Journal of Mathematical Sciences, Vol. 99, No. 4, 2000

APPLICATION OF THE FOKKER–PLANCK EQUATION TO DATA ASSIMILATION INTO HYDRODYNAMICAL MODELS

K. Belyaev (Moscow, Russia), **S. D. Meyers, and J. J. O'Brien** (Florida, USA) UDC 519.2

A numerical ocean model with an original data assimilation technique is considered. The data-assimilation approach is based on the well-known Kalman filter theory, but the method of computation of the covariation matrix is new. This method uses a diffusion stochastic representation of the error (difference between the model and observative values), and then the Fokker–Planck equation is solved for determination of the joint distribution of errors in each of two different space points. Compared to the ordinary Kalman-filter technique, which requires n^2 operations, where n is the number of grid points, this method requires only m^2 operations, where m is the number of observations. Also, the method does not require linearity of the model. This method is considered in conjunction with a thermodynamic ocean model based on the primitive equations. Some model experiments have been carried out. The stability of this technique is examined, and possible applications to other models is also discussed.

1. Introduction

During the past several years there has been an increase in the development of data-assimilation methods and their application. The first data-assimilation applications were in meteorology, where data assimilation is now an important component of numerical weather forecasting. The data revolution in oceanography is bringing the daily practice of physical oceanography closer to that of dynamic meteorology. There now exist large observational data sets of temperature and salinity in the North Atlantic (from global projects such as WOCE and SECTIONS), in the tropical ocean (TOGA, COARE), and elsewhere.

Improved assimilation techniques are needed in order for ocean models to fully exploit the new observation networks. Data assimilation improves the estimate of a state by extracting a maximum amount of information from both the measurements and the dynamic model, combining them in an optimal way. Assimilation may be used to improve initial and boundary conditions and to estimate badly known model parameters.

Data assimilation in ocean models has been ongoing in the scientific literature for approximately 30 yr (for a review, see [9, 16]). It is necessary to understand which processes can be described by the oceanographic model and what role measurements may play in the model. Two extreme opinions exist. One considers the ocean as a quasi-stationary medium and believes that it is sufficient to make measurements only once in order to have a good description of the ocean state. From this point of view, the numerical models are useless (or simply wrong) and data-assimilation techniques reduce to a statistical correction process. The second opinion is that the ocean is a highly turbulent fluid that does not have a memory about its previous state. In this case, information generated by measurements becomes meaningless in a short time interval, and it is necessary to continuously monitor the ocean. Thus, only well-designed resolution models are able to describe the actual processes in the ocean, and data assimilation is useful for obtaining correct initial (and possible boundary) conditions, needed for prescribing the development of the ocean. Reality, however, lies somewhere between these two extreme points of view. Thus, the scientific question is how to represent this situation numerically. The answer depends on both the model and the data-assimilation techniques as well as the local dynamic.

There are two general concepts that have been discussed for data assimilation. The first, the "variational/ad-joint" method, has been the most popular scheme (e.g., [23–25, 27, 29]). An example of this technique is to assume the initial conditions to be unknown parameters for the model, which, given a set of measurements distributed in some time interval, result in a model trajectory that best fits the measurements in some optimal sense. This can be formulated as a constrained minimization problem where the ocean model operates as a strong constraint of a cost function measuring the difference between the model solution and the measurements.

The other class of methods is "sequential data assimilation." Starting from some initial condition, the model solution is sequentially updated at every time step where measurements are available. The model solution will approach the observed state under certain conditions. This group of methods requires an updating scheme that combines the

Proceedings of the Seminar on Stability Problems for Stochastic Models, Vologda, Russia, 1998, Part II.

1072-3374/00/9904-1393$25.00

model solution and the measurements to find the "best" state estimate. The Kalman-filter method belongs to this class of data-assimilation technique.

The Kalman filter is derived in a number of books on control theory (e.g., [2, 11]). In oceanography, the Kalman filter has been used in [4, 8, 19, 26] and by other authors. A detailed review devoted to the Kalman-filter application is presented in [9]. The main idea of this method is to solve dynamic equations using the equation for the error covariance matrix, where the error is the distance between the model and observation values. It faces serious problems because the equation requires large computations and assumes, moreover, a linear operator for the dynamic system. There have been some attempts to extend the Kalman-filter theory to the nonlinear system operator, but this requires additional simplifications. These have been presented in [10, 19] and are now commonly used.

In this paper, another method for the definition of the covariance matrix of error is presented. This idea is based on the stochastic-process theory and partial parabolic differential equations.

The application of the theory of stochastic processes has been used and is well known in turbulence theory (see [14]) and in climate models with stochastic forcing (see [20]) but has not been dealt with in data assimilation. The method is relatively simple and does not require large computations, but nevertheless reflects the basic features of the physical interaction in the region of interest. The main idea is to consider the time variability in phase space, while avoiding the necessity of following each spatial point. The diffusion approximation is used for describing the time variability of the error, and then the Fokker–Planck equation is solved in order to obtain the joint probability distribution for each pair of measurements. Unlike the ordinary Kalman-filter method that requires $O(n^2)$ operations, where n is the number of grid points, this method requires only $O(m^2)$ operations, where m is the number of measurements. The method does not depend on the linearity of the model. Naturally, it has its own weaknesses, and a discussion about the advantages and disadvantages compared with ordinary data-assimilation methods is presented.

