional forecast models have a horizontal resolution of 5–25 km and are most often based on the finite-difference approach. The exception is the ALADIN model (France, Morocco, and Eastern Europe) which is based on a spectral representation. Many regional models use the semi-Lagrangian approach, and most of these are based on the semi-implicit scheme. The descriptions of these models can be found in [20–25].

The ECMWF model is without a doubt the leader among global models. This model is a spectral semi-Lagrangian model with a two-time-level semi-implicit time-integration scheme. The efficiency estimate for this model is given in [25]: if it would remain an Eulerian semi-implicit model with a three-time-level time scheme, the one-day forecast would be fulfilled ten times slower on the same computer system (the factor 5 is due to a large time step in the semi-Lagrangian method; the factor 2 is due to the application of the two-time-level time scheme on time instead of a three-time-level leapfrog scheme). Currently, this model has a resolution of T_L511 (the grid resolution is approximately 40 km) and 60 vertical levels, and it is planned to increase the horizontal resolution to 25 km (T_L799) and the vertical resolution to 91 levels in 2005. (T_L means the "linear" grid, when the number of grid points in longitude is related to the spectral truncation N as $2N + 1$. Its application is possible only in combination with the semi-Lagrangian advection [25].)

In Russia, the Eulerian spectral model with the three-time-level semi-implicit time scheme [22, 26] is currently applied for medium-range forecasting. It has a resolution of about 100 km in midlatitudes (T85) and 31 vertical levels.

In [27], the semi-Lagrangian model for numerical weather prediction is proposed. Its distinctive features are the application of compact fourth-order differences to approximate nonadvective terms and the use of the vertical component of the absolute vorticity and the divergence as prognostic variables. This model has been successfully tested at the Russian Hydrometeorological Research Center with the resolution 0.9 degrees in longitude, 0.72 degrees in latitude (70 km in midlatitudes), and 28 vertical levels.

There are two approaches for constructing the regional forecast models.

(1) A model with a constantly high resolution formulated in a limited area. In this case, the lateral boundary conditions are taken from another model, and they should be interpolated not only in space but also in time; thus an ill-posed problem arises. The methods of constructing the boundary conditions for these models are investigated in [28]. We note that it is necessary to have two models for this approach—a global model with a coarser resolution and a regional model—and also a block for interpolation of lateral boundary conditions.

(2) A global model with a locally high resolution. This approach is self-consistent because it does not require the formulation of boundary conditions on the lateral boundaries, and, thus, the problem of constructing the data assimilation system becomes simpler. The Canadian model of the medium-range weather forecast [20] and the French model [21] are based on this approach.

2.2. Nonhydrostatic Models

When the horizontal resolution of the forecast model is increased to 10 km or higher, the account of nonhydrostatic effects becomes important. The incompressible equations, which do not contain sound-wave solutions, were mostly considered in nonhydrostatic modeling in the 1970s, which allowed a certain increase in the time step. Now, the majority of models are based on the full compressible gas equations (see, for example, [29]). This is because the modern models mostly use an implicit time-integration scheme for those terms of the full equations that correspond to fast waves. In this case, the introduction of sound waves in the system of equations to be solved increases the computational cost only slightly.

Modern nonhydrostatic models are often either the extensions of the previous research models to a larger computational domain (for example, [30]) or nonhydrostatic generalizations of hydrostatic operational models. An exception is the American project of the development of a new nonhydrostatic model [31], which should join the best qualities of research and operational models.

Up to now, the majority of nonhydrostatic models use parametrizations of subgrid-scale processes developed basically for hydrostatic models, under the assumption that the horizontal solvable scale is much greater than the vertical one. The area of applicability of such parametrizations extends to a horizontal resolution of about 8 km. Fundamentally different parametrizations are necessary at a horizontal resolution of 1–2 km, as indicated in the Introduction. At the same time, there is no strict background what to do when some process is resolvable by the grid only partially (for example, processes of the deep convection at a horizontal resolution of 4–5 km).

2.2.1. Choice of a Vertical Coordinate

For the vertical coordinate, the majority of hydrostatic models use either a hybrid coordinate or the sigma coordinate based on the pressure normalized by its surface value. This type of the coordinate makes it possible to easily take into account a contour of the terrestrial surface, but the geopotential gradient is expressed as a small difference of two large terms. When the discrete orographic gradient reaches large values typical of nonhydrostatic models with a high horizontal resolution, the discretization error for each of these terms can give rise to a false circulation above mountains. In nonhydrostatic models, the generalization of this coordinate is applied when the hydrostatic part of pressure (i.e., the pressure that satisfies the hydrostatics equation) is used as a coordinate. Various forms of such a coordinate are given in [30, 32–34].

The simplicity of adaptation of existing packages for postprocessing model meteorological fields is an evident advantage of the pressure-based vertical coordinates.

