ISSN 0001-4370, Oceanology, 2012, Vol. 52, No. 5, pp. 593–603. © Pleiades Publishing, Inc., 2012.
Original Russian Text © K.P. Belyaev, C.A.S. Tanajura, N.P. Tuchkova, 2012, published in Okeanologiya, 2012, Vol. 52, No. 5, pp. 643–653.

MARINE PHYSICS

Comparison of Methods for Argo Drifters Data Assimilation into a Hydrodynamical Model of the Ocean

K. P. Belyaev[a], C. A. S. Tanajura[b], and N. P. Tuchkova[c]

[a] *Shirshov Institute of Oceanology, Russian Academy of Sciences, Moscow, Russia*
E-mail: kb@sail.msk.ru
[b] *Universidade Federal da Bahia, Salvador, Brasil*
E-mail: cast@ufba.br
[c] *Dorodnicyn Computing Center, Russian Academy of Sciences, Moscow, Russia*
E-mail: tuchkova@ccas.ru

Received August 5, 2011; in final form, February, 20, 2012

Abstract—Different data assimilation methods such as an extended Kalman filter, the optimal interpolation method, and a method based on the Fokker–Planck equation applications are considered. Data from the ARGO drifters are assimilated into the HYCOM shallow water model (University of Miami, USA). Throughout the study, the schemes and methods of parallel computations with an MPI library are used. The results of the computations with assimilations are compared between themselves and with independent observations. The method based on the Fokker–Planck equation and the extended Kalman filter are preferable because they give better results than the optimal interpolation scheme. The various model characteristics of the ocean, such as the heat content fields and others, are analyzed after the data assimilation.

DOI: 10.1134/S0001437012050025

INTRODUCTION

The problem of the assimilation of observational data in the hydrodynamical ocean circulation model belongs to the most urgent problems of present-day oceanology. The difficulty is in the optimal combining of the computational characteristics of a hydrodynamical model with the observational information. The problem of the assimilation emerged in meteorology in the early 1960s, and, presently, assimilation methods are an important component of numerical weather forecasting. Present-day oceanology needs further improvement of the methods and computational schemes for observational data assimilation in the context of the explosion of observational techniques and mathematical tools.

The main goal of any assimilation method is the optimal accounting for observational information in the equations of a model and in using it as a basis for correcting the model's calculations. The assimilation procedures are distinguished precisely by the specificity of the implementation of these ideas. From the viewpoint of physics, the problem concerns such a change in the model field and in the adjustment of the latter to observations resulting in the validity of the principle of least action. Having this in mind, one can subdivide all of the methods of data assimilation into two large groups. The first group of dynamical-stochastic assimilation techniques involves the estimation (optimal linear filter) of an unknown signal from the observations in the background of "noise" having specified stochastic characteristics. In this case, the principle of least action is valid at the minimum dispersion of the sought-for estimate (or filter). The second group of methods implies the searching for the optimal field as the minimum of a specified functional that comprises both the model calculations and the difference between the model field and the observations. The first group of methods involves, for instance, the Kalman–filter [12, 13], while the second one comprises the modern 3D-Var and 4D-Var schemes developed as a consequence of the theory of inverse problems [1, 6, 7].

In the present study, we compare the assimilation procedures from the first group. Figure 1 illustrates the geometry of the assimilation scheme based on the dynamical-stochastic ideology. Let us assume that the measurement is performed at the instant t_1 at the point of the coordinate plane A_2 and that the model's value at this point equals A, while the measured value is A_1. It is required to correct the model's value at point B_2 at the instant t_2 where the model's before-correction value is B but there are no measurements. To do this requires knowledge of the association of the points A and B, which depends on a multitude of factors. This dependence may be expressed by the model's equations, but, as a rule, the respective formulas are essentially unusable in the calculations. In the linear approximation, this association is determined as a covariance function between the points A_2 and B_2 or between the values A and B.

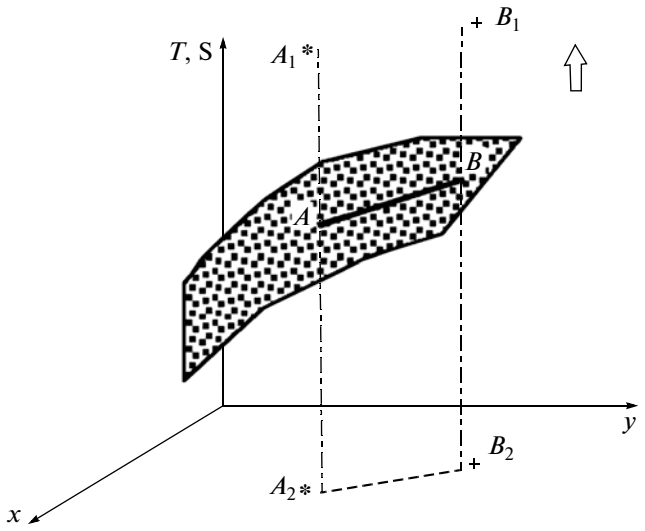

Fig. 1. Geometry of the general scheme of the data assimilation. See the text for details.

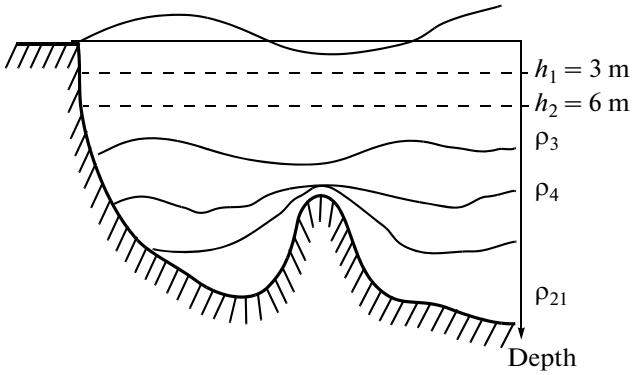

Fig. 2. Example of dividing the model's calculation domain into equal vertical layers of density ρ ($\rho_1 = 24.0$; $\rho_2 = 24.5$; $\rho_3 = 25.0$; ...).