This method is applied to the primitive-equation hydrodynamic model of Sarkisyan (see [18]). The goal is to demonstrate the usefulness of the assimilation technique. This model is used only as a tool for calculation of thermodynamic fields and their prediction and variability. This model has had wide applications, and many papers have discussed the model and its modifications (e.g., [6]). A regional version of this model was considered by Knysh (see [13]) in the tropical Atlantic and by Moiseenko and Chernov (see [12]) in the North Atlantic. Also, Knysh (see [13]) considered this model with data assimilation based on linear Kalman-filter theory.

The next section presents the ocean model and modifications to the model based on a Rossby number expansion. The assimilation technique is discussed in Sec. 3, and the application results are presented in Sec. 4. The last section is a summary and conclusion.

2. The Model

The following equations describe the thermodynamics in the ocean model:

$$\frac{\partial u}{\partial t}+u\frac{\partial u}{\partial x}+v\frac{\partial u}{\partial y}+w\frac{\partial u}{\partial z}-fv=-\frac{1}{\rho_0}\frac{\partial p}{\partial x}+\mu\Delta u+\nu\frac{\partial^2 u}{\partial z^2}, \tag{1.1}$$

$$\frac{\partial v}{\partial t}+u\frac{\partial v}{\partial x}+v\frac{\partial v}{\partial y}+w\frac{\partial v}{\partial z}+fu=-\frac{1}{\rho_0}\frac{\partial p}{\partial y}+\mu\Delta v+\nu\frac{\partial^2 v}{\partial z^2}, \tag{1.2}$$

$$\frac{\partial p}{\partial z}=\rho g, \tag{1.3}$$

$$\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}+\frac{\partial w}{\partial z}=0, \tag{1.4}$$

$$\frac{\partial T}{\partial t}+u\frac{\partial T}{\partial x}+v\frac{\partial T}{\partial y}+w\frac{\partial T}{\partial z}=\mu_T\Delta T+\nu_T\frac{\partial^2 T}{\partial z^2}, \tag{1.5}$$

$$\frac{\partial S}{\partial t}+u\frac{\partial S}{\partial x}+v\frac{\partial S}{\partial y}+w\frac{\partial S}{\partial z}=\mu_S\Delta S+\nu_S\frac{\partial^2 S}{\partial z^2} \tag{1.6}$$

in a Cartesian coordinate system, where x, y, z are axes positive to the east, north, and with depth, respectively; P is the pressure; T and S are temperature and salinity. The viscosity coefficients μ, ν, μ_T, μ_S, ν_T, and ν_S are assumed to be constant in space and time. Further, f is the Coriolis parameter, and the β-plane approximation is used, $f=f_0+\beta y$. The velocity components are u, v, ν, and w; ρ is the density; Δ is the horizontal Laplacian.

The UNESCO equation of state (UNESCO 1981) is used to close the system (1.1)–(1.6):

$$\rho=\rho(T,S). \tag{1.7}$$

This model is applied to a limited area, so boundary conditions are necessary on the lateral boundaries, the sea surface and the bottom. On the equilibrium sea-surface we have

$$\nu\frac{\partial u}{\partial z} = -\tau_x, \qquad \nu\frac{\partial v}{\partial z} = -\tau_y, \qquad w = 0, \qquad z = 0.$$

On the bottom, when $z = H(x,y)$, where H is the bottom topography, we have

$$\frac{\partial u}{\partial z} = \frac{\partial v}{\partial z} = 0, \qquad w = 0.$$

Here τ_x and τ_y are the wind stress components.

There are two kinds of boundary conditions for temperature considered at the sea surface. First, the temperature can be kept constant (in time) and, second, temperature can be prescribed by the relation $\nu_T(\partial T/\partial z) = F(T)$, where $F(T)$ is the heat flux. For temperature and salinity, only the first type of boundary condition is presented. On lateral boundaries, the first type of boundary condition is used, with

$$T = T_0(x,y,z), \qquad S = S_0(x,y,z), \qquad (u,v) = (u_0(x,y,z), v_0(x,y,z)),$$

respectively. Here u_0 and v_0 are the geostrophic velocities.

The model is initialized with seasonal fields that have been taken from the Levitus atlas (see [15]).

Following Marchuk and Sarkisyan (see [18]), this system can be rewritten in the following form:

$$-\mu\Delta^2\xi + \frac{\partial}{\partial t}\Delta\xi + \frac{g}{f}J(\xi,\Delta\xi) + \frac{1}{2\gamma H}\Delta\xi + \frac{f}{H}J(H,\xi) + \beta\frac{\partial\xi}{\partial x} = \frac{f}{\rho_0 g H}\operatorname{rot}\tau + \frac{f}{\rho_0 g H}\left(\beta\tau_x + \frac{\partial}{\partial t}\operatorname{div}\tau\right) + \overline{f}, \qquad (2.1)$$

where ξ is the sea level, J is the Jacobian, $\gamma = \sqrt{f/2\nu}$ is the inverse of the Ekman layer thickness, and the function $\overline{f}$ is shown in the Appendix.