Another possibility is to use the vertical coordinate related to the height from the sea level. One of these coordinates is the hybrid coordinate proposed in [35]:

$$\sigma = \begin{cases} \dfrac{z-h}{z_F-h} z_F, & h \le z \le z_F \\ z, & z_F \le z \le z_T. \end{cases}$$

Here, z_F is a certain level located above the maximum of the orography $h(x, y)$ and z_T is the height of the model atmosphere. The uniform resolution in the boundary layer is achieved with this choice of the vertical coordinate; in addition, the computational domain is rectangular, which simplifies programming. On the other hand, this coordinate system is nonorthogonal and strongly deformed near steep mountains, which leads to false orographic cyclones similarly to the case of the vertical hybrid coordinate based on pressure. In addition, the Helmholtz equation arising from the semi-implicit discretization in time becomes nonsymmetric when the coordinate of this type is used, and this requires more expensive algorithms to solve this equation.

Thus, the vertical coordinates based on both the pressure and the height from the sea level have their advantages and disadvantages. So far, the pressure-based vertical coordinates are used more frequently.

2.2.2. Numerical Methods for Solving the System of Nonhydrostatic Equations

As in hydrostatic models, the second-order finite differences on a staggered grid are used to discretize the equations in the vertical direction. The majority of nonhydrostatic models use the second-order finite differences on the staggered grid for discretization in the horizontal plane as well. Exceptions are the French model ALADIN and the Canadian regional model MC2 [36], where the semi-Lagrangian advection is applied. The high-order finite differences and the third-order Runge–Kutta time-integration scheme are used in the American model WRF [31].

The integration of the unfiltered nonhydrostatic equations of the atmosphere (with sound waves included) by an explicit scheme would mean a time step of a few seconds. Therefore, all nonhydrostatic models use some form of splitting in physical processes while integrating the terms responsible for fast processes by an implicit scheme. Either an explicit–implicit splitting algorithm (see, e.g., [37]), or a modified semi-implicit time-integration scheme [36, 38] are employed. The review of numerical methods used in nonhydrostatic regional models as of 2001 is published in [39].

2.2.3. Boundary Conditions

The same boundary condition at the upper boundary of the atmosphere, the rigid-lid condition, is applied currently to the hydrothermodynamics equations in the majority of nonhydrostatic models, as in hydrostatic models. To prevent wave reflection from the artificial upper boundary in the vertical, various damping mechanisms are used. It is evident that the transition to the upper-boundary radiation condition will be required as the horizontal and the vertical resolution increase. The correct formulation of lateral boundary conditions [28] is also necessary for regional models; however conditions similar to those used in hydrostatic models are currently applied.

3. PROBLEM OF INITIAL CONDITIONS

It is known from daily experience of operational meteorological centers that the accuracy of weather forecasts is determined by the errors of the objective analysis of the current state of the atmosphere and its underlying surface. The results of objective analysis are used as the initial conditions for the prognostic model. The error of the objective analysis in turn consists of the observation error, first-guess field error (also called background error), and error of the scheme of integrating observations and the first-guess fields into the objective-analysis fields (data-assimilation system). Thus, a significant potential for increasing the weather-forecast quality is concentrated in the reduction of initial-condition error. We will consider some most promising approaches to the solution of this problem, keeping in mind that the paths to increase the accuracy of prognostic models have been already discussed in the first part of this paper.

3.1. Development of the Observation System

The global observation system of the World Weather Watch (WWW) has passed some stages of development, starting from the moment when the first meteorological satellite of the Earth emerged in the mid-1960s. The basic achievements are covered in [40].

Due to the high cost of meteorological observations, the efficiency of individual components of WWW global observation systems is an important issue. To study this issue, it is possible to use the method of "twin" experiments, i.e., to carry out two integrations of the same prognostic model with and without considering data from a particular observation subsystem. The numerical experiments with the method of "twins" clearly demonstrate [41, 42] that the forecast quality becomes much worse without precise data from contact (in situ) measurements of vertical profiles of temperature, humidity and wind velocity even when a large volume of satellite information is available. Therefore, it is necessary to learn how to more efficiently use the satellite information together with data of ground-based observations. The modernization of satellite instruments and an increase in the volume of the satellite information will not give desirable results without solving this problem.

3.2. Methods of Optimal Interpolation and Discrete Four-Dimensional Data Assimilation

The first schemes for preparing initial fields were oriented mainly toward solving the problem of interpolating the measured values of meteorological parameters to the points of the computational grid of a prognostic model. These schemes have received a theoretical substantiation and a strong impetus for development owing to the statistical (optimal) interpolation method of observational data proposed by L.S. Gandin [43] in the early 1960s.

With the help of the optimal-interpolation method, it has become possible to implement the joint three-dimensional objective analysis of geopotential, temperature, and wind fields; to assess the quality of observational data; and to evaluate interpolational weights for all measurements of meteorological parameters with their accuracy and proximity to the analysis grid point taken into account.

The suggestion to use short-term numerical forecasts for 3 to 6 h as the first-guess field in the optimal interpolation scheme has appeared as another productive idea. Some algorithms of this type were created in the early 1980s. These algorithms are referred to as "discrete four-dimensional data-assimilation systems" [44, 45]. In these systems, a prognostic model is used as a spatiotemporal interpolant relaying the desired signal from the preceding analysis (i.e., analysis of observations carried out earlier) to the current analysis of the atmospheric state.