Different procedures are being proposed for calculations of the covariance function. The simplest practically feasible approach used in meteorology in the 1960s is the optimal interpolation method, according to which the covariance was given as a known function of two points in the coordinate plain [5]. The chain lines in Fig. 1 show the relation between A_2 and B_2. In the Kalman-filter theory, the relation is given not between points A_2 and B_2 in the coordinate plain but in the hyperplane of the A and B values. To this end, a variety of calculations are performed using different initial conditions (an ensemble). Next, the respective values of A and B are found for each calculation, and the covariance between these values is computed from a standard expression for the sampled data covariance; i.e., one applies the Monte-Carlo method well known in statistics [10]. The bold line in Fig. 1 shows the relation between A and B. The procedure based on the theory of random diffusion processes and the Fokker–Plank equation implies the searching for the relation between the points A and B using the method of the generalized Kalman filter. However, in contrast to the preceding scheme, this dependence is expressed through the coefficients with the help of which one describes the "relief" of the model surface shown in Fig. 1. The determination of these coefficients and the necessary mathematical tools are described at length in a number of works [2, 3, 4, 8].

When analyzing the proposed assimilation procedure, one has to have in mind the advantages and disadvantages: a priori specifying the covariance function in the scheme of the objective analysis does not actually reflect the intricate physics of the interaction of the parameters in question. Therefore, the correction of a model calculation can result in strongly mutilated or simply wrong fields of oceanographic characteristics. At that, this scheme is easy to implement and its results are easy to interpret.

When using generalized Kalman filtering, the dependence between the physical quantities at points A and B is evaluated as reliably as the model itself reproduces these relations. However, the implementation of such an approach is difficult because it requires sophisticated calculations and higher computer power (memory, time for the computations, etc). Besides, the computational accuracy grows proportionally to $(\sqrt{P})^{-1}$, where P is the number of necessary operations. This means that a 10-fold increase in accuracy is achievable at the expense of a 100-fold increase in the number of calculations, which weights the calculation process.

The scheme based on the Fokker–Planck equations is easier to implement as compared to the generalized Kalman filter, although it uses the same type of relation between points A and B determined by the surface geometry. It is necessary, however, that the time gaps between the subsequent assimilations were short against the total integration time of the model. In addition, the hyperplane given by the model has to be fairly smooth involving no deep depressions and elevations. In practice, this means that the scheme is inapplicable to a model that describes an environment with sharply changing characteristics such as shock waves, strong separated jets, and the like.

In the present work, we use a modern version of the HYCOM hydrodynamical model (version 2.2 of the Hybrid Coordinate Ocean Model) [10]. The original description of this model is available in [9]. This is an isopycnal model in which the whole thickness of the ocean is subdivided into layers of constant density. The dynamics of the basic hydrophysical characteristics occur inside these layers (Fig. 2). At that, layers of fixed thickness are considered instead of constant density layers at the ocean's surface and in marginal shallow areas (that is why the model is referred to as a hybrid one). The complete equations of the ocean's dynamics describe the motion of these layers and their

configuration relative their position inclusive. The number of layers is to be set in advance and does not change in the process of the calculations. A detailed description of the model's version used can be found on the Internet at http://hycom.org/attachments/063_hycom_users_guide.pdf.

By now, there are a number of studies dedicated to the assimilation of observational data into the HYCOM model. Let us notice studies [10, 11, 14]. The expendable bathimetric thermograph data (XBT) were assimilated in [14] in accordance with the optimal interpolation scheme. It was expressly noticed in the above work that it is necessary to know both the temperature and salinity for the correct reproduction of the calculated field. Relatively recent studies [10, 11] have demonstrated considerable progress in this direction. Advantage was taken of the Kalman filter scheme when assimilating the altimetry, the XBT data, and the others.

The present work is dedicated to the assimilation of the ARGO drifters' data in the HYCOM model (version 2.2) and based on the above methods of assimilation. The ARGO drifters' data (involving profiles of the temperature, salinity, and pressure) were taken from the data base (www.argo.ucsd.edu) for 2008–2010. Prior to the assimilation in the model, the data were checked for their quality and rejected in the case of profiles with wrong coordinates, the wrong rated time of the data transmission, and with the fraction of unreliable data exceeding the tolerable level (60% in our research). The data of the observations that passed the quality check were assimilated in the model according to the computational schemes described in what follows.

DESCRIPTION OF THE MODEL AND THE DATA ASSIMILATION TECHNIQUES

As was noted in the Introduction, we used the Hybrid Circulation Ocean Model (HYCOM), which is the next generation of the Miami University Ocean Circulation Model of 1981 (MICOM) [9, 10]. Version 2.2.14 is used in the present study.

The scheme of breaking into the vertical layers is based on [11]. Use has been made of 21 vertical layers. The depth of the three upper layers is fixed, while the remaining 18 layers were chosen as layers of equal density ρ.

The whole model area is located between 78°S–55° N, 100° W–20° E. It covers the Caribbean Sea and the main part of the Atlantic Ocean. The spatial resolution is 0.25° in latitude and slightly varies around this value in longitude. The relaxation condition is set on the ocean's surface, and the values of the observed temperature are taken from the climatic atlas. More specifically, the boundary condition for the temperature on the ocean's surface is set in the model as $\frac{dT}{dz} = F + \kappa_{rlx}(T_o - T)$, where F are the fluxes of heat (explicit and latent), and T and T_o are the model and observable temperatures, respectively. The relaxation coefficient κ_{rlx} is set constant and equals 0.8 in the respective units. The relaxation of the temperature and salinity at the side boundaries is taken from the climatic atlas too. Moreover, the condition of the constancy of the fluxes is set for the velocities: zero flux through the northern boundary, a constant eastbound flux of 110 Sv in the region of the Drake Passage, and the westbound fluxes of 10 Sv each at 12 points of the grid lengthwise the 20° E line south of Southern Africa.

MPI libraries were used in the model for the paralleling. The studied region was subdivided into 64 subregions for parallel computing. The description of this part of our study is omitted since the details of the parallel computing are of minor importance for our paper.

The climatic atlas of Levitus (NOAA, http://www.noaa.gov/) for the values of the temperature and salinity at zero values of the velocity was chosen as the initial condition for the model's integration. Next, the model is integrated for 40 years under the forcing of data from the COADS archive (http://icoads.noaa.gov); i.e., the spin-up run is performed. After the spin-up run, every month from January to December for the last decade is used as the initial condition for creating the 20-year ensemble of the modeling. At that, ten members of this ensemble are created under the forcing of the atmospheric data taken from the NCEP archive (http://www.ncep.noaa.gov/) for 2007–2008 with 6 hour resolution. Another ten members of the ensemble are created under the forcing of the data from the Global Forecast System National Climatic Data Center (GFS) archive with 24 h resolution for the same period. The HYCOM model is forced by data on the wind, the precipitation, and the fluxes of heat and moisture at the ocean's surface with the data being correspondingly transformed into the boundary conditions for the equations of motion. The discharge of the main rivers of the Atlantic Ocean are taken into account too. These values of the ensemble are subsequently used for computing the statistics in the scheme of the Kalman filter.

