



ELSEVIER

Applied Mathematical Modelling 25 (2001) 655–670

APPLIED
MATHEMATICAL
MODELLING

www.elsevier.nl/locate/apm

A data assimilation method used with an ocean circulation model and its application to the tropical Atlantic

Konstantin P. Belyaev^a, Clemente A.S. Tanajura^{a,*}, James J. O'Brien^b

^a *Laboratório Nacional de Computação Científica (LNCC/CNPq), Av. Getulio Vargas 333, Petropolis, 25651-070 RJ, Brazil*

^b *Center of the Ocean-Atmosphere Prediction Studies (COAPS), Florida State University (FSU), Johnson Building, Innovation Park, Tallahassee, FL 32306-2840, USA*

Received 25 October 1999; received in revised form 10 October 2000; accepted 4 December 2000

Abstract

A data assimilation method based on the Kalman-filter theory is presented and tested with the NOAA/GFDL Modular Ocean Model (MOM_2). The method is used to assimilate observed tropical Atlantic Ocean surface and subsurface temperatures obtained from the PIRATA moored buoy array. Comparison between model results with and without assimilation shows that the method works properly, and that it has some advantages in relation to a standard Kalman-filter technique. It provides better agreement with the observed temperature intraseasonal variability, and it does not require linearity of a numerical model. © 2001 Elsevier Science Inc. All rights reserved.

Keywords: Data assimilation; Fokker–Planck equation; NOAA/GFDL MOM_2 ocean circulation model; PIRATA project

1. Introduction

In recent years, data assimilation methods have been developed and used in many research areas. The first application of data assimilation techniques were in meteorology, and today it is a key component of numerical weather and climate forecasts. Satellites and in situ measurements are routinely providing new oceanographic data and bringing the daily practice of physical oceanography closer to that of dynamic meteorology. There are large observational data sets of temperature and salinity covering oceanic areas because of international projects such as WOCE, SECTIONS, TOGA-TAO, COARE, PIRATA, and others.

Improvements in assimilation techniques may contribute to the full exploitation of the new observational facilities. Data assimilation improves the estimate of the ocean and atmosphere physical state by combining the data from measurements and from dynamic models in an optimal way. In this sense, it extracts as much as possible information from the available resources. Data assimilation may be used to improve initial and/or boundary conditions, to create new data for

* Corresponding author. Tel.: +55-24-233-6164; fax: +55-24-231-5595.

E-mail addresses: kostia@lncc.br (K.P. Belyaev), cast@lncc.br (C.A.S. Tanajura), obrien@coaps.fsu.edu (J.J. O'Brien).

diagnostic studies of the climate system, and to evaluate poorly known model parameters. Therefore, it can improve model weather and climate predictability.

Data assimilation in ocean models has been discussed in the scientific literature for approximately 30 years (see [1] for a review). It is necessary to understand which processes can realistically be described by ocean models, and what is the role of the measurements. There are two extreme opinions about this issue. The first one considers the ocean as a quasi-stationary medium, such that it would be sufficient to make few measurements to get a good description of the ocean state. From this point of view, numerical models are not important and data assimilation techniques are restricted to a statistical correction process. The second opinion considers the ocean as a highly turbulent fluid, which does not have a memory of its previous states. In this case, information generated by measurements would become meaningless in short time intervals. Assimilation methods would be mostly needed to provide initial conditions to complex dynamical models, which would describe the real processes in the ocean. Reality, however, lies somewhere in between these two extreme points of view. Therefore, there is the scientific question on how to represent this situation numerically. The answer depends on both the model and the data assimilation techniques, as well as on the local dynamics.

Two general concepts on the mathematical formulation of data assimilation methods have been discussed in the literature. The first is the variational/adjoint method, which has been the most popular scheme, e.g. [2–5]. As an example of this technique, let the model initial and/or boundary condition be unknown and the observational data be distributed over some time-interval. The technique seeks an optimal initial and/or boundary condition with respect to some criteria by comparing the model trajectory with measurements. This can be formulated as a constrained minimization problem. The present paper does not deal with this class of methods.

The other class of methods is the sequential data assimilation. Starting from some initial condition, the model solution is sequentially updated whenever measurements are available. The model solution approaches the observed state under certain conditions. This class of methods requires an updating scheme, which combines the model solution and the measurements to find the “best” state estimate. The Kalman-filter method belongs to this class and is used in this paper.

The Kalman-filter theory is derived in a number of books on the control theory (e.g. [6,7]). In oceanography, Kalman-filter has been applied in [8,9] and others. The main idea of this method is to write and solve the dynamic equations for the error covariance matrix, where error is defined as the difference between the model and the observational values. Solving these equations is computationally expensive, and, in addition, linearity of the dynamic system operator is assumed. There are some attempts to extend the Kalman-filter theory to non-linear operators, but they also require other simplifications, e.g. [10,11].

In this paper, another method to solve the dynamic equation system on the error covariance matrix is developed. The present work is a continuation of the discussion presented in [12]. This method is based on stochastic process theory and parabolic partial differential equations.

Stochastic processes are commonly used in the turbulence model proposed by Kolmogorov [13] and in climate models with stochastic forcing (e.g. [14]), but they have not been very much applied to data assimilation. Some discussions in this direction have been shown in [11], but it remained unclear of how to apply it. The method presented here is relatively simple. It does not require intense computations, but contains the basic features of the physical processes in the region of interest. The main idea is to consider the time variability in phase space, avoiding the necessity to follow each spatial point. The diffusion approximation is used to describe the time variability of the error, and then the Fokker–Planck (FP) equation is solved to obtain the joint probability distribution of each pair of measurements. The ordinary Kalman-filter method requires $O(n^2)$ operations to invert the model operator to obtain the covariance function, where n is the number

of grid points. This method requires the inversion of the operator only with respect to the number of measurement points. It is much simpler because the number of observational points is commonly less than the number of grid points. The method does not depend on whether the model is linear or not. However, it has some weaknesses. A discussion about the advantages and disadvantages of the method compared with another data assimilation methods is presented below.

The present method is used with the NOAA/GFDL primitive equation ocean model MOM_2 [15,16]. The goal is to demonstrate the usefulness of the assimilation technique. MOM_2 has been widely applied in forecast and simulation experiments, e.g. [17]. Here, it is used only as a tool to calculate the thermodynamic fields and their variability.

This paper also uses the Pilot Research Moored Array in the Tropical Atlantic (PIRATA) data set [18], which contains surface and subsurface temperatures at several levels up to 500 m depth and some other quantities. This project is being carried out by three countries, USA, France and Brazil. Data are collected by a number of moored buoys in the Tropical Atlantic since the end of 1997. PIRATA also provides data to diagnostic studies on the Atlantic circulation, model development, model validation and initialization of weather and climate forecasts. Data used here were taken from the public website www.ifremer.fr/orstrom/pirata. The authors have done no analysis with these data.