Equations (1.1)–(1.3) are rewritten in the following form:

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} + w\frac{\partial u}{\partial z} - fv = -g\frac{\partial \xi'}{\partial x} - \frac{g}{\rho_0}\int_0^z \frac{\partial \rho}{\partial x}\,dz + \mu\Delta u + \nu\frac{\partial^2 u}{\partial z^2}, \qquad (2.2)$$

$$\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + w\frac{\partial v}{\partial z} + fu = -g\frac{\partial \xi'}{\partial y} - \frac{g}{\rho_0}\int_0^z \frac{\partial \rho}{\partial y}\,dz + \mu\Delta v + \nu\frac{\partial^2 v}{\partial z^2}, \qquad (2.3)$$

where $\xi' = \xi - \rho_0^{-1}\int_0^H \rho\,dz$ is called the adynamic sea level. All other equations of system (1) are unchanged. Equation (2.1) requires boundary and initial conditions for ξ or for ξ'. We assume ξ' equal to 0 on all boundaries and at all points where t is equal to 0.

This system is considered on a grid with spatial resolution $0.25° \times 0.25°$ and with 30 standard levels in depth. The research area is in the mid-latitudes and the Rossby parameter $\varepsilon = U_0/fL$ is small ($\varepsilon \approx 0.1$) ($U_0 = 10$ cm/sec and $L = 10^6$ cm are the typical scales of velocity and space resolution, respectively.) System (2) is expanded with respect to ε:

$$u = \sum_{i=0}^{\infty} u_i\varepsilon^i, \qquad v = \sum_{i=0}^{\infty} v_i\varepsilon^i,$$

which reduces (2.2)–(2.3) to the sequence of linear equations

$$\frac{\partial u_i}{\partial t} - fv_i = \nu\frac{\partial^2 u_i}{\partial z^2} + F_i, \qquad (2.4)$$

$$\frac{\partial v_i}{\partial t} - fu_i = \nu\frac{\partial^2 v_i}{\partial z^2} + G_i, \qquad (2.5)$$

where the functions F_i and G_i depend on $u_0,\ldots,u_{i-1}$ and $v_0,\ldots,v_{i-1}$. Their detailed form is shown in the Appendix. The calculation of the expansion is stopped when $|u_i + v_i|\varepsilon^i \le 10^{-2}$. In practice, we considered only the second order of decomposition ($i \le 2$).

The Δ^2 term in (2.1) is omitted because it is small as compared to the other terms.

An explicit time-difference scheme is used in (2.4)–(2.5) with a time step equal to 2 h. Equation (2.1) is solved using the iterative Peaceman–Rachford scheme (see [22]). Equations (1.5)–(1.6) are solved using standard numerical methods for parabolic equations (see [17]). In order to ensure the stability of the calculations, the method of simple iteration had been used for the system (2).

3. Data-Assimilation Technique

The approach presented below is new, so it will be described in detail. It is based on the well-known Kalman-filter theory, where the optimal linear estimation of the unknown field is determined in conjunction with an evolution equation for the covariance matrix of error.

Those model equations are integrated in time from $t_0 = 0$ to t_f. In our experiments, t_f equals 100 days. A number of observations were made for this period, for instance, temperature and/or salinity. The problem was then to construct new temperature and salinity fields so that they satisfy the following requirements:

1. The ensemble average of the new fields must be a solution of (2) for any time step.
2. These constructed fields have to minimize the difference with respect to the measurements, i.e., to be as close to the observations as possible.

The following notations must be introduced. Let τ and t be two arbitrary time moments, $\tau < t$, and $k(t)$ be the number of observations at moment t. Let $\overline{x}$ denote a spatial point, $\overline{x} = (x, y, z)$, and $\overline{x}_j$ be the spatial point of observation. All these points are assumed to belong to the model domain but generally not to coincide with those of the grid. Then $\xi_m(t, \overline{x})$ and $\xi_0(t, \overline{x})$ denote model and observed values, respectively. Let $\theta = \theta(t, \overline{x}) = \xi_m(t, \overline{x}) - \xi_0(t, \overline{x})$ be an error. It is reasonable to assume that θ is a random field with $\mathbf{E}\theta = 0$ for any t. ($\mathbf{E}$ is the symbol of mathematical expectation.) This suggestion means that the average of all errors at any arbitrary moment equals zero, but certainly at any separate spatial point θ it does not necessarily equal zero.

The optimal linear estimation of the unknown field $\xi(t, \overline{x})$, satisfying requirements 1 and 2, is given by the expressions (see [30])

$$\xi(t, \overline{x}) = \xi_m(t, \overline{x}) + \int_0^t \sum_{j=1}^{k(\tau)} \alpha_j(\tau, \overline{x}, \overline{x}_j)\theta(\tau, \overline{x}_j)\, d\tau, \tag{3.1}$$

$$K(t, \overline{x}, \overline{x}_i) = \int_0^t \sum_{j=1}^{k(\tau)} \alpha_j(\tau, \overline{x}, \overline{x}_j) K(t - \tau, \overline{x}_i, \overline{x}_j)\, d\tau, \tag{3.2}$$

where K is the covariance kernel of θ, i.e.,

$$K(t, \overline{x}_1, \overline{x}_2) = \mathbf{E}\theta(t, \overline{x}_1)\theta(t, \overline{x}_2). \tag{3.3}$$

The optimal (unknown) weight coefficients are $\alpha_j(t, \overline{x}, \overline{x}_j)$. For brevity, scalar values are considered. In the general case, θ is a vector, α is a matrix, and (3.1) may be rewritten in matrix form.

