

Methods for Observation Data Assimilation in Problems of Physics of Atmosphere and Ocean

V. P. Shutyaev^{a, b, *}

^a*Institute of Numerical Mathematics, Russian Academy of Sciences, Moscow, 199333 Russia*

^b*Marine Hydrophysical Institute, Russian Academy of Sciences, Sevastopol, 299011 Russia*

**e-mail: victor.shutyaev@mail.ru*

Received June 14, 2018; revised July 25, 2018; accepted November 28, 2018

Abstract—This paper presents a review and analysis of approaches to data assimilation in problems of geophysical hydrodynamics, from the simplest sequential assimilation schemes to modern variational methods. Special attention is paid to the study of the problem of variational assimilation in a weak formulation, in particular, to the construction of an optimality system and the estimation of the covariance matrices of the optimal solution errors. This is a new direction of research in which the author has obtained some results.

Keywords: observation data assimilation, variational problem, optimal solution, weak formulation, covariance matrix

DOI: 10.1134/S0001433819010080

INTRODUCTION

In recent decades, data have been collected, processed, and interpreted; retrospective analysis has been carried out; and high-quality models of subsystems of the global biogeochemical system have been developed within a number of international research projects on global changes on the Earth in order to create predictive models taking into account modern models of the global climate system. Currently, a coordinated approach to the development and application of global climate system models (including the atmosphere, oceans, the cryosphere, and the biosphere), as well as the estimation of the sensitivity of climate predictability using such coupled models, is being developed.

There is increasing interest in the problems of assimilating and processing observational data for retrospective analysis in different branches of knowledge in connection with studies of global changes on Earth. In problems of geophysical hydrodynamics, in particular, in meteorology and oceanography, mathematical models are used to study and predict hydrodynamic fields. These models are based on the laws of hydrodynamics, which follow from the conservation of mass, momentum, energy, etc., which leads to systems of nonlinear partial differential equations. These equations, although necessary, are insufficient to predict the evolution of fields. Additional information is required, including, in particular, initial conditions and model parameters. This information can be obtained by observations. Data assimilation methods are used to predict the state of the flow at the right moment based on all available observations.

In recent decades, significant progress has been made in Earth sciences due to improved observation systems and understanding of laws of geosystems. A fairly accurate description of the initial conditions is one of the fundamental requirements for successful forecasting in oceanography. Data assimilation aims to obtain the best (in a certain sense) estimation of the state of a physical system from its observations and an adequate mathematical model.

The data assimilation method is widely used in Earth sciences. It is most popular in meteorology and oceanography, where atmospheric and ocean observations are assimilated into atmospheric and oceanic models in order to obtain initial conditions (or other model parameters) for further modeling and forecasting. In recent years, data assimilation methods have been also used to analyze other observations of the geosystem including the biosphere, the cryosphere, and the soil surface.

Researchers have always wanted not only to know and understand the climatic and current states of hydrodynamic fields in the atmosphere and the ocean, but also to be able to predict them. It is necessary to estimate the current state, which, in turn, depends on a certain state in the past, to make a forecast for the future. The first attempts to estimate the state of the system based on the analysis of observational data were made in meteorology in the middle of the 19th century by Vice Admiral Robert FitzRoy, founder of the British Meteorological Service. Subjective analysis (the simplest data interpolation) was then used by Richardson [1], Charney [2], and Phillips [3].

Eventually, objective analysis replaced manual graphical interpolations of observational data with more rigorous mathematical methods, from polynomial interpolation and sequential estimation algorithms to modern variational methods.

Data assimilation methods were most developed in dynamic meteorology and physical oceanography, as well as in the real-time numerical prediction of atmospheric and oceanic fields. To date, theoretical and practical ideas of data assimilation can be found in technical [4, 5], mathematical [6–9], and geophysical literature [10–13]. The Seventh International Symposium of the World Meteorological Organization on observation data assimilation in meteorology and oceanography (Brazil, September 2017: <http://www.cptec.inpe.br/das2017/>) showed significant progress in the practical application of modern assimilation methods based on the optimal control approach (variational data assimilation) and on the sequential estimation approach (statistical methods), as well as on a combination of both approaches (hybrid methods).

Currently, intensive studies on the development of information computation systems (ICS) using observation data assimilation procedures (satellite, shipboard, etc.) are being conducted in a number of countries. The development of modern information and computing systems can rightly be attributed to the interdisciplinary fundamental problems of computer science, mathematics, physics, and many other areas of science and technology. The development of such ICS's is nowadays necessary from the point of view of economy, national security, and other needs of the state and society. The most important problems here are the implementation of real-time short-term and long-term weather forecasts, determining the areas of high biological productivity, ensuring the safety of navigation and selection of optimal ship routes, controlling the ecology of the sea, detecting and monitoring especially dangerous phenomena (such as storm surges and tsunamis), and predicting marine disasters and estimating possible damage they may cause and the risks arising from them. The problems of monitoring and predicting the state of the environment are of vital importance for human society. New geoinformation technologies, including the technology for developing variational observation data assimilation, make it possible to develop a unified system for monitoring and forecasting geosystems for global monitoring programs.

In recent years, there have been qualitative changes in measurement systems. The global scientific community is receiving more and more measurements of various characteristics of our geosystem. Therefore, the development of technologies for variational observation data assimilation based on modern approaches is an urgent problem.

In this paper, we review and analyze approaches to data assimilation in problems of geophysical hydrodynamics, from the simplest sequential assimilation

schemes to modern variational methods. Special attention is paid to the study of the problem of variational assimilation in a weak formulation, in particular, to the construction of an optimality system and the estimation of the covariance matrices of the optimal solution errors. This is a new direction of research in which the author has obtained some results.

1. METHODS AND APPROACHES TO OBSERVATION DATA ASSIMILATION

1.1. Basic Notation and Formulation of the Problem

Consider a mathematical model that describes the evolution of a hydrodynamic system (atmospheric, oceanic, or coupled) as follows

$$\begin{cases} \frac{dx}{dt} = M(x, t), & t > 0 \\ x|_{t=0} = x_0, \end{cases} \quad (1.1)$$

where x is the state vector of the model, M is the corresponding dynamic operator of the model, and x_0 is the vector of the initial state. In numerical simulation or prediction, dynamic operator M is generally nonlinear and deterministic, while the true flow field differs from (1.1) by a random or systematic error. In geophysical hydrodynamics, (1.1) is usually a system of nonlinear partial differential equations, which is often called a *distributed parameter system* in mathematical literature. The dependent variable x is called the *field*.

Observations are given by some vector function $y^0(t)$, which satisfies the following equation:

$$y^0(t) = H(x^t) + \varepsilon, \quad (1.2)$$

where H is the observation operator, x^t is the true flow field, and ε is the error function (noise). Function $y^0(t)$ is assumed to be given, while there is usually no information about function ε . Operator H , as well as M , can be nonlinear. It sets the mapping of the state vector to the observation space.

Strictly speaking, Eqs. (1.1) and (1.2) should be considered in the corresponding function spaces, and in each specific case it is important to investigate questions of solvability and properties of the solution of the problem for the development of numerical algorithms.

When model (1.1) is discretized over time using finite differences, finite elements, or (pseudo) spectral methods, a discrete model describing the transition from time t_i to time t_{i+1} is often obtained:

$$x(t_{i+1}) = M_i(x(t_i)), \quad (1.3)$$

where $x(t_i)$ is the state vector with a dimension of n , i is the number of the time step, and M_i is the difference operator of the state vector dynamics. In dis-

crete model (1.3), observations y^0 at time t_i are given by the following equation:

$$y_i^0 = H_i(x^t(t_i)) + \varepsilon_i, \quad (1.4)$$

where H_i is the observation operator at time $t = t_i$, x^t is the true state, and ε_i is the error function. Vectors y_i^0 have dimensions p_i . In most practical problems, p_i is much smaller than n .

Additional information (for example, initial conditions and unknown parameters of the model, which can be obtained using observational data) is required to predict the evolution of flows in problems of geophysical hydrodynamics. Thus, the *data assimilation problem* arises: for a given observation function $y^0(t)$, it is required to find, for example, an unknown a priori initial condition so that the state vector x satisfies problem (1.1) and vector $H(x)$ is in any sense close to $y^0(t)$. The resulting solution of x is called a *state estimate* (or *analysis*) and is denoted by x^a .

1.2. Objective Analysis and Its Generalizations

The first attempt at objective data analysis was made by Panovsky [14] using two-dimensional (2D) polynomial interpolation of observational data. Later, this approach was developed by Gilchrist and Cressman [15], who introduced the area of influence for each observation and suggested using the so-called initial approximation field (background), the field from the previous forecast.

The approach of Bergthorsson and Doos [16] is based on an analysis of the difference between the observed data and the initial approximation and optimization of the weight assigned to each observation. This approach was later modified by Cressman [17], who proposed the successive correction method (SCM), which is an iterative algorithm for determining the state vector:

$$x^{k+1} = x^k + W(y^0 - H(x^k)), \quad k = 0, 1, \dots, \quad (1.5)$$

where k is the iteration number, $x^0 = x^b$ is the initial approximation, W is the weight operator, and H is the observation operator from (1.2). After a sufficiently large number of iterations, $x^a = x^k$ is the state estimation. Successive iterations approximate observational data on ever smaller scales, as was shown in [17]. The disadvantage of the method is that the observational data, the errors of which are not taken into account, are approached more and more accurately during iterations. Nevertheless, it is widely used for real-time weather forecasting.