The assimilation technique is discussed in Section 2, and its numerical realization in Section 3. The results are discussed in Section 4, and the last section presents the summary of the work and the conclusions.

2. Data assimilation technique

Let $\zeta(t, \bar{x})$ be an unknown real or “true” variable, which is sought in some ocean domain Ω . The model approximation for ζ is written in Cartesian coordinate system as

$$\frac{\partial \zeta_m}{\partial t} = A(t)\zeta_m, \quad (1)$$

where $A(t)$ is the known model operator, in general, non-linear, t represents time and $\bar{x} = (x, y, z)$ denotes an arbitrary spatial point of the domain. The symbol ‘bar’ above is used to distinguish a spatial point from its coordinate.

To obtain the model variable $\zeta_m(t, \bar{x})$ Eq. (1) is integrated from the initial known state at $t_0 = 0$ up to $t = T$. During this period, it is assumed that the observed values $\zeta_o(\tau, \bar{x}(\tau))$ of the variable ζ are available at time τ , $0 < \tau < T$. The problem is to construct the optimal estimation of the variable $\zeta(t, \bar{x})$ using both model and observed data.

Hereafter, it is assumed there is no difference between the observed and the real values of variable $\zeta(t, \bar{x})$. In general, it is possible to consider the case where the real value is equal to the observed one plus an error of measurements. However, there is no need to do it, because it does not add anything essentially new, but complicates the formula. And, also the error due to modeling is normally much greater than the error in the measurements.

A basic hypothesis of the Kalman-filter theory is that the variable $\zeta(t, \bar{x})$ satisfies the expression:

$$\frac{\partial \zeta}{\partial t} = A(t)\zeta + \eta, \quad (2)$$

where the operator $A(t)$ is given in formula (1), and η is a random noise with a known stochastic distribution. Let $\theta = \theta(t, \bar{x}) = \zeta(t, \bar{x}) - \zeta_m(t, \bar{x})$ be the error due to modeling, simply referred to as error. If $A(t)$ is linear, from Eqs. (1) and (2) it is obtained that

$$\frac{\partial \theta}{\partial t} = A(t)\theta + \eta. \quad (3)$$

If the operator is non-linear, Eq. (3) may also be valid to represent the error, but of course with a different distribution of a noise. The following standard conditions for the noise are assumed to hold

$$E\eta = 0, \quad E\eta(\bar{x}, \tau)\eta(\bar{y}, t) = R(r)\delta(t - \tau). \quad (4)$$

In formula (4), E denotes the mathematical expectation, $\delta(t)$ is the Dirac-function, and

$$r = \text{dis}(\bar{x}, \bar{y}) = [(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2]^{1/2}$$

is the distance between two spatial points \bar{x}, \bar{y} . The function R is supposed to be known.

In the present paper, two kinds of averages are considered. Let symbol E be the ensemble average, and the symbol $\langle \rangle$ be the spatial average.

If in addition it is assumed that

$$E\theta(0, \bar{x}) = 0; \quad A(0) = 0,$$

then and by (3) and (4) the equality

$$E\theta(t, \bar{x}) = 0 \quad (5)$$

follows for any t .

The following problem is considered ([19,20] and others): find the optimal estimation $\hat{\zeta}(t, \bar{x})$ or the optimal filter of $\zeta(t, \bar{x})$ satisfying the conditions:

- (a) $E(\hat{\zeta} - \zeta) = 0$;
- (b) $E(\hat{\zeta} - \zeta)^2 = \min E(\omega - \zeta)^2$ for any estimation ω .

This is the well-known problem of searching an unbiased estimation with minimum variance.

Condition (a) and Eq. (5) lead to

$$E\hat{\zeta}(t, \bar{x}) = \zeta_m(t, \bar{x}).$$

The optimal filter is sought by the expression:

$$\hat{\zeta}(t, \bar{x}) = \zeta_m(t, \bar{x}) + \int_0^t \sum_{i=0}^{N(\tau)} \alpha(\tau, \bar{x}, \bar{x}_i) \theta(\tau, \bar{x}_i) d\tau, \quad (6)$$

which is an optimal linear filter with unknown weight coefficients $\alpha_i = \alpha(\tau, \bar{x}, \bar{x}_i)$. In this formulation, without loss of generality, observation times are continuously distributed, and the locations of observation are discrete in space. Let $N(\tau)$ be the number of observations at time τ . The unknown weight-coefficients $\alpha_i = \alpha(\tau, \bar{x}, \bar{x}_i)$ should be determined using condition (b). The optimal linear filter can be found out as in [19], where it was proved that, under some conditions, the optimal linear filter would be the best among all filters with the same risk function.

After some mathematical manipulations, the necessary minimum condition $(\partial/\partial\alpha_i)E(\hat{\zeta} - \zeta)^2 = 0$ leads to the following equation for α :

$$K(t, \bar{x}, \bar{x}_i) = \int_0^t \sum_{j=0}^{N(\tau)} \alpha(\tau, \bar{x}, \bar{x}_j) K(t - \tau, \bar{x}_i, \bar{x}_j) d\tau, \quad (7)$$

where $K = K(t, \bar{x}, \bar{y})$ is the covariance function of the error, given by

$$K(t, \bar{x}, \bar{y}) = E\theta(t, \bar{x})\theta(t, \bar{y}) - E\theta(t, \bar{x})E\theta(t, \bar{y}), \quad i = 1, \dots, N(t).$$

When the covariance is known, Eq. (7) is solved with respect to α , and the solution is inserted into formula (6) to obtain the optimal linear filter. However, the main problem of the Kalman-filter theory is to find the error covariance function and its variability in time. If the operator $A(t)$ is linear, (3) leads to the equation for the covariance function

$$\frac{\partial}{\partial t} K(t, \bar{x}, \bar{y}) = A(t)K(t, \bar{x}, \bar{y}) + R(t, \bar{x}, \bar{y}) \tag{8}$$

(see e.g. [1,6]). If the operator is non-linear, (8) is not valid. But even in the linear case, the problem of solving Eq. (8) is very difficult. First, this equation should be solved to all pairs of grid points. It is easy to calculate that if the number of grid points is around 10^4 – 10^5 , the number of operations to invert the operator matrix exceeds 10^9 and even modern supercomputers are unable to do it. In addition, this is not justified, because Eq. (7) needs the covariance only at the measurement points. Second, Eq. (8) requires initial and boundary conditions to the covariance. Because of the absence of real information, any initial value is set up artificially and this will influence the final result. It is possible to say even more about the boundaries.