The following algorithm underlies the assimilation schemes.

Let X be the vector of the model's state, i.e., the values of the model's variables at the grid points $X \in R^N$, and Y be the vector of the observations at $Y \in R^{N_{OBS}}$. Each of the elements of Y, Y_i, $i = 1$, N_{OBS} represents an r-dimensional vector because we consider the joint assimilation of several quantities, for instance, the temperature and salinity. The corrected model field (or the so-called analysis) is constructed according to the formula

$$X_a = X + K(Y - HX). \quad (1)$$

The designations in (1) are as follows: K is the unknown weighting matrix (the Kalman gain matrix), and HX is the horizontal projection of the model's state vector at the site of the observations with the projection being produced with the help of the operator H. If the dimensionality of the state vector (N) exceeds that of the observations vector (N_{OBS}), the redundant dimensionalities are believed to be equal to zero. The weighting matrix is found from the following equation:

$$BH' = K(HBH' + R). \quad (2)$$

Equation (2) comprises the matrix $B = E[(X - EX)(X - EX)']$, which represents a covariance matrix of the model's state, and the matrix R as a covariance matrix of the measurement errors. As usual, the symbol " ' " designates the transposition of a vector or a matrix, while the symbol EX indicates the mathematical expectation (the mean over the ensemble of states) of the random variable X. To avoid overloading of the text, we refrain from strict definitions of the respective quantities but instead provide constructive methods of the calculations.

The matrix R is supposed to be specified; it is believed that the measurement errors are mutually independent and that their dispersion is known. Thus, the matrix R represents a diagonal matrix whose diagonal is occupied by known quantities (dispersions of measurements errors) while all of the off-diagonal elements are equal to zero.

The determination of the covariance matrix B is the key problem of the assimilation. In the present study, we examine three assimilation schemes, each of which is distinctive in the method of the determination of this matrix.

1. Scheme of the Extended Kalman Filter (the Kalman Ensemble Filter (EnKF)). In this scheme, the parameterization of the matrix $B = [B_{ij}]$, $B \in R^{N \times N}$ is performed in the following way:

$$B = \frac{1}{M-1}\Phi\Phi' \quad (3)$$

$$\Phi = [\phi^1\phi^2 \ldots \phi^M], \quad \phi^k = \left(X^k - \frac{1}{M}\sum_{m=1}^{M} X^m \right).$$

Here, $X^k \in R^N$ is the implementation of the k-th member of the ensemble, which comprises in total M implementations: $k = 1, \ldots M$, $M = 20$. For this method, B depends on the number of grid points and on M. This scheme requires considerable computational resources because the number of grid points is of the order of 10^6 and, respectively, the size of the matrix is 10^{12}. In order to minimize the required computer memory, we computed this matrix in the following sequence: First, we calculated and stored the anomalies $\Phi = [\phi^1\phi^2 \ldots \phi^M]$ for every instant of the assimilation. Next, we serially computed the projection HBH' as $H\Phi$ and $\Phi'H'$, and, finally, we multiplied these matrices. Thanks to the fact that the projection substantially reduces the dimensionality, the described scheme allows us to calculate the matrix HBH' required in (2) but not the matrix B itself. The same idea is used to calculate the matrix BH'. In addition, the so-called truncation radius is introduced equaling ten grid points. This means that the covariance between any two quantities at a distance exceeding two truncation radii is supposed to be equal to zero. Experiments show that no substantial difference occurs if the radius is 15 points large, but, increasing it up to 100 points, results in numerical instability and difficulties concerning the covariance matrix inversion.

2. Optimal Interpolation (OI). According to this scheme, the matrix B is simply specified independently of the grid point and time. In the present work, B is specified as $B(x, y) = \sigma^2\exp(-\lambda d)$, where

$$\sigma^2 = \frac{\sum_{i=1}^{N_{OBS}} (HX_i - Y_i)^2}{N_{OBS}}.$$

The quantity d means the distance between the grid points and is measured in units of the grid. Therefore, the λ parameter is of inverse dimensionality. It was chosen according to the distance between the grid's nodes and equals 0.25.

3. The Hybrid Method Based on the Application of the Theory of Stochastic Diffusion Processes and the Fokker–Planck Equation (FP). This is an original method of authorship first published in [8]. According to this scheme, the covariance matrix is specified as

$$B_{ij} = \int_{-\infty}^{\infty} (u_i - Eu_i)(u_j - Eu_j)p(u_i, u_j)du_i du_j. \quad (4)$$

In (4), the quantity $u = (u_i, u_j) = (u(x), u(y))$ represents the value of the error at an arbitrary point of the domain x, y, and $p(u_i, u_j)$ is the probability density of this error. As was shown in [3], under certain conditions, this density satisfies the Fokker–Planck equation (Kolmogorov's second equation):

$$\frac{\partial p}{\partial t} = -\frac{\partial(Ap)}{\partial u} + \frac{1}{2}\frac{\partial^2(b^2 p)}{\partial u^2}. \quad (5)$$

In (5), the quantities A and b^2 (the drift vector and the matrix of diffusion) are determined from the values of the model and the observations. The procedure for their determination is described at length in [8] and is omitted here. It is worth noting that the present work involves a multiprocessor modification of this method based on the use of the MPI library.