Owing to inconsistency of initial fields of the geopotential and wind obtained with the help of the optimal interpolation method, there are high-frequency inertia–gravitational oscillations in models. These oscillations make short-term forecasts "noisy" and lead to the rejection of real observational data. Various methods for nonlinear adjustment of geopotential and wind fields are applied to filter out fast-growing high-frequency perturbations by taking into account nonadiabatic factors: the method of nonlinear normal-mode initialization [46, 47] and the method of digital filtering [48, 49].

In the mid-1990s, the overwhelming majority of operational meteorological centers used the systems for discrete four-dimensional data assimilation based on the joint three-dimensional optimal-interpolation method. One can find a detailed description of this system created at the USSR Hydrometeorological Center (since 1992 the Russian Hydrometeorological Research Center) in [50].

3.3. Variational Methods for Data Assimilation

The optimal interpolation method encountered serious difficulties in assimilating information from meteorological satellites and radars. It is difficult to present this information as meteorological parameters without noticeable loss of accuracy. These difficulties stimulated the development of variational data assimilation methods. The fundamentals of these methods were incorporated in the perturbation theory developed for discrete models of the dynamics of the atmosphere and ocean by Marchuk [51, 52]. While developing this theory, the formulation of variational data assimilation problem was proposed in [53] as the problem of minimization for functionals defining the measure of deviations between meteorological fields calculated with the help of models and measured in real conditions.

Some examples of solving simplified problems of the objective analysis of meteorological fields by using the variational principle were constructed in the Soviet Union in the late 1970s–early 1980s [54, 55]. However, due to serious methodological problems and lack of high-efficiency computer systems, it was not possible to create practical variational schemes for four-dimensional data assimilation of meteorological observations in the Soviet Union.

3.3.1. Three-Dimensional Variational Data Assimilation

Abroad, works [56, 57] on the barotropic vorticity equation and work [58] on the shallow-water equations have served as a starting point for research on the variational data assimilation for meteorological observations. The interest in variational assimilation has increased considerably in the late 1980s because of an insufficiently high accuracy of restoring the vertical profiles of temperature and humidity from HIRS data acquired by the radiometers that were installed onboard the polar-orbital satellites (TOVS data). To solve these problems, an algorithm for one-dimensional variational analysis (1D-VAR) was proposed in [59], and its success was confirmed with the operational model [60].

Let us briefly consider the basic particularities of the algorithm for statistical three-dimensional variational assimilation of data of meteorological observations (3D-VAR), in which all the observations acquired in some time interval are treated simultaneously and the time evolution is completely determined by the numerical solution of a prognostic atmospheric model.

Let $\mathbf{x}$ be the column vector of true values for the atmospheric parameters determined at the instant t at points of a regular grid $\mathbf{r}_i$, where $1 \le i \le I$. Without loss of generality, it is possible to use the coefficients of expansion in terms of the set of orthogonal functions instead of grid-point values of the atmospheric variables because there is a one-to-one correspondence between them. Following [61], we designate the column vector of prognostic (first-guess field) values of atmospheric parameters as $\mathbf{x}_b$ and the column vector of analyzed values of atmospheric parameters given on the same grid as $\mathbf{x}_a$.

The vector of the first-guess field error $\mathbf{e}_b$ is $\mathbf{e}_b = \mathbf{x}_b - \mathbf{x}$. As in the optimal interpolation method, it is assumed that the first-guess field errors are not biased but can correlate with one another; i.e., $\langle \mathbf{e}_b \rangle = 0$ and $\langle \mathbf{e}_b(\mathbf{e}_b)^T \rangle = \mathbf{P}_b$, where $\mathbf{P}_b$ is the square positive definite covariance matrix of the first-guess field error. The angular brackets show the average over an ensemble (expectation) and the superscript T specifies the transposition of a matrix.

Let observational data be given by a vector $\mathbf{y}$. The dimension L for the column vector of observations $\mathbf{y}$ generally differs from the number of points of the prognostic grid I and, in addition, the list of variables of the vector $\mathbf{x}$ can differ from the list of variables of the vector $\mathbf{y}$. In particular, $\mathbf{x}$ can be the temperature and the wind velocity, while $\mathbf{y}$ is the outgoing radiation measured in various parts of the spectrum or data on the radio-wave refraction angle obtained with the help of GPS satellites. To establish the correspondence between observed, prognostic, and analyzed variables, one defines the direct operator H translating the prognostic and analyzed variables $\mathbf{x}_b$ and $\mathbf{x}_a$ into the observed variables $\mathbf{y}$. For example, if $\mathbf{x}$ is the temperature of the atmosphere and $\mathbf{y}$ is the outgoing radiation, the operator H will represent the radiative-transfer equation.

Thus, it can be written that $\mathbf{y} = H(\mathbf{x}) + \mathbf{e}_r$, where $\mathbf{e}_r$ is the vector of the observation error. The observation error has two components—the instrumental error and the error of a method restoring meteorological parameters (representation error). In the limiting case when $\mathbf{x}$ and $\mathbf{y}$ define the same set of variables and when observational data do not coincide with the points of the grid on which the analysis is carried out, the operator H represents a simple interpolation operator and the error of the method of restoring meteorological parameters is merely the representation error of observational data. It is assumed that the observation error, as well as the first-guess field error, has an unbiased estimator and the positive definite covariance error matrix $\mathbf{R} = \langle \mathbf{e}_r(\mathbf{e}_r)^T \rangle$.