Another way is suggested to solve the equation for the covariance function and its time evolution. This approach can be used for non-linear operators. According to the definition, the error covariance is

$$K(t, \bar{x}, \bar{y}) = E\theta(t, \bar{x})\theta(t, \bar{y}) - E\theta(t, \bar{x})E\theta(t, \bar{y}).$$

Let $p(t, s_1, s_2)$ be joint distribution of the error of two spatial points, i.e., $\theta(t, \bar{x}) = s_1$, $\theta(t, \bar{y}) = s_2$ with the probability $p(t, s_1, s_2)$. Then, the covariance can be written as

$$K(t, \bar{x}, \bar{y}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} s_1 s_2 p(t, s_1, s_2) ds_1 ds_2 - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} s_1 p(t, s_1, s_2) ds_1 ds_2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} s_2 p(t, s_1, s_2) ds_2 ds_1. \tag{9}$$

It is important to note that, in general, the average value of the error at a fixed point \bar{x} does not equal zero. Equality (5) is correct for an a priori considered point. But if this point is taken a posteriori, Eq. (5) is not valid. In other words, Eq. (5) is correct if an average is taken over a field. Following (9), the problem of the determination of the covariance is solved after the joint distribution $p(t, s_1, s_2)$ is found.

Let $\bar{\theta}(t) = (\theta(t, \bar{x}), \theta(t, \bar{y}))$ denote the vector of a pair of errors at two points. The increment of this vector in phase space can be represented by the following stochastic differential equation, the so-called Langevin equation,

$$d\bar{\theta}(t) = \bar{a}(t, \bar{\theta}) dt + B(t, \bar{\theta}) dW, \tag{10}$$

where $\bar{a} = (a_1, a_2)$ is the drift vector, $B = (b_{ij})$, $i, j = 1, 2$ is the diffusion matrix, and W is a two-dimensional Wiener process [20]. Parameters of Eq. (10) are defined by [20]

$$\bar{a}(t, \bar{s}) = \frac{\partial}{\partial t} E(d\bar{\theta}(t)/\bar{\theta}(t) = \bar{s}), \tag{11.1}$$

$$B(t, \bar{s}) = \frac{\partial}{\partial t} E[d\bar{\theta}(t)(d\bar{\theta}(t))' / \bar{\theta}(t) = \bar{s}], \tag{11.2}$$

where the symbol “ $'$ ” denotes the transpose of the vector.

Formulas (11.1) and (11.2) use the conditional average and the conditional variance to determine the drift vector and the diffusion matrix. Using Eq. (3), formulas (11.1) and (11.2) can be rewritten after simple mathematical transformations as

$$\bar{a}(t, \bar{s}) = E\left(A(t)\bar{\theta}(t)/\bar{\theta}(t) = \bar{s}\right), \quad (12.1)$$

$$B(t, \bar{s}) = E\left[A(t)\bar{\theta}(t)(A(t)\bar{\theta}(t))'/\bar{\theta}(t) = \bar{s}\right] + R(t, r) \quad (12.2)$$

in which it was taken into account that $E(\eta\eta') = R(r)$.

Formulas (12.1) and (12.2) allow obtaining the values \bar{a} , B . They show that knowing the value $\bar{\theta}$ at time t the operator A can be applied, the average and the variance are calculated. Despite the error field $\bar{\theta}$ being unknown, its values are known under the given conditions. In other words, it means that any variable with the same values can be used and the same results will follow. The methodology applied to obtain the solution will be considered in the next section.

After the determination of the drift vector \bar{a} and the diffusion matrix B , the probability distribution is found by the FP equation

$$\frac{\partial p(t, \bar{s})}{\partial t} = -\frac{\partial(\bar{a}(t, \bar{s})p(t, \bar{s}))}{\partial \bar{s}} + \frac{1}{2} \frac{\partial^2}{\partial \bar{s}^2} (B(t, \bar{s})p(t, \bar{s})). \quad (13)$$

The following notation was used

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

$$\frac{\partial^2}{\partial \bar{s}^2} (Bp) = \frac{\partial^2}{\partial s_1^2} (b_{11}p) + 2 \frac{\partial^2}{\partial s_1 \partial s_2} (b_{12}p) + \frac{\partial^2}{\partial s_2^2} (b_{22}p).$$

The FP equation (13) requires initial and boundary conditions. For the boundary conditions, Sommerfeld conditions are naturally imposed, with the probability vanishing as t tends to infinity, i.e.,

$$p(t, +\infty) = p(t, -\infty) = 0.$$

The initial conditions for the probability distribution are defined as follows. If two observations are done at the same time t^0 , it means that at this moment the error is known. Let the error be $\bar{s}^0 = (s_1^0, s_2^0)$. Hence, the initial probability density at moment t^0 is the two-dimensional Dirac δ -function, $p(t^0, \bar{s}) = \delta(\bar{s} - \bar{s}^0)$. It is easy to be generated when two measurements are done consecutively, at time $t_1^0, t_2^0, t_1^0 < t_2^0$. So, the problem to set up the initial distributions of any pair of all measurement points is solved without artificial assumptions. However, it is more difficult to define the initial distribution between an arbitrary spatial point of the domain and a measurement point according to the left side of Eq. (7). The simplest way to do it is to make an interpolation of the covariance from the nearest measurement point. The accuracy of this interpolation should have the same order of the model accuracy. But a good coverage of the domain by observations is needed. Within some selected radius h_c , where the error field is assumed to be homogeneous (local homogeneous assumption) at least three observations should be available, and they should not lie on the same straight line (non-degenerated case). Then, the new covariance can be reconstructed by

$$K(\bar{x}, \bar{x}_i) = \sum_j \gamma_{ij} K(\bar{x}_i, \bar{x}_j), \quad (14)$$

where \bar{x}_i, \bar{x}_j are points of observations, \bar{x} is an arbitrary grid point, and γ_{ij} are the interpolation coefficients. Only, the linear interpolation is considered. There is no need to write a precise formula for these coefficients, since the linear interpolation is commonly used. The radius h_c cannot be chosen very large because of both the interpolation accuracy and the local homogeneous assumption.

If there are insufficient observations or three or more points belonging to the same line, this method does not work. In this case, it is necessary to introduce additional hypothesis. One of them, which is used in the experiment realization is as follows. An additional “artificial” point of observation is introduced in the neighborhood of two real observational points, in such a way that the three points do not lie over the same straight line. At that point, the distribution of error is taken a priori, e.g., as a Gaussian with zero average and variance equal to the variance of the noise. Then, the FP equation is applied to obtain the covariance between this additional point and its two real neighbors. This reduces the problem to the previous case. Of course, this approach is not ideal, but at least it strongly restricts the indefiniteness on the initial conditions. Naturally, it is desirable to introduce as less additional artificial observations as possible.

In the end of this section a summary of the method considered here is presented. Step-by-step the procedure is:

1. Perform the model integration according to Eq. (1) from the known initial field $\zeta_0(\bar{x})$ until the first assimilation time t ;
2. Compute the model field $\zeta_m(t + dt, \bar{x})$;
3. Determine the drift vector \bar{a} and the diffusion matrix B using both model and observed values (formula (11.1) and (11.2));
4. Apply the FP equation to find the covariance at time t (Eq. (13));
5. Interpolate the covariance from the points of observations to an arbitrary grid point of the domain (formula (14));
6. Calculate the weight-coefficients α with respect to Eq. (7);
7. Determine the optimal filter $\hat{\zeta}(t, \bar{x})$ (formula (6)).