Equation (5) is solved under the Somerfeld boundary conditions $p(t, \pm\infty) = 0$ and the initial condition $p(0, u) = f(u)$ for the specified $f(u)$. In practice, the initial condition is specified at points of the observations where the initial error is known. In other words, the initial function is $f(u) = \delta(u - u_i^0, u - u_j^0)$ if the values of the errors are u_i^0, u_j^0. In this way, the right side of (2),

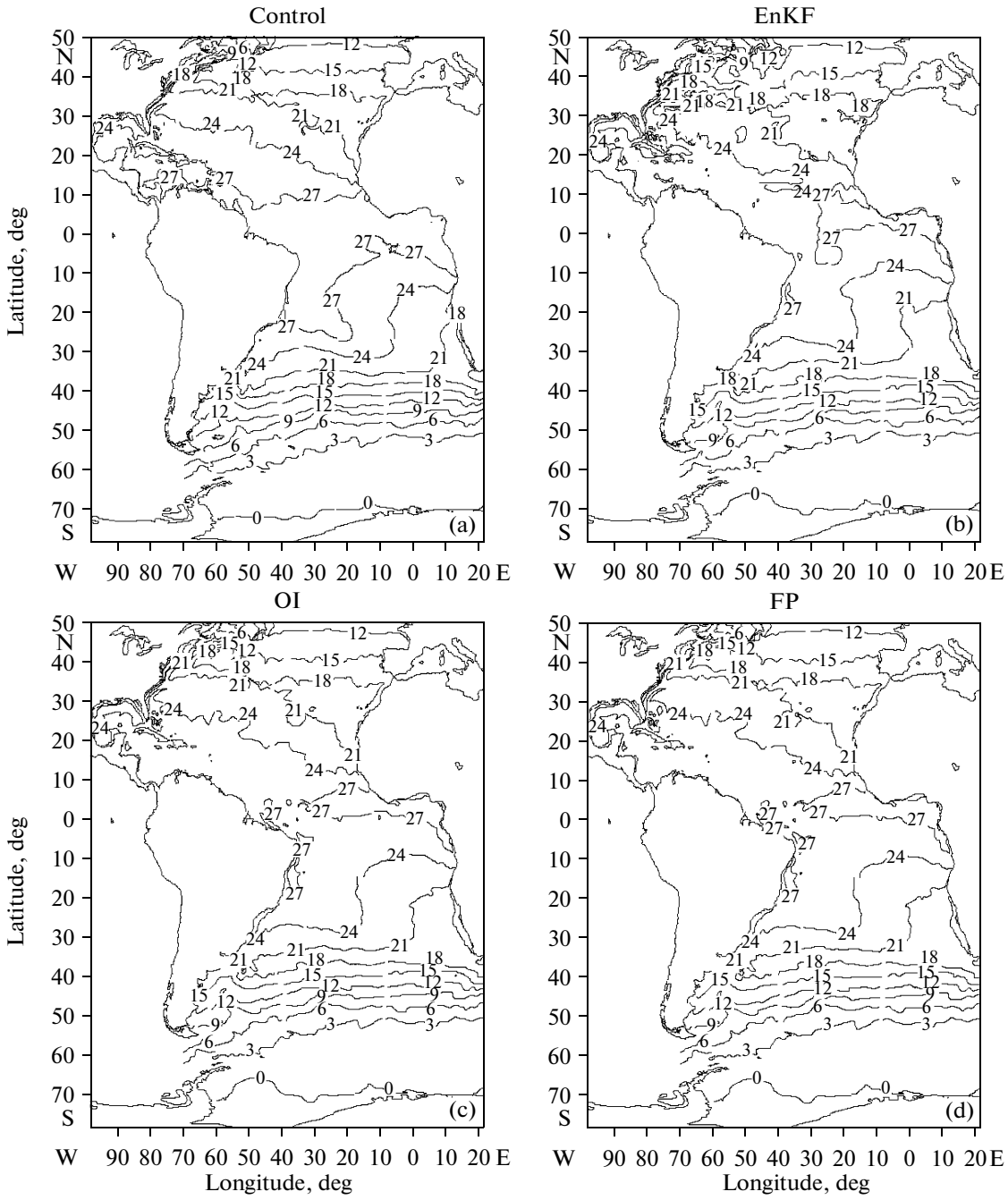

Fig. 3. SST (°C) for January 29, 2008. (a) control; (b) EnKF; (c) OI; (d) FP.

namely, the matrix ($HBH' + R$), is determined. In order to determine the left side of (2) of the matrix BH', it is necessary to specify the distribution of the initial error at an arbitrary point of the grid. It can be specified as a Gaussian one having a zero mean and dispersion equal to the average error's dispersion over the whole domain. This completely determines the error distribution for any instant of time and, consequently, the weighting matrix K.

It can be noticed that the OI scheme is a particular case of general Scheme 3. Indeed, if the drift vector and the diffusion matrix are constant, then the Gaussian distribution with a covariance function as in the OI scheme is the solution of equation (4).

These three assimilation schemes were applied jointly with the HYCOM model and the vertical profiles of the temperature and salinity from the ARGO drifters. In total, we used about 1800 profiles of the temperature and salinity (about 30 profiles per day on the average). After passing the aforementioned quality control procedures, the values of the temperature and salinity were interpolated to the middle of the respec-

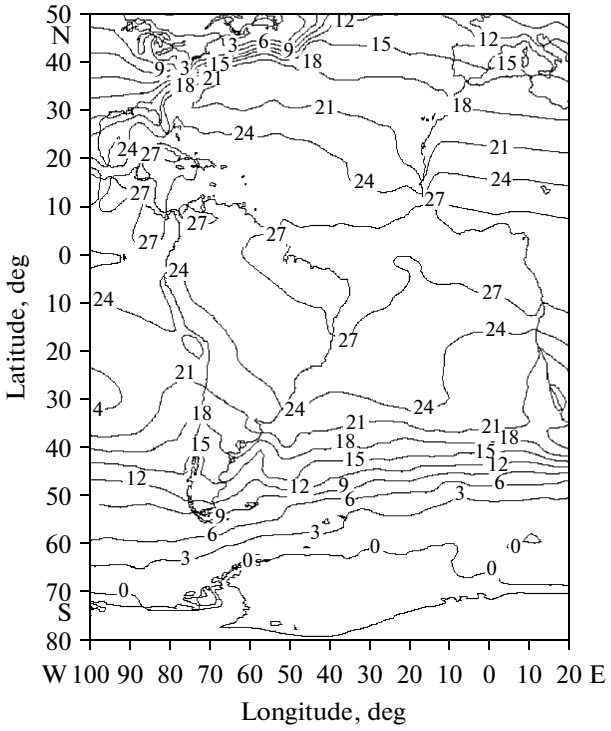

Fig. 4. SST (°C) from the Reynolds archive (Reynolds SST- http://www.nhc.noaa.gov/aboutsst.shtml) for January 29, 2008.

tive level determined by the model density and converted into the potential temperature relative to the surface. The difference $(Y_i - H_iX)$ for every station i was independently calculated for each level.