For normally distributed first-guess field errors and observation errors, the most probable vector of the current state of the atmosphere $\mathbf{x}_a$ is obtained from the minimum condition on the scalar cost functional $J(\mathbf{x}_a)$ in the variable $\mathbf{x}_a$ provided that the first-guess field errors and the observation errors do not correlate with each other:

$$\begin{aligned} J(\mathbf{x}_a) &= 1/2[\mathbf{y} - H(\mathbf{x}_a)]^T \mathbf{R}^{-1} [\mathbf{y} - H(\mathbf{x}_a)] \\ &\quad + 1/2[\mathbf{x}_b - \mathbf{x}_a]^T \mathbf{P}_b [\mathbf{x}_b - \mathbf{x}_a]. \end{aligned} \tag{1}$$

The minimum of the functional $J(\mathbf{x}_a)$ is reached at the point $\mathbf{x}_a$, where the gradient of the functional $\nabla J(\mathbf{x}_a)$ is equal to zero. The second derivative of the functional $J(\mathbf{x}_a)$ with respect to $\mathbf{x}_a$ results in the square Hessian matrix that gives the curvature measure of this functional. It is known from the theory of functional analysis that the extremum of a scalar function will be a minimum if the Hessian matrix is positive definite. This condition is fulfilled if the observation-error covariance matrix $\mathbf{R}$ and first-guess error covariance matrix $\mathbf{P}_b$ are positive definite. We note that the implementation of this condition can present some difficulties in practice if three-dimensional covariance functions are constructed from empirical data.

In practice, finding the minimum of functional (1) is not a simple task because of a large dimension of the problem. For this purpose, its vector, the gradient $\nabla J(\mathbf{x}_a)$, and the second derivative $\nabla^2 J(\mathbf{x}_a)$ are estimated using prognostic values $\mathbf{x}_b$ in the appropriate Jacobi and Hessian matrices under the assumption that $\mathbf{x}_b$ is fairly close to $\mathbf{x}_a$. Various approaches to the solution of this problem are described in [61].

3.3.2. Connection between Optimal Interpolation and Three-Dimensional Variational Analysis Methods

An explicit expression for the optimal vector $\mathbf{x}_a$ that gives the minimum to functional (1) can be written as [45]

$$\mathbf{x}_a - \mathbf{x}_b = \mathbf{P}_b H^T [H\mathbf{P}_b H^T + \mathbf{R}]^{-1} [\mathbf{y} - H(\mathbf{x}_b)]. \quad (2)$$

Equation (2) makes it possible to find the relation between the 3D-VAR method and the optimal interpolation algorithm. The optimal interpolation explicitly does not include the operator H accounting for direct transformation of variables. However, if the observation variables $\mathbf{y}$ and the grid variables $\mathbf{x}$ coincide, the operator H represents a simple spatial interpolation. If the operator H is linear, the term $H\mathbf{P}_b H^T$ can be approximated by the first-guess-error covariance matrix calculated between the observation points and the term $\mathbf{P}_b H^T$ can be approximated by the first-guess-error covariance matrix calculated between observation points and grid points. Thus, the matrix $H(\mathbf{x}_b)$ characterizes the operator of spatial interpolation of the first-guess field into observation points.

The following advantages of the 3D-VAR algorithm in comparison with the multivariate optimal-interpolation algorithm are visible from Eq. (2). First, all observations influence the analysis at any grid point, while special hypotheses are introduced in the optimal interpolation to select the number of influencing observations. Second, one can use the much more general operator H for a direct transformation of variables. Owing to this, the observational data that are not coupled directly to the analysis variables can be easily included in the assimilation scheme and their coupling relations can be weakly nonlinear. Third, a more advanced and more realistic formulation for the statistical model of errors is allowed. This model determines the weights prescribed by observational data.

In practice, one of the methods of variational data assimilation was applied for the first time to the ECMWF prognostic system [62–64]. Among important innovations, it is necessary to note the application of nonseparable functions to describe the statistical structure of the first-guess field's errors. The nonseparability allows one to introduce a number of improvements in comparison with the optimal interpolation, such as the increase of the correlation scale with height for the mass field, a smaller correlation scale for temperature than for mass, and a more "acute" form of the correlation function in the vertical for wind than for mass. Another important particularity is the application of Hough functions to the description of the cross covariances of the pressure and wind-velocity fields. To separate Rossby and gravity waves, the geostrophic condition is used.

Note that, interestingly, despite a number of theoretical advantages in the mathematical formulation, the authors of [62–64] did not manage to show an advantage in the accuracy of restoring three-dimensional meteorological fields in the Northern Hemisphere's troposphere in comparison with the optimal-interpolation method. There is also no progress in the tropical area, where, as in the optimal interpolation, the objective analysis of the wind field is carried out under the assumption that this field is nondivergent and interacts weakly with the pressure field. Probably, this can be explained by the fact that the main practical advantage of the 3D-VAR method lies in the opportunity to directly assimilate the data of vertical–humidity satellite sounding. As was already noted, the influence of these soundings is exhibited mainly in the Southern Hemisphere.