This completes the description of the methods.

3. The numerical realization

This section focuses on the description of the algorithm to perform the numerical realization with the method discussed above.

Let the model equation (1) be integrated numerically from t_0 until t , and let t_1, \dots, t_n be the times at which data assimilation or correction is done. Suppose that the new corrected fields $\hat{\zeta}_i(\bar{x})$ have already been constructed until time $t_n = t$. The transition from time t_n to time t_{n+1} is presented according to the sequence below. The numbers labeling each step refer to the steps 1–7 of the previous section. The numerical realization requires:

1–2. The numerical solution of equation (1). This does not need a special description since it is not the goal of this paper.

3. The computation of the drift vector and the diffusion matrix. This requires the construction of the conditional probability of the error $p(\bar{\theta}_{n+1} = \bar{u}/\bar{\theta}_n = \bar{s}) = p_n(\bar{u}/\bar{s})$ according to formulas (11.1) and (11.2). The known field $\varepsilon_n = \hat{\zeta}_n(\bar{x}) - \langle \hat{\zeta}_n(\bar{x}) \rangle$ is taken as a condition. As it has already been mentioned above, the only requirement this field should have is an average equal to zero.

Instead of the unknown “ensemble” mean value, the spatial average $\langle \hat{\varepsilon} \rangle$ is taken as its estimation. Then, all grid points, where $\varepsilon_n = \bar{\varepsilon}$ are selected. Let these points be $N_{\bar{\varepsilon}} \neq 0$. At time t_{n+1} , grid points are sought among the selected grid points where the model field $\varepsilon_{n+1} = \bar{u}$. Let these points be $N_{\bar{u}}$. Then $p_n(\bar{u}/\bar{\varepsilon}) = N_{\bar{u}}/N_{\bar{\varepsilon}}$. Obviously, $0 \leq p_n(\bar{u}/\bar{\varepsilon}) \leq 1$, i.e., it is really probability. Then, the conditional average and the variance are defined with respect to formulas

$$\bar{a}_n(\bar{\varepsilon}) = \lambda_n^{-1} \int_{-\infty}^{\infty} (\bar{u} - \bar{\varepsilon}) p(\bar{u}/\bar{\varepsilon}) d\bar{u},$$

$$B_n(\bar{\varepsilon}) = \lambda_n^{-1} \int_{-\infty}^{\infty} (\bar{u} - \bar{\varepsilon})(\bar{u} - \bar{\varepsilon})' (p(\bar{u}/\bar{\varepsilon})) d\bar{u} + R,$$

$$\lambda_n = t_{n+1} - t_n.$$

It is important to note that the number of computations to define the conditional probability is not large. Indeed, it is easy to note that $p_n(\bar{u}/\bar{\varepsilon}) = 0$ when $\text{abs}(\bar{u} - \bar{\varepsilon}) > L$ for some not very large L . Also, because of the symmetry of the conditional probability $p_n(\bar{u}/\bar{\varepsilon}) = p_n(\bar{\varepsilon}/\bar{u})$, there is no need to calculate this value twice. Both of these properties strongly restrict the necessary number of operations.

4. The numerical solution of the FP-equation. This is a two-dimensional parabolic differential equation on an unbounded plane. Well-known numerical methods were developed to solve it. Peaceman–Rachford’s algorithm [21] has been applied in the present paper, but with one modification. The numerical scheme is written as

$$\frac{p^{l+1/2} - p}{\tau} = \frac{1}{2} \left[(b_{11}p)_{\bar{x}\bar{x}}^{l+1/2} + 2(b_{12}p)_{\bar{x}\bar{y}}^l + (b_{22}p)_{\bar{y}\bar{y}}^l \right] + (a_1p)_{\bar{x}}^l + (a_2p)_{\bar{y}}^l,$$

$$\frac{p^{l+1} - p}{\tau} = \frac{1}{2} \left[(b_{11}p)_{\bar{x}\bar{x}}^{l+1/2} + 2(b_{12}p)_{\bar{x}\bar{y}}^l + (b_{22}p)_{\bar{y}\bar{y}}^{l+1} \right] + (a_1p)_{\bar{x}}^l + (a_2p)_{\bar{y}}^l.$$

Here, l is the iteration number. As usual, the following notations are introduced to describe the finite-difference approximation:

$$p_x = \frac{p_{i+1j} - p_{ij}}{h_1}; \quad p_{\bar{x}} = \frac{p_{ij} - p_{i-1j}}{h_1}; \quad p_{\bar{x}\bar{x}} = \frac{p_{i-1j} - 2p_{ij} + p_{i+1j}}{h_1^2}; \quad p_x = \frac{p_{i+1j} - p_{i-1j}}{2h_1}.$$

The same notations are used in the y direction with the substitution of index i by j . Here, i, j are indices of the grid in a phase-space, $h_1(h_2)$ is the grid-size in the $x(y)$ direction in a phase-space, p is the known function at the previous time-step and τ is the time-interval. The calculations end when $\text{abs}(p^{l+1} - p^l) < \varepsilon$ for a small-defined value ε . This determines the probability distribution at any time n from the previous time-step s with the initial probability distribution p^0 .

The two-dimensional Dirac δ -function used as initial condition is approximated as

$$f(s_1, s_2) = (2\pi)^{-1} k^{-1/2} \exp\left(\frac{-1}{2k}(s_1^2 + s_2^2)\right)$$

with some small parameter k .

This scheme gives the solution of FP equation at each time-step. But because the solution is the probability density, it should satisfy the conditions

$$p_{ij}^l \leq 0, \quad \int p^l ds = 1.$$

The first inequality always holds, but the second condition may not be satisfied because of numerical errors. To keep the second condition true, the following correction has been introduced at each time-step. The numerical values p_{ij}^l have been substituted by $\hat{p}_{ij}^l = p_{ij}^l / \int p^{l-1} ds$ in which information from the previous iteration is taken. This stabilizes the procedure and accelerates the computations.

5. This step does not require special discussion.

6–7. The numerical solution of the linear system of equations with a symmetric matrix. This is a well-known procedure. Here, the algorithm of “maximum element” was applied [22].

An estimate of the necessary number of operations required by the algorithm is provided below. It will be assumed that the forward computations with a rank N matrix need $O(N)$ operations, and the inversion algorithm with the same matrix needs $O(N^2)$ computations. Let N be the number of grid-points in the model domain, L be the grid-size in the phase-space and M be the number of observations. Then, an estimate of the number of operations per time-step is

1–2. $O(N)$ for the forward problem with N points.

3. $O(L^2)$ for the computation of conditional average and variance.

4. $O(M^2L)$ for the FP solution, since this equation is solved for each pair of observations with respect to the grid-size L .

5. The number of operations in this step can be neglected.

6–7. $O(M^2)$ for the inverse problem to solve Eq. (7).