In order to construct the new fields ξ, it is necessary to solve (3.2) for the relatively unknown functions α_j (Wiener–Hopf equation), and then insert the solutions into (3.1).

Note that the right-hand side of (3.2) contains the covariances only between the measurement points. Therefore, in order to seek a solution, the matrix of dimensionality proportional to the number of observations needs to be inverted.

The central problem of this theory is to determine the covariance matrix of error $K(t, \overline{x}_i, \overline{x}_j)$ between two arbitrary points of measurement and, between each grid point and point of observation with respect to (3.2). Function K cannot be defined directly from the data because of an insufficiency in and an irregular network of observations. In the standard theory, the equations for K obtained from system (3.2) are very complicated and cannot be solved without some restrictions and simplifications (see [9, 10]). Without going into detail, there are at least three main problems that do not allow for the use of the ordinary method. First, it is necessary to set up the initial fields for the error covariance matrix. Second, it can only be solved currently for a linear system, as a nonlinear extension requires additional major assumptions. And third, the numerical solution requires a large number (proportional to the square of the number of grid points) of computations. Thus, a standard approach faces serious problems for realization. The method described below is free of these difficulties. Its own weaknesses will be discussed in Sec. 5.

According to the definition,

$$K(t, \overline{x}_1, \overline{x}_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sup(t, s, u)\, ds\, du, \tag{3.4}$$

where $s = \theta(t, \overline{x}_1)$ and $u = \theta(t, \overline{x}_2)$. A pair of measurement points is $(\overline{x}_1, \overline{x}_2)$, and the joint distribution density of error at this pair is $p(t, s, u)$. This distribution is considered in a phase of space that represents a two-dimensional unlimited plane for each pair of observations.

Some assumptions about the time variability of the stochastic process $\theta(t, \cdot)$ have to be made. Suppose it is governed by the equation

$$\theta(t + dt, \cdot) = \Gamma\theta(t, \cdot) + \psi, \tag{3.5}$$

where Γ represents the model operator, i.e., system (3.2), and ψ is a Gaussian random variable with zero average and variance $\sigma^2(t)$, depending on time. In other words, the error is assumed to be satisfied by the same equations as the temperature and salinity plus random forcing. Physically this model reflects the situation where the next time-step error from two components is added; the first one is the result of the model acting on the previous error (error of boundary conditions, spatial resolution, and so on), and the second is a Gaussian random variable independent of previous errors. With respect to this model, it is possible to rewrite (5) in the form of a Langevein equation (stochastic diffusional equation) for each pair of points considered:

$$d\overline{\theta} = \overline{a}(t, \overline{\theta})\, dt + B^2(t, \overline{\theta})\, dW.$$

Denoted here and further:

$$\overline{\theta}(t) = (\theta(t, x_1), \theta(t, x_2)), \qquad \overline{a}(t, \overline{\theta}) = (a_1(t, \overline{\theta}), a_2(t, \overline{\theta})), \qquad \overline{s} = (s_1, s_2),$$

$$B^2(t, \overline{\theta}) = \begin{pmatrix} b_{11}^2(t, \overline{\theta}) & b_{12}(t, \overline{\theta}) \\ b_{21}(t, \overline{\theta}) & b_{22}^2(t, \overline{\theta}) \end{pmatrix}.$$

In accord with the stochastic-process theory (see [7]), the drift vector $\overline{a}$ and diffusion matrix B^2 are defined with the formulas

$$a(t, \overline{s}) = \frac{\partial}{\partial t}\mathbf{E}(\overline{\theta}(t + dt) - \overline{\theta}(t) \mid \overline{\theta}(t) = \overline{s}), \tag{3.6}$$

$$B^2(t, \overline{s}) = \frac{\partial}{\partial t}\mathbf{E}[(\overline{\theta}(t + dt) - \overline{\theta}(t))(\overline{\theta}(t + dt) - \overline{\theta}(t))^T \mid \overline{\theta}(t) = \overline{s}], \tag{3.7}$$

where the notations $\mathbf{E}(y \mid x)$, $p(y \mid x)$ mean the conditional average and conditional probability of a random value y with condition x, respectively (see [7]), and $(\cdot)^T$ means a transpose.

The mathematical proof of the transition to the stochastic differential equation requires several additional assumptions. But all of them are not essential and this point is omitted.

Using (3.5) for the drift vector $\overline{a}$, we have

$$a(t, \overline{s}) = \frac{\partial}{\partial t}\mathbf{E}(\overline{\theta}(t + dt) - \overline{\theta}(t) \mid \overline{\theta}(t) = \overline{s}) = \frac{\partial}{\partial t}\mathbf{E}(\Gamma\overline{\theta}(t) + \psi - \overline{s} \mid -\overline{\theta}(t) = \overline{s}) = \frac{\partial}{\partial t}\int_{-\infty}^{\infty} (\overline{u} - \overline{s}) p(\overline{u} \mid \overline{s})\, d\overline{u}.$$

Obviously, $\overline{u}$ can be chosen the same as $\overline{\xi}_m(t + dt) - C$, where $\overline{\xi}_m(t + dt)$ is a pair of model values and C is its spatial average at this moment, because $\mathbf{E}\overline{\theta}$ should be equal to zero.