The nudging method, which consists of adding a term such as (1.5) to the right-hand side of dynamic

system (1.1), is a generalization of the method of successive corrections for nonstationary problem (1.1):

$$\begin{cases} \frac{dx}{dt} = M(x, t) + W(y^0 - H(x)), & t \in (0, T) \\ x|_{t=0} = x_0. \end{cases} \quad (1.6)$$

This term causes the model solution to approach the observational data as accurately as possible. This method was first used in meteorology in [18] and later in oceanography in [19–21]. This method is still of interest. Its new versions have appeared, in particular, the BFN algorithm [22].

1.3. Statistical Methods, Sequential Assimilation Algorithms

The use of statistical interpolation methods was a very important breakthrough in solving data assimilation problems. This approach goes back to A.N. Kolmogorov (1941), the works of N. Wiener (1949), and in Earth sciences it became known thanks to the monograph by L.S. Gandin [23]. This approach is also called optimal interpolation (OI) [24, 25]. Observations are assigned weights that are associated with observational errors. At the same time, the initial approximation field is not the first approximation for analysis, as previously, but is used together with its error characteristic along with other observational data.

Let observation operator H be linear and observation function y^0 and the field of the first approximation x^b be given as follows:

$$y^0 = Hx^t + \varepsilon, \quad x^b = x^t + \varepsilon_b, \quad (1.7)$$

where errors ε and ε_b are assumed to be random Gaussian vectors with zero expectation and covariance matrices

$$R = E(\varepsilon\varepsilon^T), \quad B = E(\varepsilon_b\varepsilon_b^T). \quad (1.8)$$

The problem of optimal interpolation is to find the estimate x^a minimizing deviation $x^t - x^a$ based on data (1.7)–(1.8), for example, in the sense of the minimum trace of the covariance matrix of the analysis:

$$\text{tr}E[(x^t - x^a)(x^t - x^a)^T] \rightarrow \min. \quad (1.9)$$

Then the optimal interpolation method consists of determining the analysis of x^a using the following formula [11, 12]:

$$x^a = x^b + BH^*(HBH^* + R)^{-1}(y^0 - Hx^b), \quad (1.10)$$

where H^* is the operator adjoint to H .

According to (1.10), x^a is computed as the field of the initial approximation x^b plus the correction, which is nothing but the result of the action of some weight

operator on the vector $y^0 - Hx^b$. The latter is called the *innovation vector* or *residual of observations*.

It can be seen that the optimal interpolation method in the form of (1.10) is equivalent to the optimal control problem, which reduces to finding the minimum of a quadratic functional:

$$J(x) = \frac{1}{2}(B^{-1}(x - x^b), x - x^b) + \frac{1}{2}(R^{-1}(Hx - y^0), Hx - y^0).$$

To do this, its first derivative should vanish:

$$J'_x(x)\delta x = (B^{-1}(x - x^b), \delta x) + (R^{-1}(Hx - y^0), H\delta x) = 0.$$

Hence, we obtain

$$x^a = (B^{-1} + H^*R^{-1}H)^{-1}(B^{-1}x^b + H^*R^{-1}y^0) = x^b + BH^*(HBH^* + R)^{-1}(y^0 - Hx^b).$$

Optimal interpolation algorithm (1.10) can be divided into the following steps:

$$(HBH^* + R)\xi = y^0 - Hx^b, \quad (1.11)$$

$$\theta = BH^*\xi, \quad (1.12)$$

$$x^a = x^b + \theta, \quad (1.13)$$

where Eq. (1.11) is written in the space of observations and (1.12)–(1.13) is written in the state space.

The optimal interpolation method has been used in many operations centers since the late 1970s [24, 26]. Later, this method was developed in the works of Lorenz [27, 28], who used different approximations to solve equations (1.11)–(1.13) and introduced the analysis correction method, which is a “hybrid” of optimal interpolation and successive corrections.

The optimal interpolation method and its modifications have so far been most widely used for real-time data analysis in weather forecasting [27, 29–32], as well as in assimilation of oceanographic data [33–35]. Ensemble optimal interpolation (EnOI) [36, 37], which makes it possible to construct parallel data assimilation algorithms [38], has gained great popularity.

The Kalman filter, which extrapolates dynamic variables and their covariances at each step and then recursively refines the state estimate [39], is a generalization of the optimal interpolation method. The continuous analog of this method is often called the Kalman–Bucy filter [40]. There are different generalizations of this method to the nonlinear case [4]. Currently, the extended Kalman filter (EKF) [41, 42],

which uses model linearization near a certain known state, is very popular. A.S. Sarkisyan, V.V. Knysh, G.A. Korotaev, etc. [43–47] made a significant contribution to the development of Kalman filter methods and methods of 4D analysis of hydrophysical fields based on dynamic–stochastic models of the ocean. In recent years, the ensemble Kalman filter (EnKF) [48–50], which is based on the Monte Carlo method at every time step, has become very popular.

1.4. Variational Methods

The use of variational methods and, in particular, optimal control methods was significant progress in solving data assimilation problems. The idea of minimizing a certain functional related to observational data on the trajectories (solutions) of the model under consideration turned out to be very productive. Thus, the data assimilation problem is formulated as an optimal control problem. The theoretical foundations of research and solution of such problems were laid in the classical works of R. Bellman (1957), L.S. Pontryagin (1962), N.N. Krasovskiy (1969), J.-L. Lions (1968), G.I. Marchuk (1961), etc. Variational formalism was used for the first time in Sasaki meteorology [51, 52] and in problems of dynamic oceanography, Provost and Salmon [53].

It is necessary to calculate the gradient of the original functional when solving minimization problems. One important step in this direction was the use of the theory of adjoint equations (Marchuk, 1964; Lions, 1968). Adjoint equations have been widely used for the study and numerical solution of data assimilation problems (including the calculation of the gradient of the functional) by many researchers [58–68], starting from the well-known works (Penenko [54], Marchuk and Penenko [55], Le Dimet and Talagrand [56], Lewis and Derber [57]).

Three-dimensional variational data assimilation (3D-VAR) was used for real-time analysis for the first time at the National Center for Environmental Prediction (NCEP) [69] and later at the European Centre for Medium-Range Weather Forecasts (ECMWF) [70].

Currently, four-dimensional data assimilation (4D-VAR) is attracting more and more attention, in which linearized and adjoint models are used to assimilate observational data not at a specific time, but at a given time interval. The 4D-VAR system was used for the first time at the ECMWF [71].

Let us dwell on the formulation of the problem of 4D-VAR data assimilation using the example of the problem of restoring the initial condition. Consider problem (1.1) on the interval $(0, T)$:

$$\begin{cases} \frac{dx}{dt} = M(x, t), & t \in (0, T) \\ x|_{t=0} = x_0 \end{cases} \quad (1.14)$$

and introduce the functional of its solution:

$$J(x_0) = \frac{1}{2} \left(C_1 (x|_{t=0} - x_0^b), x|_{t=0} - x_0^b \right) + \frac{1}{2} \int_0^T (C_2 (Hx - y^0), Hx - y^0) dt, \quad (1.15)$$

where H is the (linear) observation operator from (1.2), y^0 is the observation function, x_0^b is the given vector, C_1, C_2 are weight operators, and (\cdot, \cdot) is the scalar product. Usually, C_1, C_2 are selected in the following form: $C_1 = B^{-1}, C_2 = R^{-1}$, where B, R is the covariance matrix of vectors $\varepsilon_b = x_0^b - x^t|_{t=0}$ and $\varepsilon = y^0 - Hx^t$, respectively: $B = E(\varepsilon_b \varepsilon_b^T), R = E(\varepsilon \varepsilon^T)$ under the assumption that ε and ε_b are random Gaussian vectors with zero expectation. Such weight operators (or their approximations) are often used in practical problems [12, 72, 73].

Suppose that initial condition x_0 from (1.14) is unknown to us. Then the simplest *data assimilation problem* is formulated as follows: find x_0, x such that they satisfy (1.14) and functional (1.15) reaches its smallest value on set of solutions (1.14). In other words,

$$\begin{cases} \frac{dx}{dt} = M(x, t), & t \in (0, T) \\ x|_{t=0} = x_0, \\ J(x_0) = \inf_v J(v). \end{cases} \quad (1.16)$$

The necessary optimality condition [6] leads this problem to a system for three unknowns x_0, x, x^* :

$$\begin{cases} \frac{dx}{dt} = M(x, t), & t \in (0, T) \\ x|_{t=0} = x_0, \end{cases} \quad (1.17)$$

$$\begin{cases} -\frac{dx^*}{dt} = (M(x, t))^* x^* - H^* C_2 (Hx - y^0), & t \in (0, T) \\ x^*|_{t=T} = 0, \end{cases} \quad (1.18)$$

$$C_1 (x_0 - x_0^b) - x^*|_{t=0} = 0, \quad (1.19)$$

where $(M(x, t))^*$ is the operator adjoint to the derivative of the dynamic operator of model M . System (1.17)–(1.19) is called the *optimality system* and plays an important role in the study and numerical solution of data assimilation problems. This system can also be obtained from the Pontryagin maximum principle formulated for problem (1.16) [61] or by the Lagrange multipliers method [74].

The solvability of nonlinear data assimilation problems and rigorous justification of numerical methods for their solution is not a simple problem. Sufficiently complete results concerning the solvability of linear

optimal control problems of form (1.16)–(1.19) were obtained by Lions using the Hilbert Uniqueness Method (HUM) that he developed. Some results about the solvability of weakly nonlinear data assimilation problems were obtained in [60, 75, 67]. Further generalizations and new applications have been proposed in recent years [76, 77, 68].

Problems of form (1.16) are currently numerically solved by well-known optimization algorithms developed in classical works. A number of new iterative algorithms for solving data assimilation problems using adjoint equations were proposed in [60–63, 67, 68], etc.