The maximum consumption to solve the FP equation is $\sim M^2L$, which can be considered as the total. In practice $N \sim M^2L \sim L^2 \ll N^2$ is used. Because of this, the present algorithm has an essential advantage when compared with the Kalman-filter.

4. The numerical experiments and results

The presented method has been applied to the observational data from the PIRATA project. Daily means of surface and subsurface temperature at 11 levels up to 500 m depth from 10 buoys since 1 January 1999 until 31 May 1999 were used in the assimilation experiments. For the purpose of this paper, it is important to have long-time series of good quality data to be assimilated into the numerical model.

The widely used GFDL global ocean model MOM_2 was used in this study in conjunction with the data assimilation technique. No detailed description of this model will be given here. For more information, see e.g. [15,16]. In the present study the focus is on the tropical Atlantic. MOM_2 was set up to a grid resolution of 0.5° in the meridional direction between 20°S and 20°N . The grid spacing increases linearly to 1.5° towards 40°S and 40°N . The zonal spatial resolution is 1.5° constant through the entire globe. This version has 20 levels from sea-top until bottom with the higher resolution in the first 15 levels from 0 to 1000 m. The temporal resolution is 30 min, 48 steps per day.

The model dynamics is governed by initial and boundary conditions. Climatological heat fluxes and wind stresses from the Oberhuber Atlas [23] are applied on the boundary at the sea-surface. The initial fields for the data assimilation experiments are prepared after 25 years of spin-up, since the model is first initialized at rest with the temperature-salinity structure from the Levitus Atlas [24].

Despite the fact that the GFDL model is global, no investigations have been done here beyond the tropical Atlantic zone. The model domain and buoys locations are shown in Fig. 1. Also, the two additional points mentioned above are displayed by the “cross” symbol.

Two kinds of experiments have been carried out. In the first kind, the observed data were assimilated day-by-day according to their realizations. In the second one, the inflow of real data was stopped after some time, and all other corrections have been done with respect to the previous information. This experiment is designed to address the memory the system has of its previous state.

Several parameters have to be defined before the practical realization of the method. First, the noise model should be set up. The covariance function of the noise at time n is defined as

$$R_n(r) = (N(n) - 1)^{-1} \left(\sum_{i=1}^{N(n)} \theta_i^2 \right) / (1 + r^2),$$

where θ_i is an error at point \bar{x}_i , r is a distance, and $N(n)$ is the number of measurements at this moment. Also, two radii are introduced. Let C_r be a cut-off radius. It means that any covariance between two points is prescribed to be zero if the distance between them exceeds C_r . And let h_r be a radius of homogeneity, i.e., within this radius, the error field can be considered as homogeneous, and it is possible to apply formula (14). In the experiments discussed below, these values are equaled to 10° and 3° , respectively.

In the first series of experiments, the model computes the temperature-salinity ocean state on the next day starting from the known initial temperature-salinity field after 25 years of spin-up. The real data are assimilated on the same day. Model predicted fields with assimilated observations create the new ocean state, which is taken as the initial conditions for the next day. This procedure is carried on for one month. For the next month the model starts again from the climate initial state.

Now, the results of the experiments will be discussed. Let $\sigma_m^2(t)$ be the variance of the model error without any assimilation, i.e., the error between model results and observations. Along with the variable $\sigma_m^2(t)$, two other variances $\sigma_b^2(t)$, and $\sigma_a^2(t)$, are introduced. They are variances before and after the correction, respectively. The variable $\sigma_b^2(t)$ shows the difference between the model

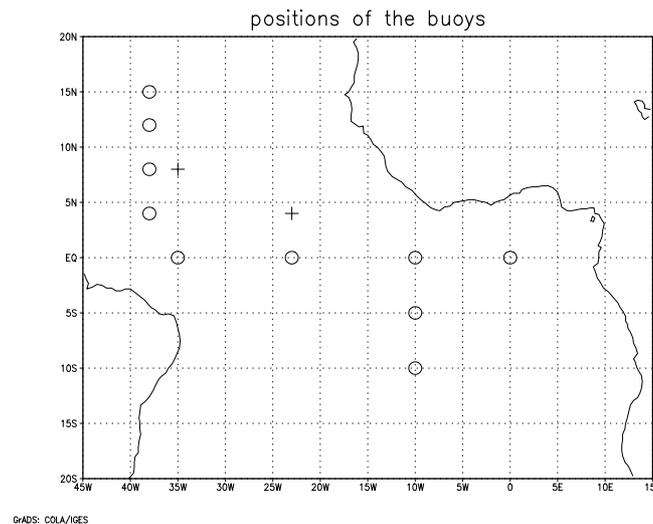


Fig. 1. Location of the buoys in the tropical Atlantic.

values and the observations at moment t , or at time-step n in the numerical realization, if the correction has been done in the previous time-step. And $\sigma_a^2(t)$ shows the difference after the correction at the same time-step. It is clear that the variable $\sigma_a^2(t)$ should be small, “close to zero” but not exactly zero, if the method works correctly.

As an example, Fig. 2(a) demonstrates the time behavior of these three variables during March 1999 at the 40 m level. The top curve shows the model error variance $\sigma_m^2(t)$. This curve defines a natural boundary of the error, which must decrease after manipulation with any technique aiming to approximate the model results to observations, e.g., by improving the boundary parameterization, initial conditions or model parameters. The variance remains almost steady during all the month, around $1.5\text{--}2^\circ C^2$. A little spike is observed in the end of the month due to the occurrence of a warm anomaly near the equator. The model could certainly not predict it, since it was forced with the climatologed data. Therefore, this spike was inevitable. The second curve demonstrates the time-behavior of the variance $\sigma_b^2(t)$. During the first four days, its values are even greater than $\sigma_m^2(t)$. It means that the model needs an adjustment time to come to an agreement with the observations. But after four days, this curve plunges under the first one and its values remain significantly smaller during all other days. Also, in the end of the month, a similar spike is observed, but the method reacts quickly and precisely. It forces the model to hold the same level of error, around $0.5^\circ C^2$. The last curve is $\sigma_a^2(t)$. Its values are almost constant, around $0.25^\circ C^2$ during all the period. This confirms that the method works correctly and it really assimilates data.

A similar picture is presented in Fig. 2(b). It shows the time behavior of the same variables, but averaged over the first 500 m depth. The errors become significantly greater, but their main features remain the same. The initial error is $6.8^\circ C^2$. This can be explained by both the ocean model deficiencies and by interpolation errors, since model and measured levels are not coincident except at the 40 m level. However, the method ignores the increasing of the initial error. The error variance of the method $\sigma_a^2(t)$ is approximately the same, and even the values of $\sigma_b^2(t)$ are not substantially different from their counterpart in Fig. 2(a).