The same is true for the variance or diffusion matrix:

$$B^2(t, \overline{s}) = \frac{\partial}{\partial t}\mathbf{E}(\Gamma\overline{\theta}(t) + \psi - \overline{s})(\Gamma\overline{\theta}(t) + \psi - \overline{s})^T \mid \overline{\theta}(t) = \overline{s}) = \frac{\partial}{\partial t}\int_{-\infty}^{\infty} (\overline{u} - \overline{s})(\overline{u} - \overline{s})^T p(\overline{u} \mid \overline{s})\, d\overline{u} + \mathbf{E}\psi\psi^T.$$

It is natural to set

$$\mathbf{E}\psi\psi^T = \sigma^2 I,$$

where $\sigma^2 = (k(t) - 1)^{-1} \sum_{j=1}^{k(t)} \theta_j^2$ and I is the unity matrix of order 2.

For clarity it is necessary to discuss the meaning of the expression $p(\overline{u} \mid \overline{s})$, because ξ_m is not assumed to be random. The description of the procedure to obtain this value follows.

Two fields at moments τ and t are taken. All grid points of the field at moment τ where $\xi_m(x) = s$ are marked. Let those points be n_s. Then among those and only those points, grid points of $\xi_m(t)$ where $\xi_m(t) = u$ are sought. Let those points be n_u. Then the conditional probability $p(u \mid s)$ will be equal to n_u/n_s.

Note that this procedure does not depend on the linearity or nonlinearity of the model. Only the results of modeling ignoring the process of their computation are dealt with.

The required number of operations to realize this algorithm appears to be large. Actually, it is easy to see that the conditional probability $p(u \mid s)$ equals zero only if the distance between values u and s is small. This strongly restricts the necessary amount of operations, which then becomes proportional to the number of grid points.

After determination of the drift vector $\overline{a}$ and diffusion matrix B^2 the probability distribution can be found and hence the covariance matrix with respect to (3.4). This can be done by using the Fokker–Planck equation

$$\frac{\partial p(t,\overline{s})}{\partial t} = -\frac{\partial(\overline{a}p(t,\overline{s}))}{\partial \overline{s}} + \frac{1}{2}\frac{\partial^2}{\partial \overline{s}^2}(B^2 p(t,\overline{s})) \tag{3.8}$$

or, in expanded form,

$$\frac{\partial(\overline{a}p(t,\overline{s}))}{\partial \overline{s}} = \frac{\partial(a_1 p(t,\overline{s}))}{\partial s_1} + \frac{\partial(a_2 p(t,\overline{s}))}{\partial s_2},$$

$$\frac{\partial^2}{\partial s^2}(B^2 p(t,\overline{s})) = \frac{\partial^2}{\partial s_1^2}\left(b_{11}^2 p(t,\overline{s}) + 2\frac{\partial^2}{\partial s_1 \partial s_2}(b_{12}p(t,\overline{s}))\right) + \frac{\partial^2}{\partial s_2^2}(b_{22}^2 p(t,\overline{s})).$$

This equation is solved with the initial and boundary conditions that the probability vanishes as $\overline{s}$ goes to infinity. These are the Sommerfeld boundary conditions, $p(t,\pm\infty) = 0$. Let us say a few words about the initial values. Consider three cases:

1. At the initial time moment t_0, suppose we have made two measurements and the observed error was $\overline{s}_0 = (s_1^\circ, s_2^\circ)$. This means that the probability density equals the Dirac δ-function $\delta(\overline{s} - \overline{s}_0)$ at this moment. Changing the time scale, t_0 can be taken to be equal to zero. So the Fokker–Planck (FP) equation is solved with Sommerfeld boundary conditions and initial condition $p(0,\overline{s}) = \delta(\overline{s} - \overline{s}_0)$.

2. Two measurements have been made at two different time moments t_1 and t_2, and for clarity suppose that $t_1 < t_2$. Then as at previous point t_1 is assumed to be zero and the one-dimensional FP equation is solved up to the moment $t = t_2 - t_1$ with the initial condition $p(0,s) = \delta(s - s_0)$ if the error at time moment t_1 was s_0. Let its solution at moment t be $p(t,s)$. Then starting from the moment t, a two-dimensional FP equation is solved with the initial function $p(t,s)\delta(s - s_1)$, where s_1 is an error at the measurement point at moment t.

There is no problem solving this equation numerically. It is simply a two-dimensional equation with fixed boundaries and it can be solved by using many developed methods. Here, the Peaceman–Rachford scheme (see [22]) was used. Moreover, the first term in (3.7) for diffusion can be neglected as compared to the second one. This strongly simplifies (3.8).

Finally, in order to use (3.2), we have to find the covariance $K(\overline{x},\overline{x}_j)$ between the arbitrary grid point $\overline{x}$ and the points of measurements $\overline{x}_j$. This is done by using the Taylor expansion, $K(\overline{x},\overline{x}_j) = K(\overline{x}_i,\overline{x}_j) + K_x(\overline{x}_i,\overline{x}_j)(\overline{x} - \overline{x}_j) + o(|\overline{x} - \overline{x}_j|)$ in some neighborhood of points $\overline{x}$, $\overline{x}_i$, $\overline{x}_j$. Only the linear part of this expansion was left in fact:

$$K(\overline{x},\overline{x}_1) = K(\overline{x}_1,\overline{x}_2) + (K(\overline{x}_1,\overline{x}_3) - K(\overline{x}_1,\overline{x}_2))[(x_2 - x_1)^{-1}(x - x_1) + (y_2 - y_1)^{-1}(y - y_1)], \tag{3.9}$$

within some radius where three observed points are available. The vertical derivatives were ignored because data on all selected levels were available.