The experiments were performed for January 2008 and January 2010 with daily assimilation according to the above scheme. A control experiment was concurrently performed; i.e., the same model was integrated for the same period but without the assimilation.

CALCULATIONS RESULTS

Figure 3 demonstrates the calculation results in the following order: the control (Fig. 3a), the EnKF (Fig. 3b), the OI (Fig. 3c), and the FP (Fig. 3d). Independently, Fig. 4 displays the SST observed on the same date (from the Reynolds SST archive at http://www.nhc.noaa.gov/aboutsst.shtml).

The comparison of these figures shows that the control of the model overestimates the temperature magnitude in reference to the observations. It is easy to see that the 27°C isotherm passes far south in Fig. 3a as compared with Fig. 4. Moreover, it is evident that the localizations of the 21°C isotherm usually occurring west of Southern Africa and associated with the Benguela Current markedly differ in Fig. 3a and Fig. 4. All the calculations involving assimilation noticeably diminish the SST in the tropical domain and west of Southern Africa near 30° S. The correction makes the analysis closer to the Reynolds' SST in the main part of the calculation domain. However, the correction in the equatorial Atlantic and in the Caribbean Sea is too strong and the fields obtained are cooler than those on the Reynolds' SST maps. The strongest correction took place in the case of the EnKF method. As a result, the EnKF produces the maximal temperature gradients and the strongest spatial variability. On the contrary, the OI scheme results in minimal changes in the control field. There is virtually no synoptic variability, and the spatial gradients are weak in the resulting SST field. The FP method yields an intermediate outcome between the OI and EnKF schemes.

To facilitate the comparison of the assimilation methods, Fig. 5 displays the vertical distributions of the corrected model temperature field across the equator at the depths down to 500 m. All the calculations show a well expressed thermocline inclination and an east-bound decrease in the temperature. Nevertheless, the outcomes of the application of these methods exhibit noticeable differences. The EnKF method gives a more fuzzy thermocline with its noticeable heating and cooling in different regions at the depths below 200 m, which leads to evident horizontal temperature gradients. The OI scheme produces a smoother pattern close to the control calculations. The well-expressed thermocline to the west resulting from the control calculations takes place in the OI experiment too, but the isotherms below the 150 m depth level obtained during the control calculations exhibit oscillations characteristic of any assimilation method because the assimilation itself generates internal slowly attenuating waves. Strong oscillations of the isotherms below the 200 m depth and the upward bending of the 8, 10, and 12°C isotherms are typical of the FP method. The FP method's results appear more realistic relative to the EnKF calculations.

COMPARING THE QUALITY OF THE METHODS

The following characteristic is traditionally introduced to compare the assimilation methods: the average dispersion of the error of the forecast and the analysis relative to all the assimilation points and its behavior in time are considered. More precisely, let σ_s be the mean square deviation of the model's error relative to the observations, i.e., $\sigma_s = \sqrt{\frac{\sum_{i=1}^{N}(Y_i - H_iX)^2}{N-1}}$ (see above for all the designations). In addition to σ_s, two more quantities are introduced, namely, $\sigma_f = \sqrt{\frac{\sum_{i=1}^{N}(Y_i - H_iX_f)^2}{N-1}}$, and $\sigma_a = \sqrt{\frac{\sum_{i=1}^{N}(Y_i - H_iX_a)^2}{N-1}}$, where X_f, X_a are the model field

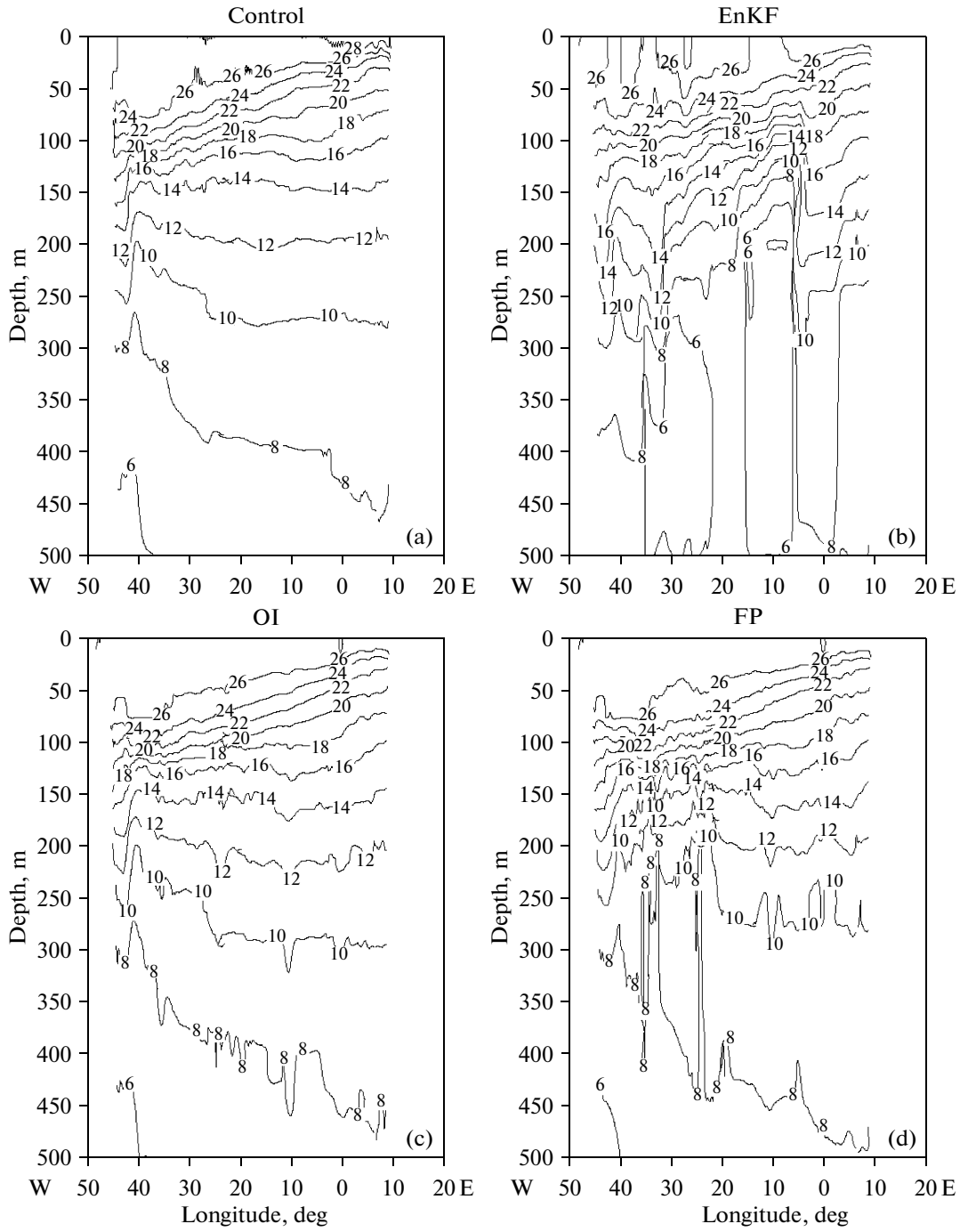

Fig. 5. Vertical temperature distribution (°C) lengthwise the equator for January 29, 2008. (a) control; (b) EnKF; (c) OI; (d) FP.