This assimilation method has been compared with another one, a version of the Kalman-filter (Eq. (8)). One of the most common approaches has been used here [10,25,26]. The covariance is assumed to be very small in comparison with the variances, which are the diagonal elements in

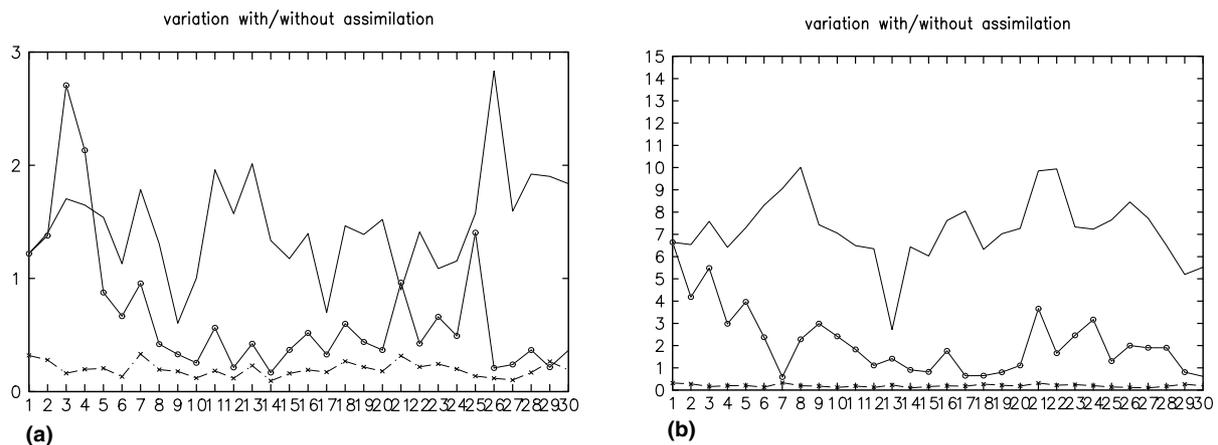


Fig. 2. (a) Time-behavior of the error variance at the 40 m level in March 1999. The solid line shows the variance of the model error without assimilation; the line marked with circles shows the error variance at time n if the correction has been done at time $n - 1$; the dashed-line shows the error variance after correction at the same time. (b) The same as in (a), but averaged through the top 500 m of the model domain.

Eq. (8). Therefore, Eq. (8) is rewritten only for the variances. Actually, the variance is found from the non-homogeneous transport equation

$$\frac{\partial \sigma^2(t, \bar{x})}{\partial t} = -(\nabla \sigma^2, U) + A \Delta \sigma^2 + B \frac{\partial^2 \sigma^2}{\partial z^2} + R(t), \quad (15)$$

where U is the three-dimensional velocity vector, A and B are the viscosity coefficients defined from the GFDL model, and the covariance function of the noise $R(t)$ was introduced before. As usual, symbol $\nabla = (\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z})$ is the gradient, Δ is the two-dimensional Laplace operator, and the symbol (\bar{a}, \bar{b}) represents the scalar product of vectors \bar{a} , \bar{b} . Eq. (15) is solved with the boundary and initial conditions for the variance equal to zero. After solving Eq. (15), the covariance between the two spatial points \bar{x} , \bar{y} is given by

$$K(t, \bar{x}, \bar{y}) = \sigma(t, \bar{x})\sigma(t, \bar{y}) \exp(-\lambda r^2), \quad (16)$$

where λ is a known parameter.

Figs. 3(a) and (b) show the comparison between the two methods. The results are presented at the 40 m level Fig. 3(a) and averaged over the first 500 m Fig. 3(b) for January 1999. As in Fig. 2, the top curve shows the model error regardless of observations. The model error grows during the month and accelerates from the day 19 on. The dashed-line is the error variance before assimilation with the Kalman-filter version, and the last curve with circles is the error variance of the method presented here. There are advantages to the latter. It gives smaller variance and faster adjustment after unexpected spikes. Fig. 3(a) shows this in the beginning of the month and around day 20. It should be pointed out that both assimilation methods work properly, but differently. Their different behaviors can be clearly seen in Fig. 3(b). In this figure, the two lines dashed and with circles are similar during first few days, but the variance of the Kalman-filter method is higher. Around day 15 they get closer. It means that the Kalman-filter needs more time to adjust to observations. Fig. 3(b) shows that the adjustment time increases with the depth, and this also confirms the correctness of the methods.

Fig. 4 shows the differences between the mean January 1999 temperature for the two methods at the 75 m level. Fig. 4(a) displays the mean January 1999 model temperature, and Fig. 4(b) the

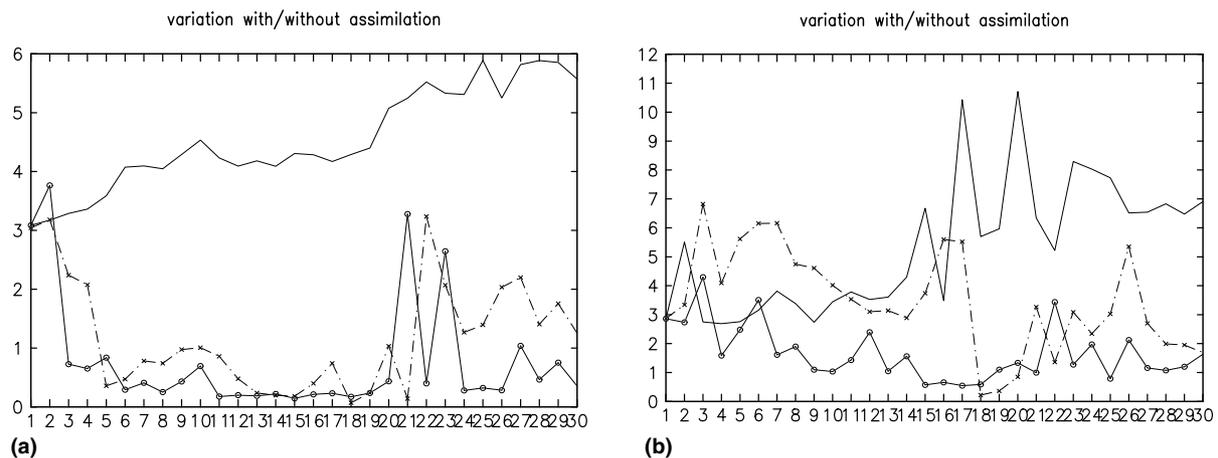


Fig. 3. (a) Comparison of the two methods at the 40 m depth during January 1999. The solid line presents the model error variance regardless of observations; the dashed-line shows the error variance at time n after the correction at time $n - 1$ with the Kalman-filter technique; and the line marked with circles illustrates the same variable with respect to the presented method. (b) The same as in (a), but averaged through the top 500 m of the model domain.