It is possible to use known minimization methods for problem (1.16) or solve optimality system (1.17)–(1.19) to construct a numerical algorithm for solving the data assimilation problem. In the numerical solution of a problem, it is often necessary to calculate the gradient of original functional J . This can be done using an adjoint problem selected in a suitable manner. In the case under consideration, the gradient of the functional is calculated as follows: for a given v , we successively find the solutions of the direct and adjoint problems

$$\begin{cases} \frac{dx}{dt} = M(x, t), & t \in (0, T) \\ x|_{t=0} = v, \end{cases} \quad (1.20)$$

$$\begin{cases} -\frac{dx^*}{dt} = (M(x, t))^* x^* - H^* C_2 (Hx - y^0), & t \in (0, T) \\ x^*|_{t=T} = 0 \end{cases} \quad (1.21)$$

and put

$$J'(v) = C_1 (v - x_0^b) - x^*|_{t=0}. \quad (1.22)$$

In the works of many authors, much attention is paid to the numerical construction of adjoint model (1.21), which can be obtained both by the discretization of continuous problem (1.21) [66, 78, 79] and the direct transposition of the code of the discrete linearized problem [68, 80, 81]. In the latter case, automatic differentiation methods are often used [82, 83, 74]. These two approaches to the construction of a discrete adjoint problem were compared, for example, in [78, 84].

The properties of the optimal solution itself play an important role, along with the study of solvability, as well as the development and justification of algorithms for the numerical solution of problems of variational data assimilation. The question of the sensitivity of optimal solutions of variational assimilation problems to the errors of observational data and model errors is extremely important. This question has been, until recently, little studied. However, a number of results have been obtained using control operators over the last few years [85–92]. Equations for the optimal solution error were obtained and investigated through the errors of the observational data in the problem of restoring the initial condition. The sensitivity of the

optimal solution was investigated using singular vectors of control operators. It turned out that fundamental control functions, which are singular vectors of response operators, play an important role in the study of errors [85, 88, 90, 92].

Currently, 4D data assimilation algorithms [10, 11, 13] seem to be the most effective. In recent years there have been many studies comparing the ensemble Kalman method (EnKF) and variational data assimilation [93–99]. In addition, a so-called hybrid approach has appeared. It combines the ensemble Kalman method and the variational data assimilation, Hybrid 4D-VAR [100–104], as well as the ensemble method of 4D-VAR assimilation, 4D-EnVar [105–110].

2. COVARIANCE MATRICES OF OPTIMAL SOLUTION ERRORS

The a posteriori covariance matrix is an important characteristic of the optimal solution obtained from the optimality system of the variational data assimilation problem. This section is devoted to the development of algorithms for the study of covariance operators of errors of optimal solutions of problems of variational data assimilation using the cost functional Hessian. The theoretical foundations of the algorithms were laid in [85, 111–113].

Consider the variational data assimilation problem by the example of initialization problem (1.14), for which optimality system (1.17)–(1.19) is valid. It is assumed that the input data are given with errors: $x_0^b = x_0^t + \varepsilon_b, y^0 = Hx^t + \varepsilon$, where $\varepsilon_b \sim N(0, B)$, $\varepsilon \sim N(0, R)$, and x^t is the exact solution of problem (1.14) for $x_0 = x_0^t$:

$$\begin{cases} \frac{dx^t}{dt} = M(x^t, t), & t \in (0, T) \\ x^t|_{t=0} = x_0^t. \end{cases} \quad (2.1)$$

Here, $\varepsilon \sim N(0, R)$ means that the random variable ε is distributed according to the Gaussian law with zero expectation and covariance matrix R . We will investigate the influence of errors $\varepsilon_b, \varepsilon$ on the optimal solution x_0 obtained by solving (1.17)–(1.19) and formulate algorithms for calculating the covariance operators of optimal solution errors through the cost functional Hessian.

System (1.17)–(1.19) with three unknowns x, x^*, x_0 can be considered one operator equation of the following form:

$$F(U, U_d) = 0, \quad (2.2)$$

where $U = (x, x^*, x_0), U_d = (x_0^b, y^0)$. A similar equation holds for the following exact solution:

$$F(\bar{U}, \bar{U}_d) = 0, \quad (2.3)$$

where $U = (x^t, x^*, x_0^t), U_d = (x_0^t, Hx^t), x^* = 0$. System (2.3) is a necessary condition for the optimality of the following minimization problem: find u and ϕ such that

$$\begin{cases} \frac{d\phi}{dt} = M(\phi, t), & t \in (0, T) \\ \phi|_{t=0} = u, \\ \bar{J}(u) = \inf_v \bar{J}(v), \end{cases}$$

where

$$\begin{aligned} \bar{J}(u) &= \frac{1}{2} (C_1 (u - x_0^t), u - x_0^t) \\ &+ \frac{1}{2} \int_0^T (C_2 (H\phi - Hx^t), H\phi - Hx^t) dt. \end{aligned}$$

From (2.2)–(2.3), we have $F(U, U_d) - F(\bar{U}, \bar{U}_d) = 0$. Let $\delta U = U - \bar{U}, \delta U_d = U_d - \bar{U}_d$. Then

$$F(\bar{U} + \delta U, \bar{U}_d + \delta U_d) - F(\bar{U}, \bar{U}_d) = 0. \quad (2.4)$$

Let $\delta x = x - x^t, \delta x_0 = x_0 - x_0^t$, and then $\delta U = (\delta x, x^*, \delta x_0), \delta U_d = (\varepsilon_b, \varepsilon)$. Assuming that operator M is sufficiently smooth, $\underline{x} = x^t + \tau(x - x^t), \tau \in [0, 1]$ exists such that $M(x, t) - M(x^t, t) = M'(\underline{x}, t)\delta x$. Then Eq. (2.4) is equivalent to the following system:

$$\begin{cases} \frac{d\delta x}{dt} = M'(\underline{x}, t)\delta x, & t \in (0, T) \\ \delta x|_{t=0} = \delta x_0, \end{cases} \quad (2.5)$$

$$\begin{cases} -\frac{dx^*}{dt} = (M'(x, t))^* x^* \\ -H^* C_2 (H\delta x - \varepsilon), & t \in (0, T) \\ x^*|_{t=T} = 0, \end{cases} \quad (2.6)$$

$$C_1 (\delta x_0 - \varepsilon_b) - x^*|_{t=0} = 0. \quad (2.7)$$

Since $\underline{x} = x^t + \tau\delta x, x = x^t + \delta x$, then, under the assumption of smallness of δx and smoothness of M in (2.5)–(2.7), we can suppose

$$M'(\underline{x}, t) \approx M'(x^t, t), (M'(x, t))^* \approx (M'(x^t, t))^*. \quad (2.8)$$

Then (2.5)–(2.7) reduces to the following system

$$\begin{cases} \frac{d\delta x}{dt} = M'(x^t, t)\delta x, & t \in (0, T) \\ \delta x|_{t=0} = \delta x_0, \end{cases} \quad (2.9)$$

$$\begin{cases} -\frac{dx^*}{dt} = (M'(x^t, t))^* x^* \\ -H^* C_2 (H\delta x - \varepsilon), & t \in (0, T) \\ x^*|_{t=T} = 0, \end{cases} \quad (2.10)$$

$$C_1(\delta x_0 - \varepsilon_b) - x^*|_{t=0} = 0. \quad (2.11)$$

System (2.9)–(2.11) is the linear data assimilation problem. For a fixed x^t , this is a necessary condition for the optimality of the following minimization problem: find u and ϕ such that

$$\begin{cases} \frac{d\phi}{dt} = M(x^t, t)\phi, & t \in (0, T) \\ \phi|_{t=0} = u, \\ J_1(u) = \inf_v J_1(v), \end{cases}$$

where

$$J_1(u) = \frac{1}{2}(C_1(u - \varepsilon_b), u - \varepsilon_b) + \frac{1}{2} \int_0^T (C_2(H\phi - \varepsilon), H\phi - \varepsilon) dt. \quad (2.12)$$

Consider the Hessian \mathfrak{H} of functional (2.12). It is defined on v by the successive solution of problems

$$\begin{cases} \frac{d\delta x}{dt} = M(x^t, t)\delta x, & t \in (0, T) \\ \delta x|_{t=0} = v, \end{cases} \quad (2.13)$$

$$\begin{cases} -\frac{d\theta^*}{dt} = (M(x^t, t))^* \theta^* - H^* C_2 H \delta x, & t \in (0, T) \\ \theta^*|_{t=T} = 0, \end{cases} \quad (2.14)$$

$$\mathfrak{H}v = C_1 v - x^*|_{t=0}. \quad (2.15)$$

Let us introduce the auxiliary operators R_1, R_2 . Let $R_1 = C_1$, and the operator R_2 be defined on the functions g by the formula $R_2 g = \theta^*|_{t=0}$, where θ^* is the solution of the adjoint problem

$$\begin{cases} -\frac{d\theta^*}{dt} = (M(x^t, t))^* \theta^* + H^* C_2 g, \\ \theta^*|_{t=T} = 0. \end{cases} \quad (2.16)$$

From (2.13)–(2.16), we conclude that system (2.9)–(2.11) is equivalent to the equation for the error δx_0 :

$$\mathfrak{H}\delta x_0 = R_1 \varepsilon_b + R_2 \varepsilon. \quad (2.17)$$

Hessian \mathfrak{H} is by definition a symmetric nonnegative definite operator. We will assume that \mathfrak{H} is positive definite and, thus, invertible. Then Eq. (2.17) can be written as follows:

$$\delta x_0 = T_1 \varepsilon_b + T_2 \varepsilon, \quad (2.18)$$

where $T_i = \mathfrak{H}^{-1} R_i$, $i=1, 2$.