The possibility of representing the covariance with respect to (3.9) depends on the homogeneity of the error field. The covariance function at an arbitrary grid point is constructed through the others, and within the chosen radius we use the same value for the spatial derivative. Without this assumption, we would need a dense observational coverage with at least three points of measurement around each grid point. In practice, the assumption of the local homogeneity of error fields within some radius is sufficient. Outside the local circle the covariance is assumed to go to zero.

Resume our algorithm from point to point.

0. We start from the known initial field.

1. Suppose the fields $\xi_m(\tau,\overline{x})$ up to the moment t have already been constructed. The transition from t to $t + dt$ is as follows:

(a) The field $\xi(t,\cdot)$ is taken as the initial, and model computation is carried out.

(b) Using the fields $\xi(t,\cdot)$ and $\xi_m(t + dt,\cdot)$, as well as the observations at moment $t + dt$, the drift vector $\overline{a}(t)$ and the diffusion matrix $B^2(t)$ are obtained, using (3.6) and (3.7).

(c) These parameters are inserted into the Fokker–Planck equation (3.8), and the new probability distribution is found.

(d) The covariance matrix is obtained with respect to (3.4) and (3.9).

2. The new field $\xi(t + dt)$ is constructed by (3.1). This completes the description of the method.

At the end of this section, we estimate the required number of operations for numerical realization of this method. Let N denote the number of grid points, and let M be the number of observations. Usually we have $N \sim M^2 \ll N^2$.

Suppose, for clarity, that the forward computations require N operations, where N means the number of grid points, and the inverse problem requires N^2 ones (standard situation). This gives the following estimation:
Step 1(a) requires $\sim N$ operations; this is a forward model computation. No inverse problem is solved at this step.
Step 1(b), as was already mentioned above, also requires $\sim N$ operations.
Step 1(c) requires $\sim M^2$ operations; it is necessary to solve the Fokker–Planck equation for each pair of measurements. The expenses for realization of this equation can be neglected as compared to other computations. The linear interpolation with respect to (3.9) also requires $\sim N$ operations.
Step 2. Requires $\sim M^2$ operations. We need to invert an $M \times M$ matrix.

Thus, the total number of operations has an order $\sim M^2$.

4. The Numerical Experiments and Results

The assimilation method is now applied to processes in the North Atlantic subpolar frontal zone. The research area is limited by $37.25N$, $47.75N$ and $37.25W$, $50.75W$. The spatial resolution in the model is $0.25° \times 0.25°$ and the time step equals 2 h. In this region, observations have been carried out by the State Oceanography Institute (Moscow, Russia) for more than 10 yr, and about 7000 hydrological stations have been recorded. A lot of papers deal with the hydrological description and seasonal and annual cycles of the processes in this area (e.g., [3, 5]). This region plays an important role in the air–sea interaction processes in the North Atlantic and influences short-period climate oscillations (see [17]). The typical hydrological structure of this area is known, and it is possible to compare the results of our data-assimilation technique to observations.

For our purposes, we chose one hydrological survey (March–April, 1985). This survey contains 105 temperature and salinity profiles to a depth of 2000 m. It was carried out from North to South on the transects with a distance between stations of 50 km.

The model experiments are carried out as follows: the STD (salinity, temperature, depth) data are assimilated according to their temporal and spatial location. If the distance between the data and grid point exceeds some value (r_c), the correlation is assumed to be equal to 0. In our experiments, the cutoff radius equals 6 grid points (approximately 150 km). This value is not critical to the model result.

The time behavior of the kinetic energy in the model without data assimilation has been studied. The minimum of this value reflects an equilibrium state that is chosen as the start time of the assimilation. This moment (approximately 10 days) depends on the viscosity coefficients. In the model,

$$\mu = 10^6 \text{ cm}^2/\text{sec}, \qquad \nu = 10^6 \text{ cm}^2/\text{sec},$$

$$\mu_T, \mu_s = 2.5 \cdot 10^6 \text{ cm}^2/\text{sec}, \qquad \nu_T, \nu_s = 2.5 \cdot 10^6 \text{ cm}^2/\text{sec}$$

is used.

The assimilated data are a source of perturbations in the model, and we examine the perturbed fields and the propagation of these perturbations. Consider in detail the behavior of the difference between the fields before and after the assimilation. We will call these differences "anomalies" or "perturbations." These anomalies occur near the measurements but then begin to move and influence the model dynamics. Outside the North Atlantic current the anomalies are almost stationary, and further assimilation of observations in this area reduces the forecasting error. The errors may be reduced further by allowing the "noise" to be a function of location.

