

Evaluating Effects of Observational Data Assimilation in General Ocean Circulation Model by Ensemble Kalman Filtering: Numerical Experiments with Synthetic Observations

V. N. Stepanov^{a*}, Yu. D. Resnyanskii^a, B. S. Strukov^a,
and A. A. Zelen'ko^a

^a*Hydrometeorological Research Center of the Russian Federation,
Bolshoi Predtechenskii per. 11–13, Moscow, 123242 Russia*

**e-mail: vlnst@hotmail.co.uk*

Received May 28, 2020

Revised September 9, 2020

Accepted October 6, 2020

Abstract—The quality of simulation of model fields is analyzed depending on the assimilation of various types of data using the PDAF software product assimilating synthetic data into the NEMO global ocean model. Several numerical experiments are performed to simulate the ocean–sea ice system. Initially, free model was run with different values of the coefficients of horizontal turbulent viscosity and diffusion, but with the same atmospheric forcing. The model output obtained with higher values of these coefficients was used to determine the first guess fields in subsequent experiments with data assimilation, while the model results with lower values of the coefficients were assumed to be true states, and a part of these results was used as synthetic observations. The results are analyzed that are assimilation of various types of observational data using the Kalman filter included through the PDAF to the NEMO model with real bottom topography. It is shown that a degree of improving model fields in the process of data assimilation is highly dependent on the structure of data at the input of the assimilation procedure.

DOI: 10.3103/S1068373921020047

Keywords: NEMO model, PDAF software product, ensemble Kalman filter, data assimilation

1. INTRODUCTION

Ocean circulation models along with atmospheric hydrodynamic models or coupled with them are the main tool for predicting the climate system state on the timescales from several days to months and centuries. However, uncertainties both in formulating and parameterizing physical processes and in specifying initial and boundary conditions lead to inaccuracies in the simulation of hydrophysical fields. The ability of the model to predict changes in the ocean state at not too large timescales significantly depends on initial conditions. Preparing such conditions is the task of data assimilation, which is an important component of operational oceanography systems [1–3, 5, 6].

The authors of [22] present the system for data assimilation using the Kalman filter implemented in the PDAF (Parallel Data Assimilation Framework) software product (also see [21], <http://pdaf.awi.de>), that can be used together with the NEMO ocean circulation model. The authors of this system showed that the incorporation of the PDAF leads only to a small reduction of the speedup of the NEMO model itself. The authors of [17, 25] successfully used the PDAF jointly with the idealized NEMO configuration corresponding to a double ocean gyre in a rectangular basin. Here, the results of the PDAF application to the global ocean model with real bottom topography are presented.

The NEMO (as well as the Kalman filter) is used by many authors to solve various problems with the same parameterizations for similar spatial resolutions. The present paper assumes that one of the modeling states obtained with low dissipation coefficients is “true” and is compared with the model values obtained

with high dissipation coefficients; it is assessed how the used algorithm approximates a general state of the simulated system to the “real” one.

The objective of the present paper is to analyze the results of assimilation of observational data of various types using the LETKF (Local Ensemble Transform Kalman Filter) [16]. It is included through the PDAF to the NEMO model with real bottom topography, which has not been considered in the previous publications. The problem statement with the brief description of the NEMO is given in Section 2. Section 3 presents the results of two numerical experiments with free model runs, whose values were subsequently used to form a set of synthetic observations and to assess data simulation efficiency. Section 4 briefly describes the assimilation technique. In Section 5 we compare the results of assimilation of different synthetic oceanographic fields with LETKF filter. Section 6 presents conclusions.