values after the start from the preceding assimilated value, respectively, before and after the next assimilation. In other words, the quantity σ_f determines the mean square deviation of the forecast error per one assimilation step, while the quantity σ_a reproduces the mean square deviation of the analysis error.

The time dependence of three variables σ_s, σ_f, and σ_a for the temperature dispersion is shown in Fig. 6 for every method in question: the EnKF (Fig. 6a), OI (Fig. 6b), and FP (Fig. 6c), respectively, during the period of January 1– 29, 2008. In Fig. 6, the upper curve describes the control value, while two the lower ones represent the quantities σ_f and σ_a, respectively. According to these figures, the EnKF method initially gives values exceeding the control estimates. This is due to the inertia, because the model does not adjust immediately to the data and for some time may yield a forecast of poorer quality than in the case of no assim-

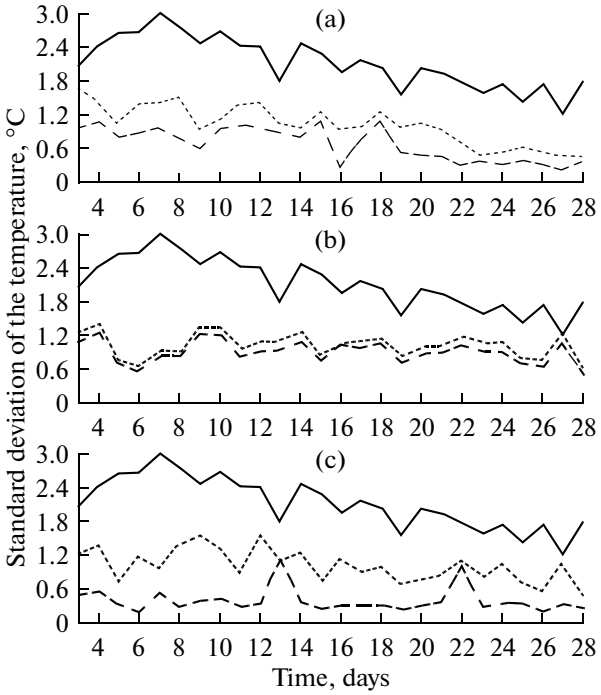

Fig. 6. Time-dependence of the standard deviation of the temperature at the 10 m depth level in January 2008. (a) EnKF; (b) OI; (c) FP. The upper curve shows the control results, while the dots and chain curves designate the one day forecasts and the analysis, respectively.

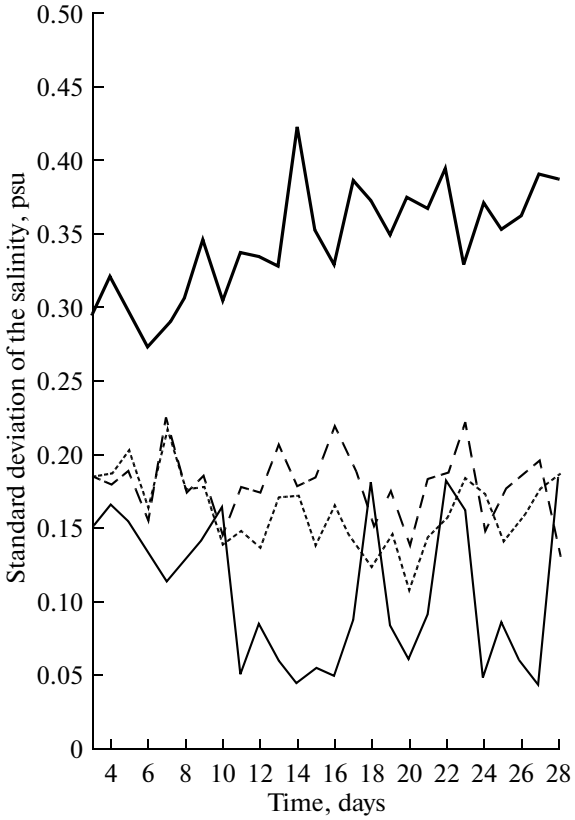

Fig. 7. Time dependence of the standard deviation of the salinity: the solid bold, chain, dot, and solid thin curves designate the control results, the OI error, the FP error, and the EnKF error, respectively.

ilation. Nevertheless, the forecast error diminishes from the second day, and the general trend of the lessening of this error with time becomes evident. At the end of the calculations, the value of the analysis error is as low as 0.8°C compared to the initial 1.5°C.

In the case of the EnKF method, the mean square value of the analysis error behaves in the same manner: starting from 1.0°C and finally dropping to 0.3°C. As for the OI method, its forecast error oscillates around values of 0.6 and 1.4°C. This error does not change in time: the amplitudes of the initial and final errors are virtually indistinguishable. The analysis error of the OI method is exactly the same in character as the forecast error; its absolute value is lower by 0.1°C. The forecast error of the FP method oscillates between 0.7 and 1.5°C during the first 15 days, but it diminishes by the end of the experiment. Among all the schemes examined, the FP method secures the lowest analysis error. It changed only slightly during the whole month but exhibited three evident peaks at 13, 22, and 29 days. This error makes up about 0.3°C not counting these spikes.

The salinity errors were examined in the same way; see Fig. 7. The latter does not show the mean square analysis error because it changes only slightly depending on the method.