The spatial distribution of perturbations is not uniform and the form of the anomalies is very complicated. In the beginning the perturbations appear to follow the direction of the cruise vessel but then forcing by advection, diffusion, and boundaries make their motions more complex. The anomalies influence one another, causing them to deform and even vanish. Eddy generation takes place in the model due to both the natural current instability and notability triggered by new perturbations advected by the flow. The anomalies propagates through two kinds of motions — wave motion, mostly westward, and advection. It is possible to show that the linearized equations of motion of the model contain damped barotropic and baroclinic Rossby waves as well as baroclinic gravity waves. Barotropic gravity waves are repressed because of our use of the cutoff radius r_c. Our experiments show that the decay time of these waves is short (two or three days) and the advective mechanism dominates. Through nonlinear interaction and diffusion, the perturbations are deformed and destroyed.

Throughout the experiment, the main oceanographic structures in this region are found in the model result: the frontal zone, the anticyclonic quasistationary eddy in the middle of domain, and the meanders of the North Atlantic current. Data assimilation alters the structure of the region. There is no quasistationary eddy in the middle of the

region and the frontal zone is shifted northward and strengthened. The result of our assimilation fits the observation of Clarke well (see [5]). The model with assimilation is able to recover the typical hydrological structures of the region.

Note the effects of the assimilation near the lateral boundaries. Up to the end of the time of integration (100 days) near the western boundary without assimilation a large temperature anomaly occurred (this is an example of the model deficiencies). These anomalies occurred due to the discrepancy between the boundary conditions (climatology) and the dynamics inside the domain. After the assimilation, near the lateral boundaries the anomaly almost disappears.

We now discuss the quality of our method. The usual concept is to estimate the variance of the forecasting error (difference between the model prediction and observed values) at each time step. This is not ideal because the number of observations is not uniform in time and their spatial location seriously influences the error. However, some interesting properties arise. Near $t = 0$, when the initial fields are far from the observations, the variance of the error is large. Then, after the first or second assimilation the variance decreases and then oscillates around some positive value. This may be explained by the fact that after some assimilation the new fields are "turning on" the measurements and afterwards it is impossible to substantially decrease the error without additional improvement of the model and/or data-assimilation technique.

Appendix

In (2.1), for the adynamic sea level ξ' the function $\overline{f}$ has the following form (see [18]):

$$\overline{f} = \left(H^{-1}\frac{g}{\rho_0 f}\right)^2 J\left(\int_0^H z\rho\,dz, \int_0^H z\Delta\rho\,dz\right) + (\rho_0 H)^{-1}\int_0^H z\partial(\Delta\rho)\,dz$$

$$-(\rho_0^2 f H)^{-1} g\int_0^H J\left(\int_0^z \rho\,dz, \int_0^z \Delta\rho\,dz\right)dz - (2\alpha\rho_0 H)^{-1} f\int_0^H \Delta\rho\,dz$$

$$-\frac{1}{\rho_0 H}\int_0^H J(H,\rho)\,dz - \frac{\beta}{\rho_0 H}\int_0^H (H-z)\frac{\partial p}{\partial x}\,dz + \mu(H\rho_0)^{-1}\int_0^H (H-z)\Delta\Delta\rho\,dz,$$

where p is the atmospheric pressure.

In practice, we neglect the terms with second-order Laplacian and atmospheric pressure derivatives.

In the Rossby number decomposition, in (2.4)–(2.5) there is the following representation of the functions in the series:

$$F_0 = -g(UL)^{-1}\frac{\partial\xi'}{\partial x} - g(UL\rho_0)^{-1}\int_0^z \frac{\partial\rho}{\partial x}\,dz,$$

$$F_1 = -f\left(u_0\frac{\partial u_0}{\partial x} + v_0\frac{\partial u_0}{\partial y}\right) + fw_0\frac{\partial u_0}{\partial z} + \mu f(LU)^{-1}\Delta u_0,$$

$$F_2 = -f\left[\left(u_0\frac{\partial u_1}{\partial x} + u_1\frac{\partial u_0}{\partial x}\right) + \left(v_0\frac{\partial u_1}{\partial y} + v_1\frac{\partial u_0}{\partial y}\right)\right] + f\left(w_1\frac{\partial u_0}{\partial z} + w_0\frac{\partial u_1}{\partial z}\right) + \mu f(LU)^{-1}\Delta u_1,$$

where U and L are the characteristic scales of velocities and space, and u_i and v_i, $i = 0, 1, 2$, are the terms of decomposition,

$$w_i = \int_0^z \left(\frac{\partial u_i}{\partial x} + \frac{\partial v_i}{\partial y}\right)dz.$$

For the function G_i we have

$$G = -g(UL)^{-1}\frac{\partial\xi'}{\partial y} - g(UL\rho_0)^{-1}\int_0^z \frac{\partial\rho}{\partial y}\,dz,$$

$$G_1 = -f\left(u_0\frac{\partial v_0}{\partial x} + v_0\frac{\partial v_0}{\partial y}\right) + fw_0\frac{\partial v_0}{\partial z} + \mu f(LU)^{-1}\Delta v_0,$$

$$G_2 = -f\left[\left(u_0\frac{\partial v_0}{\partial x} + u_1\frac{\partial v_1}{\partial y}\right) + \left(v_0\frac{\partial v_1}{\partial x} + v_1\frac{\partial v_0}{\partial y}\right)\right] + f\left(w_1\frac{\partial v_0}{\partial z} + w_0\frac{\partial v_1}{\partial z}\right) + \mu f(LU)^{-1}\Delta u_1.$$