We assume that errors $\varepsilon_b, \varepsilon$ are random, normally distributed with zero mean, and uncorrelated among themselves while, as was mentioned above, $\varepsilon_b \sim N(0, B), \varepsilon \sim N(0, R)$. Then it follows from (2.18) that error δx_0 is also normally distributed with zero expectation. Let P denote the covariance operator of

the error of the optimal solution: $P = E[\delta x_0(\delta x_0)^T]$. From (2.18), we obtain

$$P = E[\delta x_0(\delta x_0)^T] = T_1 B T_1^* + T_2 R T_2^*, \quad (2.19)$$

where T_i^* are operators adjoint to T_i , $i=1, 2$. It is necessary to find operators $T_1 B T_1^*, T_2 R T_2^*$ to construct operator P . Consider operator $T_1 B T_1^*$. Since $T_1 = \mathfrak{H}^{-1} R_1 = \mathfrak{H}^{-1} C_1 = T_1^*$, then $T_1 B T_1^* = \mathfrak{H}^{-1} C_1 B C_1 \mathfrak{H}^{-1}$. Moreover, if $C_1 = B^{-1}$, then

$$T_1 B T_1^* = \mathfrak{H}^{-1} C_1 B C_1 \mathfrak{H}^{-1} = \mathfrak{H}^{-1} B^{-1} \mathfrak{H}^{-1}. \quad (2.20)$$

Thus, the algorithm for calculating $w = T_1 B T_1^* v$ is as follows:

- (1) solve equation $\mathfrak{H}p = v$,
- (2) calculate $C_1 p$,
- (3) solve equation $\mathfrak{H}w = C_1 p$.

As a result, Eq. (2.20) gives the contribution of error ε_b to covariance operator P .

Consider now operator $T_2 R T_2^*$. Since $T_2 = \mathfrak{H}^{-1} R_2$, then $T_2 R T_2^* = \mathfrak{H}^{-1} R_2 R R_2^* \mathfrak{H}^{-1}$. Let us consider the scalar product $(R_2 g, p)$ for fixed g and p to find R_2^* . We have from (2.16) that $(R_2 g, p) = (\theta^*|_{t=0}, p) = \int_0^T (H^* C_2 g, \phi) dt = (g, R_2^* p)$, where $R_2^* p = C_2 H \phi$, and ϕ is the solution of problem (2.13) for $v = p$. Thus, operator $T_2 R T_2^*$ is determined by successively solving the following problems (for a given v):

$$\mathfrak{H}p = v, \quad (2.21)$$

$$\begin{cases} \frac{d\phi}{dt} = M(x^t, t)\phi, & t \in (0, T) \\ \phi|_{t=0} = p, \end{cases} \quad (2.22)$$

$$\begin{cases} -\frac{d\theta^*}{dt} = (M(x^t, t))^* \theta^* - H^* C_2 R C_2 H \phi, & t \in (0, T) \\ \theta^*|_{t=T} = 0, \end{cases} \quad (2.23)$$

$$\mathfrak{H}w = \theta^*|_{t=0}. \quad (2.24)$$

Then $T_2 R T_2^* v = w$. If $C_2 = R^{-1}$, then $H^* C_2 R C_2 H = H^* C_2 H$ and from (2.22)–(2.23) we have that $\theta^*|_{t=0} = \mathfrak{H}v - C_1 v$, where \mathfrak{H} is the Hessian given by formulas (2.13)–(2.15). Then $R_2 R R_2^* = \mathfrak{H} - C_1$ and

$$T_2 R T_2^* = \mathfrak{H}^{-1} (\mathfrak{H} - C_1) \mathfrak{H}^{-1}. \quad (2.25)$$

Thus, the algorithm for calculating $w = T_2 R T_2^* v$ is as follows:

- (1) solve equation $\mathfrak{H}p = v$,
- (2) calculate $(\mathfrak{H} - C_1)p$,

(3) solve equation $\zeta w = (\zeta - C_1)p$.

As a result, Eq. (2.25) gives the contribution of error ε to covariance operator P .

Equations (2.20) and (2.25) should be added to calculate the total contribution of uncorrelated errors $\varepsilon_b, \varepsilon$. Then

$$P = T_1 B T_1^* + T_2 R T_2^* = \zeta^{-1} C_1 \zeta^{-1} + \zeta^{-1} (\zeta - C_1) \zeta^{-1},$$

from which we obtain the following:

$$P = \zeta^{-1}. \tag{2.26}$$

The latter formula gives the form of covariance operator P through the Hessian defined by Eqs. (2.13)–(2.15). The rule of ζ multiplication by some function v according to (2.13)–(2.15) or the BFGS method [114, 115], which gives an approximation for ζ^{-1} as a result of iterations, can be used to calculate the inverse Hessian.

3. WEAK FORMULATION OF VARIATIONAL DATA ASSIMILATION

The formulation of the problem of variational data assimilation in the form of (1.17)–(1.19) is often called a strong formulation or a formulation with strong constraints. The strong constraints are the equations of model (1.17), which must be strictly satisfied if functional (1.15) is minimized. A significant shortcoming of formulation (1.17) is that the model is assumed to be accurate and the model errors are not taken into account. Model errors can be associated with discretization, with an inaccurate description of physical processes, and with errors in the input data. The so-called weak formulation or formulation with weak constraints is considered to take into account possible errors of the model [51, 52], [116–125].

In the case of a weak formulation of the equation, the models are no longer necessarily accurate and they are included in the original cost functional. Thus, we consider the problem of minimizing the following functional instead of problem (1.14):

$$J = \frac{1}{2} (C_1 (x|_{t=0} - x_0^b), x|_{t=0} - x_0^b) + \frac{1}{2} \int_0^T (C_2 (Hx - y^0), Hx - y^0) dt + \frac{1}{2} \int_0^T \left(C_3 \left(\frac{dx}{dt} - M(x, t) \right), \frac{dx}{dt} - M(x, t) \right) dt, \tag{3.1}$$

where C_1, C_2, C_3 are weight operators: $C_1 = B^{-1}$, $C_2 = R^{-1}$, $C_3 = Q^{-1}$, and B, R, Q are covariance matrices of vectors $\varepsilon_b, \varepsilon, \xi$, respectively: $B = E[\varepsilon_b \varepsilon_b^T]$, $R = E[\varepsilon \varepsilon^T]$, $Q = E[\xi \xi^T]$. Error vectors $\varepsilon_b, \varepsilon$ were introduced ear-

lier, and ξ is the model error: $\xi = M(x', t) - \frac{dx'}{dt}$. Let us

introduce the notation $f = \frac{dx}{dt} - M(x, t)$. Then it can be seen that functional minimization problem (3.1) by x can be formulated in an equivalent form; namely, it is necessary to find x_0, f such that

$$\begin{cases} \frac{dx}{dt} = M(x, t) + f, & t \in (0, T) \\ x|_{t=0} = x_0, \\ J(x_0, f) = \inf_{v, g} J(v, g), \end{cases} \tag{3.2}$$

where

$$J(x_0, f) = \frac{1}{2} (C_1 (x_0 - x_0^b), x_0 - x_0^b) + \frac{1}{2} \int_0^T (C_2 (Hx - y^0), Hx - y^0) dt + \frac{1}{2} \int_0^T (C_3 f, f) dt.$$

Thus, functional minimization problem (3.1) again reduces to a problem with strong constraints, but here the unknowns (controls) are not only the function of the initial condition x_0 , but also the right side f .

We calculate the gradients J in x_0 and f , respectively, to construct an optimality system. By definition of the gradient,

$$\begin{aligned} J'_{x_0}(x_0, f) \delta x_0 &= (C_1 (x_0 - x_0^b), \delta x_0) \\ &+ \int_0^T (C_2 (Hx - y^0), H \delta x_1) dt, \\ J'_f(x_0, f) \delta f &= \int_0^T (C_2 (Hx - y^0), H \delta x_2) dt + \int_0^T (C_3 f, \delta f) dt, \end{aligned}$$

where $\delta x_1, \delta x_2$ are the solutions of the following problems:

$$\begin{cases} \frac{d\delta x_1}{dt} = M(x, t) \delta x_1, & t \in (0, T) \\ \delta x_1|_{t=0} = \delta x_0, \end{cases} \tag{3.3}$$

$$\begin{cases} \frac{d\delta x_2}{dt} = M(x, t) \delta x_2 + \delta f, & t \in (0, T) \\ \delta x_2|_{t=0} = 0. \end{cases} \tag{3.4}$$

It can be seen that their sum $\delta x = \delta x_1 + \delta x_2$ satisfies the system in variations:

$$\begin{cases} \frac{d\delta x}{dt} = M(x, t) \delta x + \delta f, & t \in (0, T) \\ \delta x|_{t=0} = \delta x_0. \end{cases} \tag{3.5}$$

We introduce the adjoint problem with respect to (3.5) in form (1.18) to construct gradients explicitly. Then we obtain the following gradients using the well-known conjugacy relation [9]

$$J'_{x_0}(x_0, f)\delta x_0 = (C_1(x_0 - x_0^b) - x^*|_{t=0}, \delta x_0),$$

$$J'_f(x_0, f)\delta f = \int_0^T (C_3 f - x^*, \delta f) dt,$$

which should be set to zero for any $\delta x_0, \delta f$. Thus, the necessary optimality condition leads the problem to the system for four unknowns x_0, f, x, x^* :

$$\begin{cases} \frac{dx}{dt} = M(x, t) + f, & t \in (0, T) \\ x|_{t=0} = x_0, \end{cases} \quad (3.6)$$

$$\begin{cases} -\frac{dx^*}{dt} = (M(x, t))^* x^* - H^* C_2 (Hx - y^0), & t \in (0, T) \\ x^*|_{t=T} = 0, \end{cases} \quad (3.7)$$

$$C_1(x_0 - x_0^b) - x^*|_{t=0} = 0, \quad (3.8)$$

$$C_3 f - x^* = 0. \quad (3.9)$$

Note that adjoint problem (3.7) in the obtained optimality system coincides with adjoint problem (1.18), and condition (3.8) coincides with (1.19), the condition for equality of the gradient in x_0 to zero.