For all of the three methods, the mean square forecast error is much lower in value than the estimates from the control calculations. It should be noticed that, in contrast to the temperature, the forecast error for the control calculation does not reduce but instead grows with time. This is due to the fact that the boundary condition on the surface for the temperature differs from that for the salinity, which was mentioned above when describing the model. The EnKF method leads to reduced forecasting errors and reaches the absolute minimum of 0.05 psu among all three methods. The OI method results in oscillations of the mean square forecast error around 0.18 psu. The FP method leads to almost monotonous diminishing of the forecast error with time during the first 20 days finally reaching a value of about 0.1 psu.

In any circulation model, the assimilation of the temperature and salinity results in changes in the fields of the density, the currents, and so on. Figure 8 demonstrates the changes in the subsurface equatorial countercurrent produced by the HYCOM model after assimilation. The control calculations (Fig. 8a) clearly indicate the west-bound surface flow lengthwise the equator and the subsurface flow near the South American continent. Also evident is the east-bound flow with its core at 36° W at the 120 m depth level. Any assimilation method results in upwelling of the zero

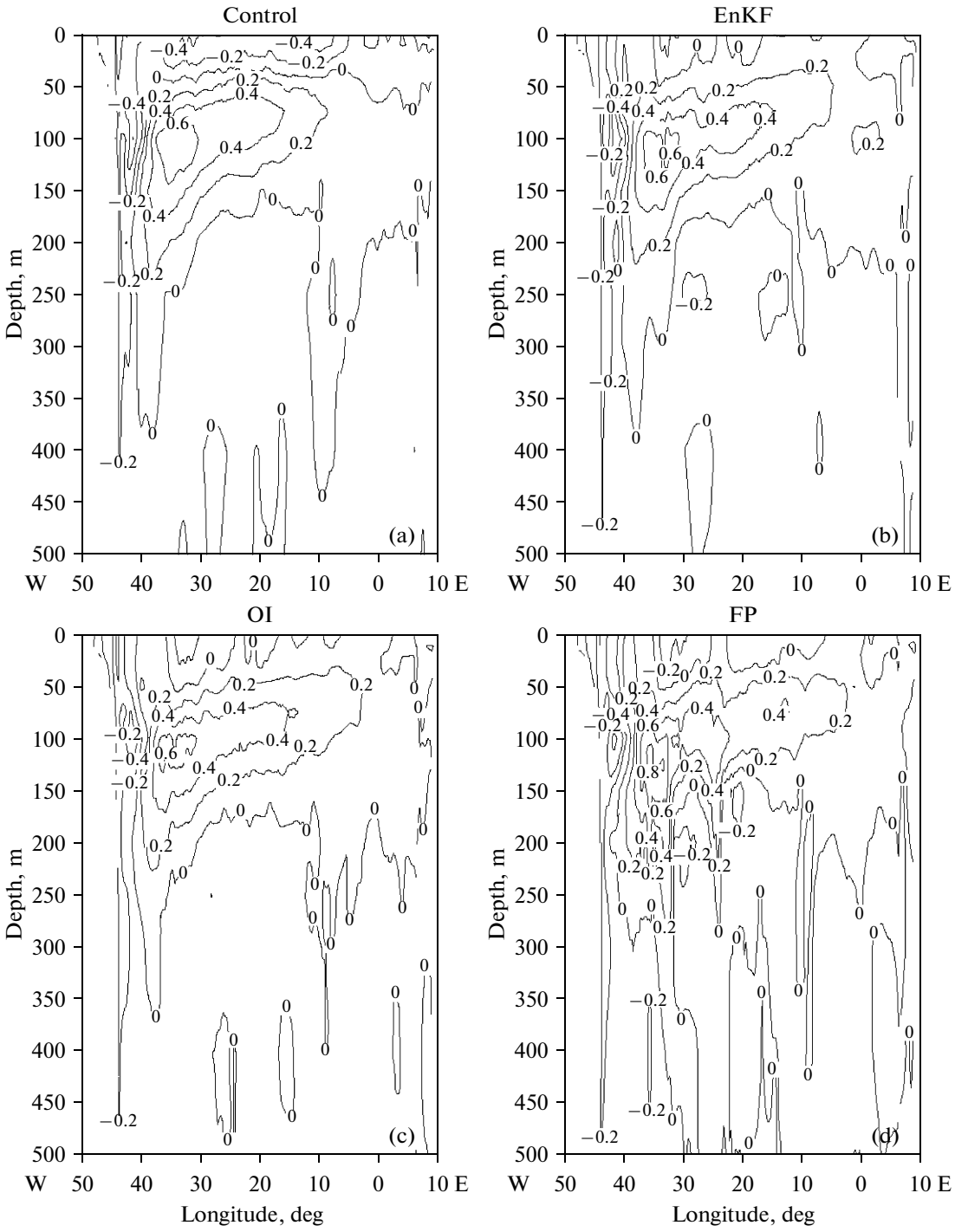

Fig. 8. Vertical distribution of the zonal velocity field (m/s) lengthwise the equator on January 29, 2008. (a) control; (b) EnKF; (c) OI; (d) FP.

velocity isoline. Therefore the west-bound flow markedly diminishes both in intensity and in area. The 0.2 m/s isoline in the subsurface current expands up to 5° to the east (Figs. 8b–8d); the assimilation of the temperature and salinity produces a weak east-bound flow in the region east of 5° W. The EnKF and OI methods are very close to each other in their results.

The surface velocities from the FP method exhibit minor differences too, but markedly stronger distinctions are characteristic of the subsurface current. For instance, this is easy to see in the isolines of 0 and 0.2 m/s, which are deepening in reference to the control at 25° W and at 35° W below the 300 m depth level. The core of the subsurface current with the 0.8 m/s

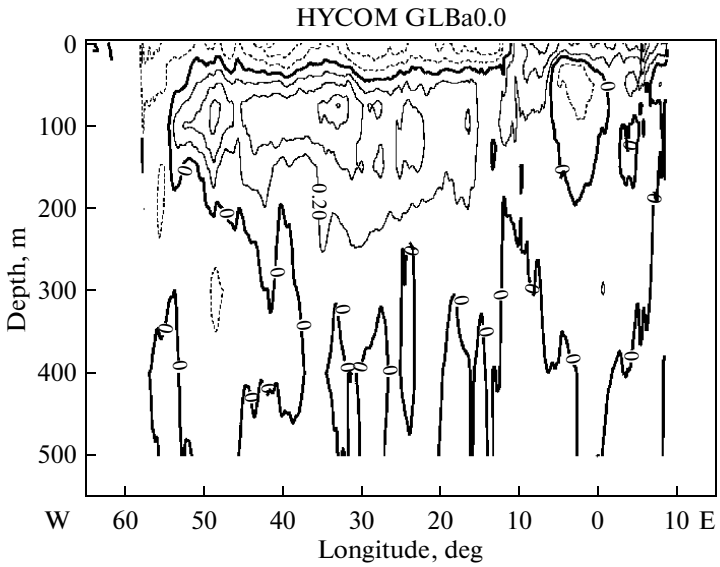

Fig. 9. Vertical distribution of the zonal current velocity across the equator from the HYCOM + NCODA (hittp://www.hycom.org).

maximum is easy to see in Fig. 8d. In total, the FP method produces markedly more intensive currents as compared to other assimilation schemes.