3.3.3. Four-dimensional Variational Data Assimilation (4D-VAR) and Methods of Kalman Filtering

The four-dimensional variational data assimilation (4D-VAR) was initially regarded as the main goal of the variational approach, the three-dimensional assimilation (3D-VAR) being the necessary intermediate step. The basic idea of 4D-VAR lies in the minimization of the span between observations, the statistical model for the first-guess (background) error, and the prognostic model of the atmosphere in some time interval called the assimilation window. All the approaches proposed up to now use the hypothesis about an ideal atmospheric model, which has no error. The model trajectory approximating statistical information and observational data in the best way within the bounds of the assimilation window is sought for in this case. The 4D-VAR algorithm uses the adjoint equations to evaluate the gradient of the cost functional giving a measure of proximity between the prognostic model and available information. The incremental formulation of the 4D-VAR algorithm proposed in [65] allowed its implementation on existing supercomputers.

One of the first practical applications of the 4D-VAR scheme was implemented in the ECMWF prognostic system [62]. The formulation of the cost functional in increments [61] can be written as

$$J(\delta\mathbf{x}) = 1/2\Sigma[H_i\delta\mathbf{x}(t_i) - \mathbf{d}_i)]^T\mathbf{R}_i^{-1}[H_i\delta\mathbf{x}(t_i) - \mathbf{d}_i)] + 1/2\delta\mathbf{x}^T\mathbf{P}_b^{-1}\delta\mathbf{x}, \quad (3)$$

where i is the time index; $\delta\mathbf{x}$ is the increment that should be added to the first-guess field $\mathbf{x}_b$ at the initial moment $i = 0$; $\delta\mathbf{x}(t_i) = \mathbf{M}(t_i, t_0)\ \delta\mathbf{x}$ is the increment, whose time evolution is characterized by the tangent-linear prognostic model of the atmosphere; $\mathbf{R}$ and $\mathbf{P}_b$ are the covariance matrices for the observation error and first-guess (background) error, respectively; and H_i is a suitable linear approximation describing the modification in time t_i for the direct observation operator H.

In Eq. (3), the innovation vector $\mathbf{d}_i = \mathbf{y}_i^0 - H_i\mathbf{x}_b(t_i)$ is calculated at each time step. The estimate of the first-guess vector is calculated from the full nonlinear prognostic model $\mathbf{M}$ from the moment t_0 to the moment t_i by using the initial field $\mathbf{x}_b$. The increment $\delta\mathbf{x}_a$ minimizing the cost functional on the time interval (t_0, t_i) is sought. For this purpose, the Newton method of coordinate descent is applied. The resulting $\delta\mathbf{x}_a$ is then added to the first-guess field $\mathbf{x}_b$ to obtain the analysis vector $\mathbf{x}_a$:

$$\mathbf{x}_a = \mathbf{x}_b + \delta\mathbf{x}_a. \quad (4)$$

The ECMWF spectral model T106L19 with the full package of parametrizations for subgrid-scale physical processes is applied as a prognostic model in [63]. The tangent–linear model T63L19 describing increment propagation in time is the same prognostic model but with a lower spatial resolution and a limited set of physical parametrizations. Clearly the tangent–linear model cannot sufficiently accurately approximate the perturbations with relatively large amplitudes generated by the physical subgrid-scale processes. The simplest way to account for some nonlinear effects in the final analysis lies in the sequential solution of minimization problems for the functional $J(\delta\mathbf{x})$ [63].

The algorithm described above consists of two nested loops. The outer loop uses the full nonlinear model of high spatial resolution to estimate the sequence of innovation vectors. The inner loop uses tangent–linear and adjoint models constructed with respect to a simpler prognostic model with a smaller resolution and simplified description of subgrid-scale physical processes to find the minimum of the cost functional. Numerical experiments on comparison of the 4D-VAR algorithms with several data-assimilation windows (6, 12, and 24 h) and 3D-VAR have shown [63] noticeable advantages of the 4D-VAR method with the assimilation windows of 6 and 12 h. For the assimilation window of 24 h, the results are somewhat worse than for the 3D-VAR method in spite of the fact that, in theory, the extension of the assimilation window should result in the improvement of the optimal-estimation results. The authors of [63] explain these results by the assumption of ideality of the full nonlinear model (the prognostic model is regarded as a strong constraint), and by the simplifications introduced in constructing the tangent–linear and adjoint models.

To reduce the role of these assumptions, it is possible to include a term in the cost functional characterizing the error of the full nonlinear model or to reduce the error of the tangent–linear and adjoint models. Intensive research is currently being carried out in these directions.

The problem of variational four-dimensional data assimilation requires huge computer resources. Therefore, interest in Kalman filters, regarded as an alternative approach to variational data assimilation methods for geophysical observations, has reappeared. A Kalman filter represents a sequential algorithm in which the solution is sought by taking into account observational data at the time of analysis, the forecast in the interval between the previous and current analysis time, and the first-guess field's covariances. The equations of the prognostic model are used for the time extrapolation of the analysis-error covariances, which include error covariances of the prognostic model itself. The principal difference of a Kalman filter from the four-dimensional variational data assimilation is that this method takes into account the prognostic-model error explicitly and does not impose restrictions on the time interval (assimilation window) in which the assimilation of the earlier observational data occurs. A review of the theory of Kalman filters is given in [66], and a detailed review of modern research in the field of variational assimilation and Kalman filters is published in [67].