Let us construct covariance matrix P of optimal solution errors $\delta u = (\delta x_0, \delta f)^T$, where $\delta x_0 = x_0 - x^t|_{t=0}$, $\delta f = f - f^t$, which is determined by formula $P = E[\delta u(\delta u)^T]$. For this purpose, we write the optimality system for errors, which under assumption (2.8) reduces to the following system, similarly to (2.5)–(2.7):

$$\begin{cases} \frac{d\delta x}{dt} = M(x^t, t)\delta x + \delta f, & t \in (0, T) \\ \delta x|_{t=0} = \delta x_0, \end{cases} \quad (3.10)$$

$$\begin{cases} -\frac{d\delta x^*}{dt} = (M(x^t, t))^* \delta x^* \\ -H^* C_2 (H\delta x - \varepsilon), & t \in (0, T) \\ \delta x^*|_{t=T} = 0, \end{cases} \quad (3.11)$$

$$C_1(\delta x_0 - \varepsilon_b) - \delta x^*|_{t=0} = 0, \quad (3.12)$$

$$C_3(\delta f - \xi) - \delta x^* = 0. \quad (3.13)$$

System (3.10)–(3.13) is nothing more than a condition of optimality for the following linear data assimilation problem: find $\delta x_0, \delta f$ such that

$$\begin{cases} \frac{d\delta x}{dt} = M(x^t, t)\delta x + \delta f, & t \in (0, T) \\ x|_{t=0} = \delta x_0, \\ J_2(\delta x_0, \delta f) = \inf_{v, g} J_2(v, g), \end{cases} \quad (3.14)$$

where

$$\begin{aligned} J_2(\delta x_0, \delta f) &= \frac{1}{2}(C_1(\delta x_0 - \varepsilon_b), \delta x_0 - \varepsilon_b) \\ &+ \frac{1}{2} \int_0^T (C_2(H\delta x - \varepsilon), H\delta x - \varepsilon) dt \\ &+ \frac{1}{2} \int_0^T (C_3(\delta f - \xi), f - \xi) dt. \end{aligned}$$

Let us introduce the Hessian \mathfrak{H} of the functional J_2 ; it is defined on $u = (v, g)^T$ by the sequential solution of problems:

$$\begin{cases} \frac{d\delta x}{dt} = M(x^t, t)\delta x + g, & t \in (0, T) \\ \delta x|_{t=0} = v, \end{cases} \quad (3.15)$$

$$\begin{cases} -\frac{d\delta x^*}{dt} = (M(x^t, t))^* \delta x^* - H^* C_2 H\delta x, & t \in (0, T) \\ \delta x^*|_{t=T} = 0, \end{cases} \quad (3.16)$$

$$\mathfrak{H}v = (C_1 v - x^*|_{t=0}, C_3 g - x^*)^T. \quad (3.17)$$

Then it can be seen that system (3.10)–(3.13) is equivalent to the equation for the error $\delta u = (\delta x_0, \delta f)^T$:

$$\mathfrak{H}\delta u = \mathfrak{R}_1 \Xi + \mathfrak{R}_2 \varepsilon, \quad (3.18)$$

where $\Xi = (\varepsilon_b, \xi)^T$, $\mathfrak{R}_1 \Xi = (C_1 \varepsilon_b, C_3 \xi)^T$, $\mathfrak{R}_2 \varepsilon = (\theta^*|_{t=0}, \theta^*)^T$, and θ^* is the solution of adjoint problem (2.16). Assuming that operator \mathfrak{H} from (3.18) is invertible, we have

$$\delta u = \mathfrak{S}_1 \Xi + \mathfrak{S}_2 \varepsilon, \quad (3.19)$$

where $\mathfrak{S}_i = \mathfrak{H}^{-1} \mathfrak{R}_i, i = 1, 2$. The latter equation can be used to construct covariance operator P for the optimal solution errors: $P = E[\delta u(\delta u)^T]$. From (3.19), we obtain the following

$$P = \mathfrak{S}_1 V_{\Xi} \mathfrak{S}_1^* + \mathfrak{S}_2 R \mathfrak{S}_2^*, \quad (3.20)$$

where $V_{\Xi} = E[\Xi \Xi^T] = \begin{pmatrix} B & 0 \\ 0 & Q \end{pmatrix}$, assuming that errors

$\varepsilon_b, \varepsilon, \xi$ are uncorrelated. Following the scheme of the proof of (2.20)–(2.25), it can be shown that

$$\mathfrak{S}_1 V_{\Xi} \mathfrak{S}_1^* = \mathfrak{H}^{-1} V_{\Xi}^{-1} \mathfrak{H}^{-1}, \quad (3.21)$$

$$\mathfrak{S}_2 R \mathfrak{S}_2^* = \mathfrak{H}^{-1} (\mathfrak{H} - V_{\Xi}^{-1}) \mathfrak{H}^{-1},$$

where \mathfrak{H} is the Hessian defined by formulas (3.15)–(3.17). Then, from (3.21) we conclude:

$$P = \mathfrak{H}^{-1}. \quad (3.22)$$

Thus, we obtain a result similar to (2.26) only with a different operator \mathfrak{H} .

Operator \mathfrak{H} , defined by formulas (3.15)–(3.17), can be written in matrix form as follows:

$$\mathfrak{H} = \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix},$$

where H_{ij} are combinations of derivatives with respect to x and f . Thus, the dimension of \mathfrak{H} in this case increases by an order of magnitude compared with (2.26), because in problem (3.2) it is necessary to find not only the function of the initial condition x_0 but also the right-hand side f , which depends on time and space variables. Other ways of accounting for model errors using reduction in the dimension of the problem were considered in [119–121, 124, 125].

4. COMPARATIVE ANALYSIS OF 4D-VAR AND ENSEMBLE KALMAN FILTER

As follows from the review presented above, the new generations of assimilation schemes are based on 4D-VAR data assimilation (4D-VAR) and the ensemble Kalman filter (EnKF). Each of these modern approaches has its advantages and disadvantages, and quite a lot of work has been devoted to their comparative analysis (see, for example, [93–99]).

In the case of a linear model, a linear observation operator, and Gaussian errors, the 4D-VAR methods and the Kalman filter give identical results at the end of the assimilation “window” if model errors are not taken into account [27]. The EnKF method approximates the Kalman filter well [36] under the same assumptions and with a sufficiently large number of elements of the ensemble. Nonlinearities of the model and the observation operator (and, as a consequence, the non-Gaussianity of errors) are a potential cause of the discrepancy in the results when using 4D-VAR and EnKF [95]. If the errors of observations and initial approximation remain Gaussian and the dynamics model is nonlinear, the 4D-VAR method gives an estimate of maximum likelihood—the mode of the distribution function of a posteriori conditional probability [126]. At the same time, in general, it is not clear how the search for such a mode is associated with the result of the EnKF method [95].

In most problems of geophysical hydrodynamics, the dimension of the state vector of the system is so large that it is necessary to find a compromise between computational capabilities and theoretically optimal approaches. For example, the EnKF method has sampling errors due to the limited size of the ensemble, and it is necessary to search for approximations of the

initial approximation covariance matrices in the 4D-VAR method due to its large dimension, which also leads to errors that are difficult to estimate using comparative analysis.

As we saw above, 4D-VAR data assimilation (4D-VAR) in form (1.17)–(1.19) uses direct and adjoint models to estimate the state of the system, which reproduces the observed data as accurately as possible at a given time interval in the sense of minimizing cost functional (1.15). It should be noted that problems (1.16) and (1.17)–(1.19) are solved at once over the entire time interval $(0, T)$ in the 4D-VAR method.

The EnKF method assimilates observations sequentially, unlike 4D-VAR. This method requires the ensemble of state vector x_i^f from the previous step t_{i-1} for given observations y_i^0 at time t_i . The EnKF method consists of constructing a correction for expectation (mean over the ensemble) \bar{x}_i^f according to the following formula:

$$\bar{x}_i^a = \bar{x}_i^f + \tilde{K} (y_i^0 - H(\bar{x}_i^f)), \quad (4.1)$$

where \bar{x}_i^a is the ensemble mean state estimate (analysis) at time t_i , H is the observation operator, and \tilde{K} is the generalization of the Kalman matrix (gain matrix):

$$K = P_i^f H^* (H P_i^f H^* + R)^{-1}, \quad (4.2)$$

where P_i^f is the covariance matrix of state errors at time t_i . The covariance matrices in definition (4.2) are replaced by the covariance matrices of the sample based on the ensemble to obtain \tilde{K} . Thus, the EnKF method constructs corrections for \bar{x}_i^f taking into account the uncertainties in the observational data y_i^0 . This scheme gives the state ensemble at time t_i , which later serves as the initial condition for the ensemble at time t_{i+1} .

Thus, the difference in the mentioned approaches is laid down already in very formulations (1.16) and (4.1)–(4.2): the 4D-VAR method minimizes the functional $J(x)$ at once over the entire time interval $(0, T)$, while the EnKF method assimilates the observations sequentially at each specific time. Unlike 4D-VAR, the covariance analysis matrix P_i^f plays a key role in assimilation by the EnKF method, where the estimates of these matrices are refined at each time step.

The processing of the covariance matrices in the EnKF method becomes a serious computational problem if dimensions of the state vector are large. The use of a limited number of ensemble elements leads to a deterioration in the approximation of the Kalman filter. On the other hand, in the 4D-VAR method, it is necessary to construct and solve linearized direct and

adjoint problems using iterative gradient methods, which is often a big problem for complex geophysical models. Thus, the construction of an adjoint model for the well-known NEMOVAR data assimilation system took many hours [73].