REFERENCES

1. A. Alexandrochkin and K. Belyaev, "The seasonal and annual variability in the North Atlantic subpolar frontal zone in a model with data assimilation," *Meteorol. Hydrol.*, **7**, 64–78 (1993).
2. B. D. O. Anderson and J. Moore, *Optimal Filtering*, Prentice-Hall, Englewood Clifts, New Jersey (1979).
3. K. Belyaev, S. Gulev, S. Lappo, and V. Tereschenkov, "Experiment Newfaex 88. The heat-balance estimation in the North Atlantic frontal zone resulting in the numerical model with data assimilation," *Izv. Akad. Nauk SSSR*, **26**, 67–77 (1990).
4. N. Budgell, "Stochastic filtering of linear shalow water equations in branched channels," *J. Geophys. Res.*, **91**, 10633–10644 (1986).
5. A. Clarke, H. Hill, R. Reiniger and B. Warren, "Current system South and East of the Grand Banks of Newfoundland," *J. Phys. Oceanogr.*, **10**, 25–65 (1980).
6. Yu. Demin and R. Ibraev, "On the boundary-value problem for the sea-surface topography in the sea current model," *Izv. Akad. Nauk SSSR*, **21**, 56–72 (1985).
7. D. L. Doob, *Stochastic Processes*, New York (1953).
8. M. Ghil, "Meteorological data assimilation for oceanographers. 1. Description and theoretical framework," *Dyn. Atmos. Ocean*, **13**, 197–218 (1989).
9. M. Ghil and P. Malanotte-Rizzoli, "Data assimilation in meteorology and oceanography," *Adv. Geophys.*, **33**, 141–266 (1991).
10. G. Evensen, "Using the Extended Kalman-filter with a multilayer quasi-geostrophic ocean model," *J. Geophys. Res.*, **97**, 17905–17924 (1992).
11. A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, Academic Press, New York (1970).
12. V. Knysh, V. Moiseenko, and V. Chernov, "Four-dimensional analysis model, theory and experiments," *Hydrometizdat, Ser. Ocean. Atm. Space*, 102–150 (1991).
13. V. Knysh, "The four-dimensional analysis model and its application to the research of processes in the North Atlantic," *Hydrometizdat, Ser. Ocean. Atm. Space*, 56–78 (1988).
14. A. N. Kolmogorov, *Sammelband zur Statistische Theorie der Turbulenz*, Akademie-Verlag, Berlin (1958).
15. S. Levitus, *Climatological Atlas of the World Ocean*, NOAA Prof. Pap 13 Natl. Oceanic Atmos. Admin., Rockville, Maryland (1982).
16. P. Malanotte-Rizzoli and W. Holland, "Data constraints applied to models of the ocean general circulation. I. The steady case," *J. Phys. Oceanogr.*, **16**, 1665–1687 (1987).
17. G. I. Marchuk, *Methods of Numerical Mathematics*, Springer, Berlin (1982).
18. G. I. Marchuk and A. S. Sarkisyan, *Mathematical Modeling in the Ocean Circulation*, Springer, Berlin (1987).
19. R. N. Miller and M. Gane, "A Kalman-filter analysis of sea level height in the tropical Pacific," *J. Phys. Oceanogr.*, **19**, 773–790 (1989).
20. D. Muller, "Bispectra of sea-surface temperature anomalies," *J. Phys. Oceanogr.*, **17**, 26–36 (1987).
21. I. M. Navon, X. Zou, J. Derber, and J. Sela, "Variational data assimilation with an adiabatic version of the spectral model-Mon," *Weather Rev.*, **120**, 1433–1446 (1992).
22. D. W. Peaceman and H. H. Rachford, Jr., "The numerical solution of parabolic and elliptic differential equations," *J. Soc. Ind. Appl. Math.*, **3**, 28–41 (1955).
23. V. Penenko and N. A. Obrastov, "A variational initialization method for the field meteorological elements," *Meteorol. Hydrol.*, **11**, 1–11 (1976).
24. A. Robinson and L. Walstad, "Model: Calibration application to dynamical process forecasting and data assimilation studies," *J. Appl. Numer. Math.*, **3**, 89–121 (1987).
25. O. Talagrant and P. Courtier, "Variational assimilation of meteorological observations with the adjoint vorticity. Equation I," *Theor. Q. J. R. Meteorol. Soc.*, **113**, 1311–1328 (1987).
26. I. E. Timchenko, *Dynamic-Stochastic Models and Their Application in Oceanography* [in Russian], Hydrometizdat, Leningrad (1986).
27. W. Tracker, "Fitting models to data by enforcing spatial and temporal smoothness," *J. Geophys. Res.*, **93**, 10655–10665 (1988).
28. UNESCO (1981).
29. L. Yu and J. J. O'Brien, "Variational data assimilation for determining the seasonal net surface heat flux using a tropical Pacific Ocean model," *J. Physic. Oceanogr.*, **25**, 2319–2343 (1995).

30. A. M. Yaglom, "Introduction to the theory of stational random processes," *Adv. Math.*, **5**, 3–168 (1953).

Shirshov Institute of Oceanology,
Moscow,
Russia

Florida State University,
USA