For comparison, Fig. 9 demonstrates the currents obtained at the Navy Supercomputer Research Center (NSRC) in Miami, USA from the global analysis of the HYCOM + NCODA system. The data were taken from the database www.hycom.org. The intensity of the subsurface core is almost the same as in the control calculations and in the assimilation experiment, but the core area itself is more extended in the NSRC calculations. The western surface current is well determined in the global analysis, but it does not cover the whole region, as was done in the control calculations; the east-bound flow in the east takes place in the global analysis. The weakening of the surface current and the surfacing of the zero isoline eastwards observed in the assimilation experiments demonstrates the coincidence of the surface currents' pattern with the global analysis. In the west, the control calculations and the EnKF and OI methods demonstrate different depth levels of the 0 m/s isoline from 200 to 500 m in reference to the global analysis. The pattern of the currents according to the FP method is closer to the NSRC currents than the remaining schemes.

CONCLUSIONS AND INFERENCES

Our study shows that different methods of data assimilation (the extended Kalman filter (EnKF), optimal interpolation (OI), and the schemes based on the application of the theory of random diffusion processes (FP)) produce different corrections of the model's state. However, by and large, all of the above schemes correct the model field in due direction, actually make it closer to the observational data and improve the consequent forecast when using the corrected field as the initial one. It can be noticed too that the OI correction is less significant, and the resulting field is closer to the control field as compared with the EnKF and/or FP methods.

The calculated fields of the temperature, salinity, and current's velocity constructed from the assimilation schemes demonstrate the increased sharpness of both the vertical and horizontal gradients of any quantities involved. The synoptic variability and oceanic turbulence become more evident. Quantitatively, the EnKF method gives variability 25% higher against the rest of the assimilation methods. At that, the FP calculations of the temperature and velocity better correspond to the observational data and to the data from the independent HYCOM + NCODA calculations (quantitative estimates are given in Figs. 4–7 and 9). The EnKF methods provide certain quantitative advantages in relation to the other methods (Fig. 8).

Taking into account these results, it is possible to infer that the FP and EnKF schemes deserve further development. Specifically, these schemes can be extended to the assimilation of altimetry, satellite SST data, and other available observational information. In addition, a reasonable balance is needed concerning the desired accuracy of the estimates, the computational sophistication, and the calculation efficiency.

ACKNOWLEDGMENTS

This work was supported by the FAPEX foundation, by the Agênda National do Petróleo of the Gás Natural e Biocombustíveis (ANP) of Brazil, and by the

Russian Foundation for Basic Research (project no. 11-07-00161-A).

REFERENCES

1. V. P. Agoshkov, *The Methods of Optimal Control and Adjoin Equation in the Mathematical Physics* (Inst. Computat. Modeling, Moscow, 2003) [in Russian].
2. K. P. Belyaev, V. N. Solov'ev, and C. A. Tanajura, "The Method of Oceanographic Data Assimilation in a Joined "Ocean-Atmosphere" Model System and Its Application to the Analysis of Ocean in the Tropical Atlantic," Okeanologiya **44**, No. 1, 71–81 (2004).
3. K. P. Belyaev, N. P. Tuchkova, and I. Kirkhner, "The Method of Model Output Correction by Measurements of Diffusion Approximation and Its Application to the Analysis of Hydrophysical Parameters," Mat. Model. **21**, No. 3, 53–68 (2009).
4. K. P. Belyaev, N. P. Tuchkova, and U. Kubash, "Reaction of Joint "Ocean-Atmosphere" Model on the Data Assimilation in the Tropical Pacific," Okeanologiya **50**, No. 3, 306–316 (2010).
5. L. S. Gandin, *Objective Analysis of Hydrometeorological Fields* (Gidrometizdat, Leningrad, 1963) [in Russian].
6. G. I. Marchuk, *Numerical Task Solution of the Dynamics of Atmosphere and the Ocean* (Gidrometizdat, Leningrad, 1974) [in Russian].
7. E. Andersson , J. Haseler, and P. Undén, "The ECMWF Implementation of Three Dimensional Variational Assimilation (3D-Var)," Q. J. R. Meteorol. Soc. **98**, 1100–1123 (1998).
8. K. Belyaev, C. A. S. Tanajura, and J. J. O'Brien, "A Data Assimilation Method Used with an Ocean Circulation Model and Its application to the Tropical Atlantic," Applied Math. Modeling **25**, No. 8, 655–670 (2001).
9. R. Bleck and D. B. Boudra, "Initial Testing of a Numerical Ocean Circulation Model Using a Hybrid Quasi Isopycnal Vertical Coordinate," J. Phys. Oceanogr. **11**, 755–770 (1981).
10. F. Counillon and L. Bertino, "High-Resolution Ensemble Forecasting for the Gulf of Mexico Eddies and Fronts," Ocean Dyn. **59**, 83–95 (2009).
11. J. A. Cummings, "Operational Multivariate Ocean Data Assimilation," Quart. J. Roy. Meteor. Soc. **131C**, 3583–3604 (2005).
12. G. Evensen, "Using the Extended Kalman-Filter with a Multilayer Quasi-Geostrophic Ocean Model," J. Geophys. Res. **97**, No. C11, 17905–17924 (1991).
13. R. Kalman, "A New Approach to Linear Filtering and Prediction Problems," J. Basic Engineering **82**, 35–45 (1960).
14. G. Evensen, "Using the Extended Kalman-Filter with a Multilayer Quasi-Geostrophic Ocean Model," J. Geophys. Res. **97**, 17905–17924 (1991).
15. W. C. Thacker and O. E. Esenkov, "Assimilating XBT Data into HYCOM," J. Atmos. Ocean. Tech. **19**, 709–724 (2002).