Numerical comparisons of 4D-VAR and EnKF [93–99] showed that these methods often give similar results. The EnKF provides more accurate results for small time intervals. 4D-VAR leads to smaller errors than EnKF for observations with gaps in the data, when ensemble perturbations grow nonlinearly and become non-Gaussian [94]. However, EnKF is preferable from the point of view of parallelization of computations, because computations for each member of the ensemble can be carried out independently [99].

Errors in the models describing real physical systems, such as the atmosphere and the ocean, occur due to the inaccurate forcing (of the right side or boundary conditions), the parametrization of subgrid processes, low resolution, etc. Errors can be systematic and random, as well as errors of model parameters or physical parameterizations. In this case, 4D-VAR is used in the weak formulation (weak constraint) [116] described in Section 3. On the one hand, this approach places high demands on computing systems due to the high dimension of the state vector of the system. It reaches $\sim 10^9$ in modern numerical weather forecast models. On the other hand, this approach can improve the accuracy of the forecast and increase the “window” of data assimilation by considering the extended cost functional in form (3.1).

In the EnKF method, a weak formulation appears naturally by adding errors of the model, the estimation of which is also refined at each step to the formulation of the problem of the covariance matrix [93]. This indicates the need for a further comparison of EnKF and 4D-VAR methods in a weak formulation.

A broad discussion of the comparison of 4D-VAR and EnKF [93–99] methods concluded in the recognition of the need to develop data assimilation approaches combining the best features of 4D-VAR and EnKF [103]. This is how the Hybrid 4DVar approach, which combines the ensemble Kalman filter and variational data assimilation [100–104], and the ensemble method of 4D-VAR data assimilation 4DEnVar [105–110] appeared.

CONCLUSIONS

In this paper, we reviewed and analyzed methods for solving data assimilation problems developed in recent decades. The development of data assimilation systems began with meteorology and was dictated by the need to improve weather forecasts. These methods are increasingly used in oceanography and other areas, in addition to modern complex meteorological data assimilation systems.

Qualitative changes in measurement systems occur along with the progress in solving data assimilation problems. Recent years have been marked by a continuous increase in the number of measurements of various characteristics of our geosystem. Therefore, the development of technologies for solving data assimilation problems based on modern approaches and taking into account recent advances in this direction is urgent.

The most modern and effective methods are the variational data assimilation. Thus, special attention should be paid to research in the field of a numerical solution of the problems of variational observation data assimilation for the problems of the dynamics of oceans and seas. The weak formulation of the problem of variational data assimilation makes it possible to take into account possible errors of the model and thus leads to a more accurate solution of the problem. The algorithms formulated in Sections 2 and 3 can be used to calculate the covariance matrix of the errors of the optimal solution and the individual contributions to it associated with the errors of the input data. These algorithms make it possible to investigate the sensitivity of optimal solutions of problems of variational data assimilation using the cost functional Hessian.

ACKNOWLEDGMENTS

This study was supported in part by the Russian Science Foundation (Sections 1, 2, and 4), project no. 17-77-30001, and by the Russian Foundation for Basic Research (Section 3), project no. 18-01-00267.

REFERENCES

1. L. Richardson, *Weather Prediction by Numerical Process* (Cambridge University Press, Cambridge, 1922).
2. J. G. Charney, “The use of the primitive equations of motion in numerical prediction,” *Tellus* **7**, 22–26 (1955).
3. N. A. Phillips, “On the problem of initial data for the primitive equations,” *Tellus* **12**, 121–126 (1960).
4. A. H. Jazwinski, *Stochastic Processes and Filtering Theory* (Academic, London, 1970).
5. R. S. Bucy and P. D. Joseph, *Filtering for Stochastic Processes with Applications to Guidance* (Chelsea, New York, 1987).
6. J. L. Lions, *Contrôle optimal des systèmes gouvernés par des équations aux dérivées partielles* (Dunod, Paris, 1968).
7. G. I. Marchuk, *Adjoint Equations and Analysis of Complex Systems* (Kluwer, Dordrecht, 1995).
8. P. E. Gill, W. Murray, and M. H. Wright, *Practical Optimization* (Academic, London, 1987).
9. G. I. Marchuk, V. I. Agoshkov, and V. P. Shutyaev, *Adjoint Equations and Perturbation Algorithms in Non-linear Problems* (CRC, New York, 1996).
10. A. F. Bennett, *Inverse Modeling of the Ocean and Atmosphere* (Cambridge University Press, Cambridge, 2002).

11. R. Daley, *Atmospheric Data Analysis* (Cambridge University Press, Cambridge, 1991).
12. M. Ghil and P. Malanotte-Rizzoli, "Data assimilation in meteorology and oceanography," *Adv. Geophys.* **33**, 141–266 (1991).
13. E. Kalnay, *Atmospheric Modeling. Data Assimilation and Predictability* (Cambridge University Press, Cambridge, 2003).
14. H. Panofsky, "Objective weather-map analysis," *J. Appl. Meteorol.* **6**, 386–392 (1949).
15. B. Gilchrist and G. Cressman, "An experiment in objective analysis," *Tellus* **6** (4), 309–318 (1954).
16. P. Bergthórsson and B. Döös, "Numerical weather map analysis," *Tellus* **7** (3), 329–340 (1955).
17. G. Cressman, "An operational objective analysis system," *Mon. Weather Rev.* **87**, 367–374 (1959).
18. J. Hoke and R. A. Anthes, "The initialization of numerical models by a dynamic initialization technique," *Mon. Weather Rev.* **104**, 1551–1556 (1976).
19. J. Verron, "Altimeter data assimilation into an ocean circulation model: Sensitivity to orbital parameters," *J. Geophys. Res.* **95** (C7), 443–459 (1990).
20. J. Verron and W. R. Holland, "Impact de données d'altimétrie satellitaire sur les simulations numériques des circulations générales océaniques aux latitudes moyennes," *Ann. Geophys.* **7**, 31–46 (1989).
21. E. Blayo, J. Verron, and J.-M. Molines, "Assimilation of TOPEX/POSEIDON altimeter data into a circulation model of the North Atlantic," *J. Geophys. Res.* **99** (C12), 24691–24705 (1994).
22. D. Auroux and J. Blum, "A nudging-based data assimilation method: the Back and Forth Nudging (BFN) algorithm," *Nonlinear Processes Geophys.* **15**, 305–319 (2008).
23. L. C. Gandin, *Objective Analysis of Hydrometeorological Fields* (Gidrometizdat, Leningrad, 1963) [in Russian].
24. A. C. Lorenc, "A global three-dimensional multivariate statistical analysis scheme," *Mon. Weather Rev.* **109** (4), 701–721 (1981).
25. R. D. McPherson, K. H. Bergman, R. E. Kistler, G. E. Rasch, and D. S. Gordon, "The NMC operational global data assimilation system," *Mon. Weather Rev.* **107** (11), 1445–1461 (1979).
26. W. H. Lyne, R. Swinbank, and N. T. Birch, "A data assimilation experiment, with results showing the atmospheric circulation during the FGGE special observing periods," *Q. J. R. Meteorol. Soc.* **108**, 575–594 (1982).
27. A. C. Lorenc, "Analysis methods for numerical weather prediction," *Q. J. R. Meteorol. Soc.* **112** (474), 1177–1194 (1986).
28. A. C. Lorenc, R. S. Bell, and B. Macpherson, "The Meteorological Office analysis correction data assimilation scheme," *Q. J. R. Meteorol. Soc.* **117**, 59–89 (1991).
29. H. J. Thiebaut and M. A. Pedder, *Spatial Objective Analysis* (Academic, London, 1987).
30. H. Douville, P. Viterbo, J.-F. Mahfouf, and A. C. M. Beljaars, "Evaluation of the optimum interpolation and nudging techniques for soil moisture analysis using FIFE data," *Mon. Weather Rev.* **128**, 1733–1756 (2000).
31. A. N. Bagrov and M. D. Tsirul'nikov, "Operational scheme of objective analysis of the Russian Hydrometeorological Center," in *70 Years of the Russian Hydrometeorological Center* (Gidrometeoizdat, St. Petersburg, 1999), pp. 59–69 [in Russian].
32. A. V. Frolov, A. I. Vazhnik, P. I. Svirengo, and V. I. Tsvetkov, *Global System of Data Assimilation for Atmospheric Observation Data* (Gidrometeoizdat, St. Petersburg, 2000) [in Russian].
33. J. A. Carton and E. C. Hackert, "Applications of multi-variate statistical objective analysis to the circulation in the tropical Atlantic," *Ocean. Dyn. Atmos. Oceans* **13**, 491–515 (1989).
34. J. C. Derber and A. Rosati, "A global ocean data assimilation system," *J. Phys. Oceanogr.*, **19**, 1333–1347 (1989).
35. S. Smith, J. A. Cummings, and C. Rowley, Validation Test Report for the Navy Coupled Ocean Data Assimilation 3D Variational Analysis (NCODA-VAR) System, Version 3.43 (2012).
36. G. Evensen, "The ensemble Kalman filter: Theoretical formulation and practical implementation," *Ocean Dyn.* **53** (4), 343–367 (2003).
37. P. Sakov and P. A. Sandery, "Comparison of EnOI and EnKF regional ocean reanalysis systems," *Ocean Modell.* **89**, 45–60 (2015).
38. M. N. Kaurkin, R. A. Ibrayev, and K. P. Belyaev, "ARGO data assimilation into the ocean dynamics model with high spatial resolution using Ensemble Optimal Interpolation (EnOI)," *Oceanology (Engl. Transl.)* **56** (6), 774–781 (2016).
39. R. E. Kalman, "A new approach to linear filter and prediction theory," *J. Basic Eng.* **82**, 35–45 (1960).
40. R. E. Kalman and R. S. Bucy, "New results in linear filtering and prediction theory," *J. Basic Eng.* **83D**, 95–108 (1961).
41. M. Ghil, S. E. Cohn, and A. Dalcher, "Sequential estimation, data assimilation, and initialization," in *The Interaction between Objective Analysis and Initialization (Proceedings of the Fourteenth Stanstead Seminar)*, Ed. by D. Williamson (McGill University, Montreal, 1982), pp. 83–97.
42. N. P. Budgell, "Stochastic filtering of linear shallow water wave processes," *SIAM J. Sci. Stat. Comput.* **8** (2), 152–179 (1987).
43. A. S. Sarkisyan, *Modeling of Ocean Dynamics* (Gidrometeoizdat, St. Petersburg, 1991) [in Russian].
44. B. A. Nelepo, V. V. Knysh, A. S. Sarkisyan, and I. E. Timchenko, "Study of synoptic variability of the ocean based on the dynamic–stochastic approach," *Dokl. Akad. Nauk SSSR* **246** (4), 974–978 (1979).
45. G. K. Korotaev and V. N. Ereemeev, *Introduction to Operational Oceanography of the Black Sea* (EKOSI-Gidrofizika, Sevastopol, 2006) [in Russian].
46. A. S. Sarkisyan, S. G. Demyshev, G. K. Korotaev, and V. A. Moiseenko, "An example of four-dimensional analysis of observational data from the Razrezy program for the Newfoundland energy active zone of the ocean," in *Scientific and Technological Results: Atmo-*