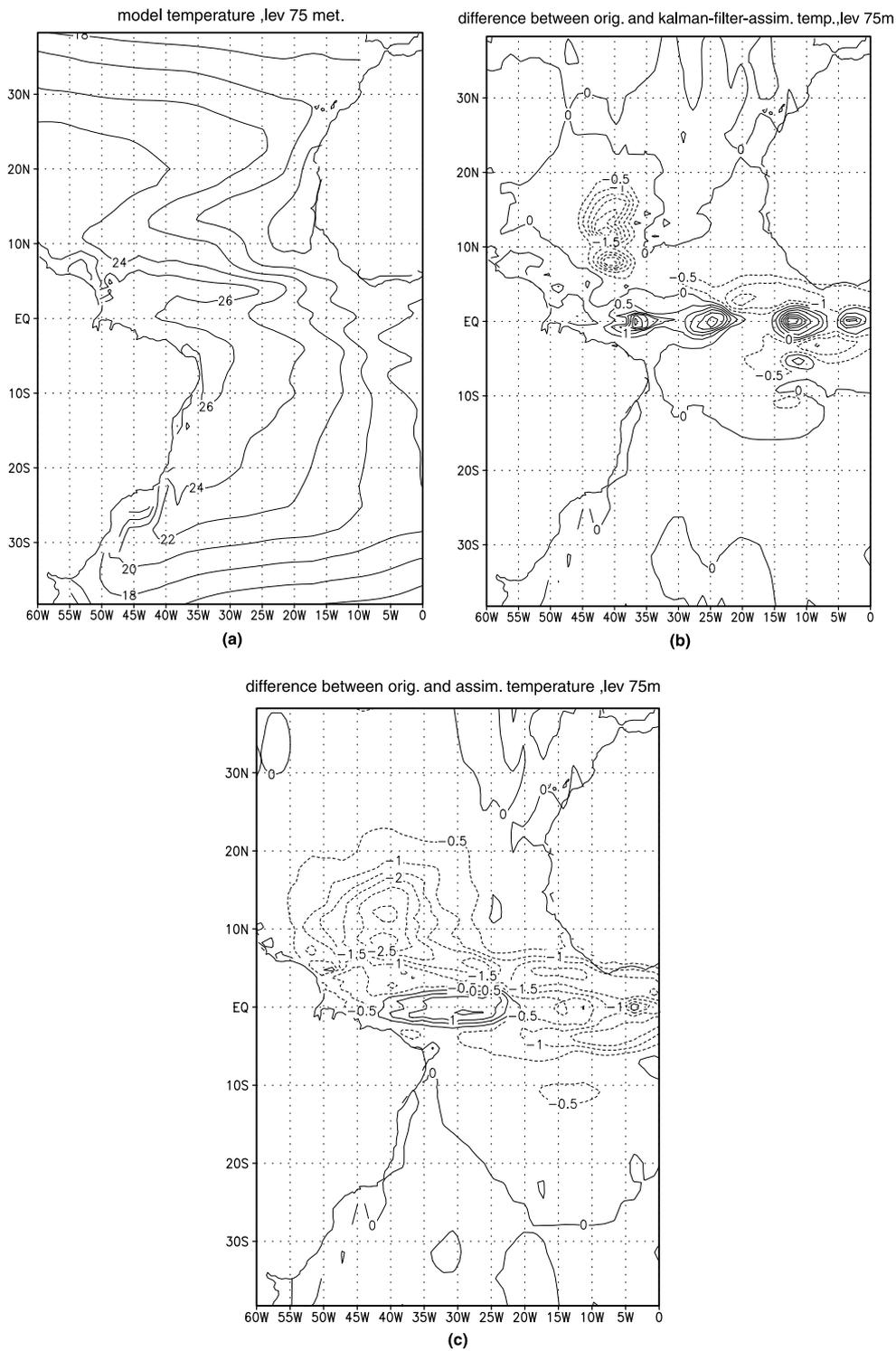


Fig. 4. (a) Monthly averaged temperature in the tropical Atlantic for January at the 75 m depth after 25-years of spin-up forced by climatological wind-stresses and heat fluxes. (b) Temperature difference between model results with Kalman-filter assimilation and model results shown in (a) for January 1999 at the 75 m depth. (c) The same as in (b), but for the presented method.

difference between the model temperature with and without assimilation with the Kalman-filter method. The latter shows that almost all changes are around the buoy locations (see Fig. 1). The differences are up to 3.5°C and have the order of the errors. This is clear because of formula (16) for the covariance. The real spatial dependence is ignored and replaced by the rapidly vanishing function away from the buoys. Note that as λ decreases in formula (16), a larger region of influence of the input data is imposed, but the covariance matrix becomes unstable. And within the reasonable range $10^{-8} \leq \lambda \leq 10^{-7} \text{ cm}^{-1}$ it weakly influences the pattern. The result of the presented method is shown in Fig. 4(c). This figure displays the difference between the model computations with and without assimilations with the discussed method. It produces smaller changes at the observational points with a maximum up to 2.5°C. Although the region of influence is larger it covers almost all the tropical zone. Also, Fig. 4(c) shows a highly non-homogeneous covariance. This is physically much more reasonable. Finally, one more detail deserves consideration. Fig. 4 shows that, even out of the cut-off radius, the correction may be different from zero. This occurs because of the feedback between currents and temperature-salinity conditions. The perturbations appeared due to data assimilation spread away from the buoy positions. This involves other dynamic phenomena. Both methods contain these characteristics, but it is much more evident in the present scheme.

Few more comments on a second series of experiments are discussed, since they may give an important information on the assimilation method. In this series, the flow of observational data was interrupted after some time, but the assimilation continued on along time according to Eq. (13). The results showed the maximum of the probability, solution of Eq. (13), became very small after 3–6 days, and made the covariance negligible. This means that the maximum period between data collection in the assimilation process should be around 3–6 days, to allow the model to keep memory of the data. This also limits the spatial region of influence because of the Lorenz relation $L \leq U_0 T$. If the order of the velocity is $\sim 100 \text{ cm/s}$, the Lorenz relation imposes a radius of influence of $\sim 600 \text{ km}$. But the influence due to wave propagation might be greater than this. The problem of connections between different ocean locations and the non-linear interaction among ocean variables is indeed an open research area.

5. Discussion and conclusions

The major advantage of the presented method is its ability to represent the real dependence in time and space according to Eq. (3). The FP equation reflects this dependence through the parameters a and B . But the diffusion approximation (10) is valid only under some restrictions. First, the time-interval between two sequential corrections should be significantly smaller than the time of integration. Second, the error variance should be limited. This condition can be expressed mathematically, but there is no need to do it here. Physically, it simply means that the model should not be too wrong or be completely unrealistic. These two conditions are physically reasonable.

Few points should be discussed on the practical application of the method. It is necessary to prove its stability with respect to perturbations on the data. Also, it is desirable to show its possible application to other research areas.