CONCLUSIONS

The level reached in leading world meteorological weather-forecast centers allows an efficient use of prognostic information to solve many practical problems in the forecast range of one to eight days [68], and the quality of the five-day forecast is the same as that of the two-day forecast 25 years ago (or the same as the quality of four-day forecast five years ago). At the same time, a wide circle of problems remain unsolved, and some of these were considered in this work. Solving any of these problems may help to achieve a better quality of numerical weather prediction and extension of its practical range.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research, project nos. 04-05-64638 and 04-07-90183.

REFERENCES

1. L. F. Richardson, *Weather Prediction by Numerical Process* (Cambridge University Press, Cambridge, 1922).
2. I. A. Kibel, "Application of the Equations of Baroclinic Fluid Mechanics to Meteorology," Izv. Akad. Nauk SSSR, Ser. Geogr. Geofiz., No. 5, 627–638 (1940).
3. I. A. Kibel, "Mode of Short-Term Forecasting of Meteorological Elements," in *Selected Works of I.A. Kibel on Dynamic Meteorology* (Gidrometeoizdat, Leningrad, 1984) [in Russian].
4. I. A. Kibel, "Finite-Difference Scheme for Solving the Complete System of Equations of Short-Term Forecasting and Quasi-Geostrophic Relations," in *Selected Works of I.A. Kibel on Dynamic Meteorology* (Gidrometeoizdat, Leningrad, 1984) [in Russian].
5. I. A. Kibel, "Hydrodynamic Short-Term Forecast in Mesometeorological Problems," Tr. GMTs, No. 48, 3–33 (1970).