- sphere, Ocean, Space—the Razrezy Program* (VINITI, Moscow, 1986), vol. 6, pp. 88–89 [in Russian].
47. V. V. Knysh, G. K. Korotaev, A. I. Mizyuk, and A. S. Sarkisyan, “Assimilation of hydrological observation data for calculating currents in seas and oceans,” *Izv., Atmos. Ocean. Phys.* **48** (1), 57–73 (2012).
 48. G. Evensen, *Data Assimilation: The Ensemble Kalman Filter* (Springer, Berlin, 2007).
 49. E. Klimova, “A suboptimal data assimilation algorithm based on the ensemble Kalman filter,” *Q. J. R. Meteorol. Soc.* **138**, 2079–2085 (2012).
 50. A. V. Shlyayeva, M. A. Tolstykh, V. G. Mizyak, and V. S. Rogutov, “Local ensemble transform Kalman filter data assimilation system for the global semi-Lagrangian atmospheric model,” *Russ. J. Numer. Anal. Math. Modell.* **28** (4), 419–441 (2013).
 51. Y. K. Sasaki, “An objective analysis based on the variational method,” *J. Meteorol. Soc. Jpn.* **36**, 77–88 (1958).
 52. Y. Sasaki, “Some basic formalisms in numerical variational analysis,” *Mon. Weather Rev.* **98**, 875–883 (1970).
 53. C. Provost and R. Salmon, “A variational method for inverting hydrographic data,” *J. Mar. Res.* **44**, 1–34 (1986).
 54. V. V. Penenko and N. V. Obraztsov, “Variational method for the fields of meteorological variables,” *Meteorol. Gidrol.* **11**, 1–11 (1976).
 55. G. I. Marchuk and V. V. Penenko, “Application of optimization methods to the problem of mathematical simulation of atmospheric processes and environment,” in *Modelling and Optimization of Complex Systems: Proc. of the IFIP-TC7 Working Conf.*, Ed. by G. I. Marchuk (Springer, New York, 1978), pp. 240–252.
 56. F.-X. Le Dimet and O. Talagrand, “Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects,” *Tellus, Ser. A* **38**, 97–110 (1986).
 57. J. M. Lewis and J. C. Derber, “The use of adjoint equations to solve a variational adjustment problem with advective constraints,” *Tellus, Ser. A* **37**, 309–322 (1985).
 58. P. Courtier and O. Talagrand, “Variational assimilation of meteorological observations with the adjoint vorticity equation. II. Numerical results,” *Q. J. R. Meteorol. Soc.* **111**, 1329–1347 (1987).
 59. I. M. Navon, “A review of variational and optimization methods in meteorology,” in *Variational Methods in Geosciences*, Ed. by Y. K. Sasaki (Elsevier, New York, 1986), pp. 29–34.
 60. V. I. Agoshkov and G. I. Marchuk, “On solvability and numerical solution of data assimilation problems,” *Russ. J. Numer. Anal. Math. Modell.* **8** (1), 1–16 (1993).
 61. G. I. Marchuk and V. B. Zalesny, “A numerical technique for geophysical data assimilation problem using Pontryagin’s principle and splitting-up method,” *Russ. J. Numer. Anal. Math. Modell.* **8** (4), 311–326 (1993).
 62. G. I. Marchuk and V. P. Shutyaev, “Iteration methods for solving a data assimilation problem,” *Russ. J. Numer. Anal. Math. Modell.* **9** (3), 265–279 (1994).
 63. G. I. Marchuk and V. P. Shutyaev, “Adjoint equations and iterative algorithms in variational data assimilation problems,” *Tr. Inst. Mat. Mekh. Ural. Otd. Ross. Akad. Nauk* **17** (2), 136–150 (2011).
 64. V. I. Agoshkov, E. I. Parmuzin, V. B. Zalesny, V. P. Shutyaev, N. B. Zakharova, and A. V. Gusev, “Variational assimilation of observation data in the mathematical model of the Baltic Sea dynamics,” *Russ. J. Numer. Anal. Math. Modell.* **30** (4), 203–212 (2015).
 65. V. I. Agoshkov, M. Assovskii, V. B. Zalesny, N. B. Zakharova, E. I. Parmuzin, and V. P. Shutyaev, “Variational assimilation of observation data in the mathematical model of the Black Sea taking into account the tide-generating forces,” *Russ. J. Numer. Anal. Math. Modell.* **30** (3), 129–142 (2015).
 66. M. Ventsel’ and V. B. Zalesny, “Data assimilation in the one-dimensional model of heat convection–diffusion in the ocean,” *Izv. Akad. Nauk: Fiz. Atmos. Okeana* **32** (5), 613–629 (1996).
 67. V. P. Shutyaev, *Control Operators and Iterative Algorithms in Problems in Variational Data Assimilation Problems* (Nauka, Moscow, 2001) [in Russian].
 68. V. I. Agoshkov, E. I. Parmuzin, and V. P. Shutyaev, “Numerical algorithm for variational assimilation of sea surface temperature data,” *Comput. Math. Math. Phys.* **48** (8), 1293–1312 (2008).
 69. D. F. Parrish and J. C. Derber, “The National Meteorological Center’s spectral statistical interpolation analysis scheme,” *Mon. Weather Rev.* **120**, 1747–1763 (1992).
 70. P. Courtier, E. Andersson, W. Heckley, J. Pailleux, D. Vasiljevic, M. Hamrud, A. Hollingsworth, F. Rabier, and M. Fisher, “The ECMWF implementation of three-dimensional variational assimilation (3D-Var). I. Formulation,” *Q. J. R. Meteorol. Soc.* **124**, 1783–1807 (1998).
 71. P. Courtier, J. N. Thepaut, and A. Hollingsworth, “A strategy for operational implementation of 4D-Var, using an incremental approach,” *Q. J. R. Meteorol. Soc.* **120**, 1389–1408 (1994).
 72. K. Ide, P. Courtier, M. Ghil, and A. C. Lorenc, “Unified notation for data assimilation: Operational, sequential and variational,” *J. Meteorol. Soc. Jpn.* **75**, 181–189 (1997).
 73. K. Mogensen, M. A. Balmaseda, A. T. Weaver, M. Martin, and A. Vidard, “NEMOVAR: a variational data assimilation system for the NEMO ocean model,” *ECMWF Tech. Mem.*, No. 120 (2009).
 74. Yu. G. Evtushenko, E. S. Zasukhina, and V. I. Zubov, “Numerical optimization of solutions to Burgers problem by means of boundary conditions,” *Comput. Math. Math. Phys.* **37** (12), 1406–1414 (1997).
 75. V. M. Ipatova, *Data Assimilation for an Ocean General Circulation Model in the quasi-geostrophic approximation* (VINITI, Moscow, 1992), No. 2333-V92 [in Russian].
 76. V. I. Agoshkov and V. M. Ipatova, “Solvability of the observation data assimilation problem in the three-