The stability of the method can be proved theoretically. Without loss of generality, only the one-dimensional case is considered. Let the drift coefficient $a(t, s)$ and the diffusion coefficient $b(t, s)$ be represented as

$$a(t, s) = a_0(t, s) + \mu a_1(t, s), \quad b(t, s) = b_0(t, s) + \mu b_1(t, s),$$

where $a_0(t, s)$, $a_1(t, s)$ ($b_0(t, s)$, $b_1(t, s)$) are the initial non-perturbed and perturbed drift (diffusion) coefficients, respectively, and μ is a small parameter. Using standard methods in the perturbation theory the solution of the FP equation with coefficients $a(t, s)$ and $b(t, s)$ is given by

$$p(t, s) = p_0(t, s) + \mu p_1(t, s),$$

where $p_0(t, s)$, $p_1(t, s)$ are satisfied by the equations

$$\begin{aligned} \frac{\partial p_0(t, s)}{\partial t} &= -\frac{\partial(a_0 p_0(t, s))}{\partial s} + \frac{1}{2} \frac{\partial^2(b_0 p_0(t, s))}{\partial s^2}, \\ \frac{\partial p_1(t, s)}{\partial t} &= -\frac{\partial(a_0 p_1(t, s))}{\partial s} + \frac{1}{2} \frac{\partial^2(b_0 p_1(t, s))}{\partial s^2} - \frac{\partial(a_1 p_0(t, s))}{\partial s}. \end{aligned} \quad (17)$$

Here, only the perturbation for the drift is considered, since the proof for diffusion is the same. Because $p_0(t, s)$ is a probability density, the last term in Eq. (17) is bounded in average, i.e.,

$$\text{abs} \left(\int_{-\infty}^{\infty} \frac{\partial(a_1 p_0)}{\partial s} ds \right) < \infty.$$

This gives the stability of the FP equation in average and hence, the stability of the method, in general. However, in practical applications, the question of stability is not so clear and should be tested numerically. This is an issue for further studies.

The method can be significantly improved by involving additional information such as salinity, velocities, chemical tracers, satellite observations, and so on. Of course, it will increase the computations but it will allow both an optimization of the structure of the observational array and take into account the physical connections in time and space among the different oceanic variables. This should be performed in future works.

Finally, few words about the possible application areas of the method are presented. In climatology, it can be used with long-time series to investigate the seasonal-to-interannual variability of the climate system. It can help to evaluate the heat fluxes, mass transport and vertical mixing. In weather and climate forecasts, it can be used to produce optimal initial condition to improve the predictability of ocean or coupled ocean–atmosphere models.

Acknowledgements

The authors would like to thank Dr. Ben Kirtman of the Center for Ocean-Land-Atmosphere Studies (COLA) for the invaluable help with the implementation of MOM_2; Mr. Brian Doty of COLA for the software GrADS. Also, we thank the Centro de Previsão de Tempo e Estudos Climáticos (CPTEC/INPE) for the computational facilities, and personally, its director, Dr. Carlos A. Nobre, for his support. This work has been partially supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP).

References

- [1] M. Ghil, P. Malanotte-Rizzoli, Data assimilation in meteorology and oceanography, *Adv. Geophys.* 33 (1991) 141–266.
- [2] V. Penenko, N. Obrastov, A variational initialization method for the field meteorological elements, *Meteorol. Hydrol.* 11 (1976) 1–11 (Russian version).
- [3] O. Talagrand, P. Courtier, Variational assimilation of meteorological observations with the adjoint vorticity. Equation I. Theory, *Q. J. R. Meteorol. Soc.* 113 (1987) 1311–1328.

- [4] I.M. Navon, et al., Variational data assimilation with an adiabatic version of the spectral model, *Mon. Weather Rev.* 120 (1992) 1433–1446.
- [5] L. Yu, J.J. O'Brien, Variational data assimilation for determining the seasonal net surface heat flux using a tropical Pacific Ocean model, *J. Phys. Oceanogr.* 25 (1995) 2319–2343.
- [6] A.H. Jazwinski, *Stochastic Processes and Filtering Theory*, Academic Press, New York, 1970.
- [7] B. Anderson, J. Moore, *Optimal Filtering*, in: *Inform. and Syst. Sci. Ser.*, Prentice-Hall, Englewood Cliffs, NJ, 1979.
- [8] M. Ghil, Meteorological data assimilation for oceanographers, Part 1. Description and theoretical framework, *Dyn. Atmos. Ocean* 13 (1989) 197–218.
- [9] R.N. Miller, M. Gane, A Kalman-filter analyses of sea level height in the tropical Pacific, *J. Phys. Oceanogr.* 19 (1989) 773–790.
- [10] G. Evensen, Using the extended Kalman-filter with a multilayer quasi-geostrophic ocean model, *J. Geophys. Res.* 97 (1991) 17905–17924.
- [11] G. Evensen, Advanced data assimilation for strongly nonlinear dynamic, *Mon. Weather Rev.* 125 (6) (1997) 1342–1354.
- [12] K. Belyaev, S. Meyers, J.J. O'Brien, Fokker-Planck equation application to data assimilation into the hydrodynamic models, *J. Math. Sci.* 99 (4) (2000) 1393–1403.
- [13] A. Kolmogorov, *Sammelband zur Statistische Theorie der Turbulenz*, Akademie, Berlin, 1958.
- [14] D. Mueller, Bispectra of sea-surface temperature anomalies, *J. Phys. Oceanogr.* 17 (1987) 26–36.
- [15] K. Bryan, A numerical method for the study of the circulation of the World Ocean, *J. Comput. Phys.* 4 (1969) 347–376.
- [16] R. Pakanovsky, MOM_2 documentation, Users Guide and Reference Manual, GFDL Ocean Group Tech. Rep., N3, NOAA/GFDL, 1995, Princeton, NJ.
- [17] B. Kirtman, Ocean Rossby wave dynamics and the ENSO period in a coupled model, *J. Climate* 10 (1996) 1690–1704.
- [18] J. Servain, et al., A pilot research moored array in the tropical Atlantic PIRATA, *Bull. Am. Meteorol. Soc.* 79 (1998) 2019–2031.
- [19] A. Yaglom, The introduction to the theory of stationnal random processes, *Adv. Math.* 5 (1953) 3–168 (Russian version).
- [20] D. Doob, *Stochastic Processes*, Academic Press, New York, 1956.
- [21] D. Peaceman, H. Rachford Jr., The numerical solution of parabolic and elliptic differential equations, *J. Soc. Ind. Appl. Math.* 3 (1955) 28–41.
- [22] G. Marchuk, *Methods of Numerical Mathematics*, Springer, Berlin, 1982.
- [23] J. Oberhuber, An atlas based on COADS data set: The budget of heat, bouyance and turbulent kinetic energy at the surface of the Global Ocean. MPI reprint, vol. 15, 1988.
- [24] S. Levitus, *Climatological Atlas of the World Ocean*, NOAA Prof. Pap., vol. 13, Rockville, ML, 1982.
- [25] I. Timchenko, *Dynamico-stochastic models and their application in oceanography*, Hydrometizdat, Leningrad, 1986 (in Russian).
- [26] K. Belyaev, et al., Experiment Newfaex 88. The heat-balance estimation in the North Atlantic frontal zone resulted from the numerical model with data assimilation, *Izv. AN. SSSR, Ser. Phys. Atmos. Ocean* 26 (1990) 67–77 (Russian version).