6. M. A. Petrosyants, "History of Weather Forecast Development in Russia," in *165 Years of the Russian Hydrometeorological Service* (Gidrometeoizdat, St. Petersburg, 2001), pp. 45–67 [in Russian].
7. D. Gregory, R. Kershaw, and P. M. Inness, "Parameterization of Momentum Transport by Convection II: Tests in Single-Column and General Circulation Models," Q. J. R. Meteorol. Soc. **123**, 1153–1183 (1997).
8. *Proceedings of ECMWF Seminar on Predictability of Weather and Climate* (ECMWF, Reading, UK, 2002). Available at http://www.ecmwf.int/publications/library/do/references/list/209.
9. *Proceedings of ECMWF Seminar on Key Issues of the Parameterization of Subgrid Physical Processes* (ECMWF, Reading, UK, 2001).
10. *Proceedings of ECMWF Seminar on Parameterization of Subgrid-Scale Physical Processes, Reading, UK, 1994* (ECMWF, Reading, UK, 1995).
11. A. S. Monin, *Weather Prediction As a Problem of Physics* (Nauka, Moscow, 1969) [in Russian].
12. G. I. Marchuk, *Numerical Methods in Weather Prediction* (Gidrometeoizdat, Leningrad, 1967) [in Russian].
13. A. Rober, "Semi-Implicit Method" in *Numerical Methods Used in Atmospheric Models*, Ed. by F. Mesinger and A. Arakawa (GARP Publication Series, no. 17, 1979; Gidrometeoizdat, Leningrad, 1982, vol. 2, pp. 302–315).
14. M. Taylor, J. Tribbia, and M. Iskandarani, "The Spectral Element Method for the Shallow Water Equations on the Sphere," J. Comput. Phys. **130**, 92–108 (1997).
15. W. F. Spotz, M. A. Taylor, and P. N. Swarztrauber, "Fast Shallow-Water Equations Solvers in Latitude–Longitude Coordinates," J. Comput. Phys. **145**, 432–444 (1998).
16. H.-B. Cheong, "Application of Double Fourier Series to the Shallow Water Equations on a Sphere," J. Comput. Phys. **165**, 261–287 (2000).
17. F. X. Giraldo, "Lagrange–Galerkin Methods on Spherical Geodesic Grids: The Shallow Water Equations," J. Comput. Phys. **160**, 336–368 (2000).
18. A. Staniforth and J. Côté, "Semi-Lagrangian Integration Schemes for Atmospheric Models: A Review," Mon. Weather Rev. **119**, 2206–2223 (1991).
19. K. M. Magomedov and A. S. Kholodov, *Grid-Characteristic Numerical Methods* (Nauka, Moscow, 1988) [in Russian].
20. J. Côté, S. Gravel, A. Methot, *et al.*, "The Operational CMC-MRB Global Environmental Multiscale (GEM) Model. Part I: Design Considerations and Formulation," Mon. Weather Rev. **126**, 1373–1395 (1998).
21. P. Courtier, C. Freydier, J.-F. Geleyn, *et al.*, "The ARPEGE Project at Météo-France," in *Proceedings of ECMWF Seminar on Numerical Methods in Atmospheric Models, Reading, UK, 1991* (Reading, UK, ECMWF, 1992), vol. 2, pp. 192–208.
22. G. P. Kurbatkin, A. I. Degtyarev, and A. V. Frolov, *Spectral Model of the Atmosphere, Initialization, and Database for Numerical Weather Forecasting* (Gidrometeoizdat, St. Petersburg, 1994) [in Russian].
23. T. Davies, M. J. P. Cullen, M. H. Mawson, and A. J. Malcolm, "An Overview of Numerical Methods for the Next Generation U.K. NWP and Climate Model," in *Proceedings of ECMWF Seminar on Recent Developments in Numerical Methods for Atmospheric Modelling*, Reading, UK, 1998 (Reading, UK, ECMWF, 1999), pp. 202–225.
24. D. Majewski, D. Liermann, P. Prohil, *et al.,* "The Operational Global Icosahedral–Hexagonal Gridpoint Model GME: Description and High-Resolution Tests," Mon. Weather Rev. **130**, 319–338 (2002).
25. M. Hortal, "Aspects of the Numerics of the ECMWF Model," in *Proceedings of ECMWF Seminar on Recent Developments in Numerical Methods for Atmospheric Modelling*, Reading, UK, 1998 (Reading, UK, ECMWF, 1999), pp. 127–143.
26. A. V. Frolov, A. I. Vazhnik, V. I. Tsvetkov, and E. D. Astakhova, "Global Spectral Model of the Amosphere with a High Vertical Resolution," Meteorol. Gidrol., No. 2, 12–20 (2000).
27. M. A. Tolstykh, "Semi-Lagrangian High-Resolution Atmospheric Model for Numerical Weather Prediction," Meteorol. Gidrol., No. 4, 5–16 (2001).
28. A. McDonald, "A Step toward Transparent Boundary Conditions for Meteorological Models," Mon. Weather Rev. **130**, 140–151 (2002).
29. R. Rõõm and A. Männik, "Responses of Different Nonhydrostatic, Pressure-Coordinate Models to Orographic Forcing," J. Atmos. Sci. **56**, 2553–2570 (1999).
30. J. Dudhia, "A Nonhydrostatic Version of the Penn State-NCAR Mesoscale Model: Validation Tests and Simulation of an Atlantic Cyclone and Cold Front," Monthly Weather Rev. **121**, 1493–1513 (1993).
31. J. Michalakes, S. Chen, J. Dudhia, *et al.*, "Development of a Next-Generation Regional Weather Research and Forecast Model," in *Developments in Teracomputing: Proceedings of 9th ECMWF Workshop on the Use of High Performance Computing in Meteorology*, Ed. by W. Zwiehofer and N. Kreitz (World Sci., Singapore, 2001), pp. 269–276.
32. R. Laprise, "The Euler Equations of Motion with Hydrostatic Pressure As an Independent Variable," Mon. Weather Rev. **120**, 197–207 (1992).
33. R. Bubnova, G. Hello, P. Benard, and J.-F. Geleyn, "Integration of the Fully Elastic Equations Cast in the Hydrostatic Pressure Terrain-Following Coordinate in the Framework of the ARPEGE/ALADIN NWP System," Mon. Weather Rev. **123**, 515–535 (1995).
34. Z. I. Janjic, J. P. Gerrity, and S. Nickovic, "An Alternative Approach to Nonhydrostatic Modeling," Mon. Weather Rev. **129**, 1164–1178 (2001).
35. T. Gal-Chen and R. Sommerville, "On the Use of a Coordinate Transformation for the Solution of the Navier-Stokes Equations," J. Comput. Phys. **17**, 209–228 (1975).
36. R. Benoit, M. Desgagne, P. Pellerin, *et al.*, "The Canadian MC2: A Semi-Lagrangian, Semi-Implicit Wideband Atmospheric Model Suited for Fine Scale Process Studies and Simulation," Mon. Weather Rev. **125**, 2382–2415 (1997).
37. J. Steppeler, G. Doms, U. Schattler, *et al.*, "Meso-Gamma Scale Forecasts Using the Nonhydrostatic Model LM," Meteorol. Atmos. Phys. **82**, 75–96 (2003).

38. P. Bénard, "On the Use of a Wider Class of Linear Systems for the Design of Constant-Coefficients Semi-Implicit Time Schemes in NWP," Mon. Weather Rev. **132**, 1319–1324 (2004).

39. J. Steppeler, R. Hess, U. Schattler, and L. Bonaventura, "Review of Numerical Methods for Nonhydrostatic Weather Prediction Models," Meteorol. Atmos. Phys. **82**, 287–301 (2003).

40. *Twenty-First Status Report on Implementation of the World Weather Watch* (WMO, 2003), no. 957.

41. *Proceedings of the CGC/WMO Workshop on the Impact of Various Observing Systems on Numerical Weather Prediction* (Geneva, WMO TD, 1997), no. 868.

42. *Proceedings of the Second CGC/WMO Workshop on the Impact of Various Observing Systems on Numerical Weather Prediction* (Toulouse, WMO TD, 2000), no. 1034.

43. L. S. Gandin, *Objective Analysis of Meteorological Fields* (Gidrometeoizdat, Leningrad, 1963) [in Russian].

44. A. Lorenc, "A Global Three-Dimensional Multivariate Statistical Interpolation Scheme," Mon. Weather Rev. **109**, 701–721 (1981).