- dimensional model of ocean dynamics,” *Differ. Equations* **43** (8), 1088–1100 (2007).
77. V. I. Agoshkov and V. M. Ipatova, “Existence theorems for a three-dimensional ocean dynamics model and a data assimilation problem,” *Dokl. Math.* **75** (1), 28–30 (2007).
 78. Z. Sirkes and E. Tziperman, “Finite difference of adjoint or adjoint of finite difference?,” *Mon. Weather Rev.* **125**, 3373–3378 (1997).
 79. E. I. Parmuzin, V. P. Shutyaev, and N. A. Diansky, “Numerical solution of a variational data assimilation problem for a 3D ocean thermohydrodynamics model with a nonlinear vertical heat exchange,” *Russ. J. Numer. Anal. Math. Modell.* **22** (2), 177–198 (2007).
 80. F.-X. Le Dimet and I. Charpentier, “Méthodes de second ordre en assimilation de données,” in *Équations aux Dérivées Partielles et Applications (Articles dédiés à Jacques-Louis Lions)* (Elsevier, Paris, 1998), pp. 623–639.
 81. A. S. Lawless, N. K. Nichols, and S. P. Balloid, “A comparison of two methods for developing the linearization of a shallow-water model,” *Q. J. R. Meteorol. Soc.* **129**, 1237–1254 (2003).
 82. W. C. Chao and L. -P. Chang, “Development of a four-dimensional variational analysis system using the adjoint method at GLA Part I. Dynamics,” *Mon. Weather Rev.* **120**, 1661–1672 (1992).
 83. R. Giering and T. Kaminski, “Recipes for adjoint code constructions,” *ACM Trans. Math. Software* **24**, 437–474 (1998).
 84. M. B. Giles and N. A. Pierce, “An introduction to the adjoint approach to design,” *Flow, Turbul. Combust.* **65**, 393–415 (2000).
 85. F.-X. Le Dimet and V. P. Shutyaev, “On deterministic error analysis in variational data assimilation,” *Nonlinear Processes Geophys.* **12**, 481–490 (2005).
 86. I. Gejadze, F.-X. Le Dimet, and V. Shutyaev, “On analysis error covariances in variational data assimilation,” *SIAM J. Sci. Comput.* **30** (4), 1847–1874 (2008).
 87. V. P. Shutyaev and E. I. Parmuzin, “Some algorithms for studying solution sensitivity in the problem of variational assimilation of observation data for a model of ocean thermodynamics,” *Russ. J. Numer. Anal. Math. Modell.* **24** (2), 145–160 (2009).
 88. V. I. Agoshkov, E. I. Parmuzin, and V. P. Shutyaev, “Observational data assimilation in the problem of Black Sea circulation and sensitivity analysis of its solution,” *Izv., Atmos. Ocean. Phys.* **49** (6), 592–602 (2013).
 89. V. P. Shutyaev, F.-X. Le Dimet, V. I. Agoshkov, and E. I. Parmuzin, “Sensitivity of functionals in problems of variational assimilation of observational data,” *Izv., Atmos. Ocean. Phys.* **51** (3), 342–350 (2015).
 90. V. P. Shutyaev and E. I. Parmuzin, “Studying the sensitivity of the optimal solution of the variational data assimilation problem for the Baltic Sea thermodynamics model,” *Russ. Meteorol. Hydrol.* **40** (6), 411–419 (2015).
 91. F.-X. Le Dimet, V. Shutyaev, and E. I. Parmuzin, “Sensitivity of functionals with respect to observations in variational data assimilation,” *Russ. J. Numer. Anal. Math. Modell.* **31** (2), 81–91 (2016).
 92. V. B. Zalesny, V. I. Agoshkov, V. P. Shutyaev, F. Le Dimet, and V. O. Ivchenko, “Numerical modeling of ocean hydrodynamics with variational assimilation of observational data,” *Izv., Atmos. Ocean. Phys.* **52** (4), 431–442 (2016).
 93. A. Lorenc, “The potential of the ensemble Kalman filter for NWP—a comparison with 4D-Var,” *Q. J. R. Meteorol. Soc.* **129**, 3183–3203 (2003).
 94. E. Kalnay, H. Li, T. Miyoshi, S.-C. Yang, J. Balabrera-Poy, “4D-Var or ensemble Kalman filter?,” *Tellus, Ser. A* **59**, 758–773 (2007).
 95. A. Caya, J. Sun, and C. Snyder, “A Comparison between the 4DVAR and the ensemble Kalman filter techniques for radar data assimilation,” *Mon. Weather Rev.* **133** (11), 3081–3094 (2005).
 96. N. Gustafsson, “Discussion on “4D-Var Or EnKF”,” *Tellus, Ser. A* **59**, 774–777 (2007).
 97. E. J. Fertig, J. Harlim, and B. R. Hunt, “A comparative study of 4D-VAR and a 4D ensemble Kalman filter: Perfect model simulations with Lorenz-96,” *Tellus, Ser. A* **59**, 96–100 (2007).
 98. M. Buehner, P. Houtekamer, C. Charette, H. Mitchell, and B. He, “Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part I. Description and single-observation experiments,” *Mon. Weather Rev.* **138**, 1550–1566 (2010).
 99. D. Fairbairn, S. R. Pring, A. C. Lorenc, and I. Roulstone, “A comparison of 4DVar with ensemble data assimilation methods,” *Q. J. R. Meteorol. Soc.* **140**, 281–294 (2014).
 100. X. Tian, J. Xie, and A. Dai, “An ensemble-based explicit 4D-Var assimilation method,” *J. Geophys. Res.* **113**, D21124 (2008).
 101. F. Q. Zhang, M. Zhang, and J. A. Hansen, “Coupling ensemble Kalman filter with four dimensional variational data assimilation,” *Adv. Atmos. Sci.* **26** (1), 1–8 (2009).
 102. A. M. Clayton, A. C. Lorenc, and D. M. Barker, “Operational implementation of a hybrid ensemble/4D-Var global data assimilation at the Met Office,” *Q. J. R. Meteorol. Soc.* **139**, 1445–1461 (2013).
 103. N. Gustafsson, J. Bojarova, and O. Vignes, “A hybrid variational ensemble data assimilation for the High Resolution Limited Area Model (HIRLAM),” *Nonlinear Processes Geophys.* **21**, 303–323 (2014).
 104. M. Bonavita, E. Hólm, L. Isaksen, and M. Fisher, “The evolution of the ECMWF hybrid data assimilation system,” *Q. J. R. Meteorol. Soc.* **142**, 287–303 (2016).
 105. C. Liu, Q. Xiao, and B. Wang, “An ensemble-based four-dimensional variational data assimilation scheme. Part I: Technical formulation and preliminary test,” *Mon. Weather Rev.* **136**, 3363–3373 (2008).
 106. C. Liu, Q. Xiao, and B. Wang, “An ensemble-based four-dimensional variational data assimilation scheme. Part II: Observing system simulation experiments with Advanced Research WRF (ARW),” *Mon. Weather Rev.* **137**, 1687–1704 (2009).
 107. C. Liu and Q. Xiao, “An ensemble-based four-dimensional variational data assimilation scheme. Part III:

- Antarctic applications with advanced WRF using real data," *Mon. Weather Rev.* **141**, 2721–2739 (2013).
108. G. Desroziers, J.-T. Camino, and L. Berre, "4D-EnVar: Link with 4D state formulation of variational assimilation and different possible implementations," *Q. J. R. Meteorol. Soc.* **140**, 2097–2110 (2014).
109. N. Gustafsson and J. Bojarova, "Four-dimensional ensemble variational (4D-En-Var) data assimilation for the HIgh Resolution Limited Area Model (HIR-LAM)," *Nonlinear Processes Geophys.* **21**, 745–762 (2014).
110. M. Asch, M. Bocquet, and M. Nodet, *Data Assimilation: Methods, Algorithms, and Applications* (SIAM, Philadelphia, 2016).
111. I. Gejadze, F.-X. Le Dimet, and V. P. Shutyaev, "On optimal solution error covariances in variational data assimilation problems," *J. Comput. Phys.* **229**, 2159–2178 (2010).
112. I. Gejadze, V. P. Shutyaev, and F.-X. Le Dimet, "Analysis error covariance versus posterior covariance in variational data assimilation," *Q. J. R. Meteorol. Soc.* **139**, 1826–1841 (2013).
113. I. Yu. Gejadze and V. P. Shutyaev, "On Gauss-verifiability of optimal solutions in variational data assimilation problems with nonlinear dynamics," *J. Comput. Phys.* **280**, 439–456 (2015).
114. D. C. Liu and J. Nocedal, "On the limited memory BFGS Method for large scale minimization," *Math. Programming* **45** (1–3), 503–528 (1989).
115. F. Veerse, D. Auroux, and M. Fisher, "Limited-memory BFGS diagonal pre-conditioners for a data assimilation problem in meteorology," *Optim. Eng.* **1**, 323–339 (2000).
116. Y. Trémolet, "Model-error estimation in 4D-Var," *Q. J. R. Meteorol. Soc.* **133** (626), 1267–1280 (2007).
117. A. Carrassi and S. Vannitsem, "Accounting for model error in variational data assimilation: A deterministic approach," *Mon. Weather Rev.* **138**, 875–883 (2010).
118. D. Furbish, M. Y. Hussaini, F.-X. Le Dimet, et al., "On discretization error and its control in variational data assimilation," *Tellus, Ser. A* **60**, 979–991 (2008).
119. A. K. Griffith and N. K. Nichols, "Adjoint methods in data assimilation for estimating model error," *Flow, Turbul. Combust.* **65** (3–4), 469–488 (2000).
120. A. Vidard, A. Piacentini, and F.-X. Le Dimet, "Variational data analysis with control of the forecast bias," *Tellus, Ser. A* **56**, 1–12 (2004).
121. S. Akella and I. Navon, "Different approaches to model error formulation in 4D-Var: A study with high resolution advection schemes," *Tellus, Ser. A* **61**, 112–128 (2009).
122. V. V. Penenko, "Variational methods of data assimilation and inverse problems for studying the atmosphere, ocean, and environment," *Numer. Anal. Appl.* **2** (4), 341–351 (2009).
123. M. D. Tsyrlunikov, "Stochastic modelling of model errors: A simulation study," *Q. J. R. Meteorol. Soc.* **131**, 3345–3371 (2005).
124. I. Gejadze, H. Oubanas, and V. Shutyaev, "Implicit treatment of model error using inflated observation-error covariance," *Q. J. R. Meteorol. Soc.* **143**, 2496–2508 (2017).
125. V. Shutyaev, I. Gejadze, A. Vidard, and F.-X. Le Dimet, "Optimal solution error quantification in variational data assimilation involving imperfect models," *Int. J. Numer. Methods Fluids* **83** (3), 276–290 (2017).
126. A. M. Stuart, "Inverse problems: A Bayesian perspective," *Acta Numerica* **9**, 451–559 (2010).

Translated by O. Pismenov