45. R. Daley, *Atmospheric Data Assimilation* (Cambridge University Press, Cambridge, 1991).

46. B. Machenhauer, "On the Dynamics of Gravity Oscillation in Shallow Water Model, with Application to Normal Mode Initialization," Beitr. Phys. Atmos. **50**, 253–271 (1977).

47. A. V. Pokudov, "Relaxation Method of Nonlinear Initialization on Normal Modes," Meteorol. Gidrol., No. 12, 109–111 (1989).

48. P. Lynch and X.-Y. Huang, "Initialization of the HIRLAM Model Using a Digital Filter," Mon. Weather Rev. **120**, 1019–1034 (1992).

49. P. Lynch, D. Giard, and V. Ivanovici, "Improving the Efficiency of a Digital Filtering Scheme for Diabatic Initialization," Mon. Weather Rev. **125**, 1976–1982 (1997).

50. A. V. Frolov, A. I. Vazhnik, P. I. Svirenko, and V. I. Tsvetkov, *Global Data Assimilation System for Observations of the Atmosphere* (Gidrometeoizdat, St. Petersburg, 2000) [in Russian].

51. G. I. Marchuk, "Basic and Adjoint Equations of Dynamics of the Atmosphere and Ocean," Meteorol. Gidrol., No. 2, 17–34 (1974).

52. G. I. Marchuk, *Methods of Numerical Mathematics*, 2nd ed. (Springer, New York, 1975; Nauka, Moscow, 1977).

53. V. V. Penenko, *Methods of Numerical Modeling of Atmospheric Processes* (Gidrometeoizdat, Leningrad, 1981) [in Russian].

54. V. V. Penenko and N. N. Obraztsov, "Variational–Difference Method for the Objective Analysis of the Fields of Meteorological Elements," Meteorol. Gidrol., No. 6, 15–25 (1978).

55. V. P. Svidritskii, "Multielement Spectral Analysis with Application of Normal Modes to an Atmospheric Model," Meteorol. Gidrol., No. 8, 78–83 (1984).

56. O. Talagrand and P. Courtier, "Variational Assimilation of Meteorological Observation with the Adjoint Vorticity Equation. I: Theory," Q. J. R. Meteorol. Soc. **113**, 1311–1328 (1987).

57. P. Courtier and O. Talagrand, "Variational Assimilation of Meteorological Observation with the Adjoint Vorticity Equation. II: Numerical Results," Q. J. R. Meteorol. Soc. **113**, 1329–1347 (1987).

58. P. Courtier and O. Talagrand, "Variational Assimilation of Meteorological Observation with the Direct and Adjoint Shallow-Water Equations," Tellus A **42**, 531–549 (1990).

59. J. R. Eyre, "Inversion of Cloudy Satellite Sounding Radiances by Nonlinear Optimal Estimation," Q. J. R. Meteorol. Soc. **115**, 1001–1037 (1989).

60. J. R. Eyre, G. A. Kelly, A. P. McNally, *et al.*, "Assimilation of TOVS Radiance Information through One-Dimensional Variational Analysis," Q. J. R. Meteorol. Soc. **119**, 1427–1463 (1993).

61. R. Daley and E. Barker, *The NRL Atmospheric Variational Data Assimilation System (NAVDAS) Source Book* (Monterey, Naval Research Laboratory, 1999).

62. P. Courtier, E. Andersson, W. Heckley, *et al.*, "The ECMWF Implementation of Three-Dimensional Variational Assimilation (3D-Var). I: Formulation," Q. J. R. Meteorol. Soc. **124**, 1783–1808 (1998).

63. F. Rabier, A. Mc. Nally, E. Andersson, *et al.*, "The ECMWF Implementation of Three-Dimensional Variational Assimilation (3D-Var) II: Structure Functions," Q. J. R. Meteorol. Soc. **124**, 1809–1830 (1998).

64. E. Andersson, J. Haseler, P. Unden, *et al.*, "The ECMWF Implementation of Three-Dimensional Variational Assimilation (3D-Uar). III: Experimental Results," Q. J. R. Meteorol. Soc. **124**, 1831–1860 (1998).

65. P. Courtier, J.-N. Thepaut, and A. Hollingworth, "A Strategy for Operational Implementation of 4D-Var, Using an Incremental Approach," Q. J. R. Meteorol. Soc. **120**, 1367–1388 (1994).

66. S. Cohn, "An Introduction to Estimation Theory," J. Meteorol. Soc. Jpn. **75**, 257–288 (1997).

67. *Proceedings of ECMWF Seminar on Recent Developments in Data Assimilation for Atmosphere and Ocean, Reading, UK, 2003* (ECMWF, Reading, UK, 2003). Available at http://www.ecmwf.int/publications/library/do/references/list/17111.

68. A. Simmons, "Observations, Assimilation and the Improvement of Global Weather Prediction—Some Results from Operational Forecasting and ERA-40" in *Proceedings of ECMWF Seminar on Recent Developments in Data Assimilation for Atmosphere and Ocean, Reading, UK, 2003* (ECMWF, Reading, UK, 2003), pp. 1–28.