2. PROBLEM STATEMENT

The NEMO 4.0 ocean model [20] was used as a prognostic model in the data assimilation procedure. The model is based on primitive equations describing ocean hydrothermodynamics with a free level, coupled to the SI3 thermodynamic sea-ice model [24] with elastic-viscous-plastic rheology [26]. The simulations with the ORCA1 model configuration were performed. In this configuration, the calculations are carried out on the tripolar grid with a horizontal resolution of $1^\circ \times 1^\circ$ in the mid-latitudes, with a decrease in the horizontal resolution along the latitude in the subequatorial zone to $1/3^\circ$ (≈ 37 km) at the equator and with a specific location of grid points (that differs from the latitude-longitude one) in the northern polar region (a typical horizontal resolution here is ≈ 50 km). The model has 75 nonuniform vertical levels. The ice cover at the initial time moment was assumed to be present in the areas where sea surface temperature (SST) according to WOA13 data [19, 27] did not exceed 0°C , with the initial ice concentration equal to 90%. The ice thickness in this initial distribution was assumed equal to 3 m in the Northern Hemisphere and 1 m in the Southern Hemisphere, and the snow depth in both hemispheres was 0.3 m. The more detailed description of the model configuration is given in [8, 9].

The preparation for the assimilation included two numerical experiments on the simulation of the ocean–sea ice system (hereinafter, the experiments E1 and E2) without data assimilation performed with the same atmospheric forcing (data from the DFS5.2 dataset [11] for 1996), but with different values of coefficients of horizontal turbulent viscosity and diffusion (the similar technique was applied in [7]). Model data obtained with higher values of the coefficients (in the experiment E2) were used as first guess fields in subsequent experiments with data assimilation, while model data with lower values of the coefficients (in the experiment E1) were interpreted as a true state, and a part of these data was used to generate synthetic observations, which are used at the input of the assimilation procedure. The values of the coefficients of horizontal turbulent viscosity and diffusion used by the Laplace operator in the experiment E2 were specified 1.5 and 2 times higher than those in the experiment E1 (the coefficients of horizontal turbulent viscosity for the E1 and E2 were assumed equal to 1700 and 2500 m^2/s , the diffusion coefficients were 700 and 1500 m^2/s , respectively). Both experiments also took into account the effects of subgrid eddy-induced turbulence using the algorithm of [13], whose influence depends on the slope of isopycnic surfaces and Rossby deformation radius. A degree of the eddy effects is controlled by the eddy coefficient, that was 3 times higher in the E2 than in the E1 (1000 and 3000 m^2/s , respectively, for the E1 and E2). The vertical turbulent mixing was computed using one of the NEMO schemes, namely, TKE Turbulent Closure Scheme [20], where the coefficients of vertical turbulent viscosity and diffusion are assumed proportional to kinetic energy of turbulence. Changes in this energy were evaluated from the prognostic equation, where they are determined by the generation due to the vertical velocity shear, the suppressing effect of the stable density stratification (characterized by the Brunt–Vaisala frequency), vertical diffusion and dissipation [4]. When solving the equation, the parameterization of characteristic scales of mixing length and dissipation was used (see [20]).

Each of the two experiments included two integration stages. The experiments differed, as noted above, in the model values of the coefficients of turbulent viscosity and diffusion. At the first stage, the run started at rest, with January climatological temperatures and salinities from the WOA13 atlas and was carried out for the yearly time period: January 1–December 31, 1996 with the atmospheric forcing from the DFS5.2 dataset [11]. The turbulent components of the total flux of heat, moisture, and momentum (wind stress) were calculated based on these data using the bulk formulas and the CORE (Coordinated Ocean Research Experiments) algorithm [18], where the model SST, that imitates the atmosphere–ocean interaction, was also taken. At the second stage, the integration was repeated with the same atmospheric forcings for January 1–December 31, 1996, but from initial conditions obtained at the end of the yearly integration at the

first stage (December 31, 1996). Such two-stage integration was performed to overcome the adjustment procedure during the ocean state adaptation to the external atmospheric forcing.

It was assumed that the model fields obtained in the experiment E1 indicate the true state, from which synthetic observations were subsequently extracted. Observational data created in such way were used at the input of the data assimilation procedure, where the model configuration used in the experiment E2 was used as a computing core to obtain the first guess field.

Data of these two experiments obtained as a result of simulations were used for the intercomparison, and data from the experiment E1 were subsequently compared to the simulations, where different types of synthetic observations were assimilated.

3. PREPARATORY EXPERIMENTS FOR DATA ASSIMILATION E1 AND E2

All model outputs are archived as successive five-day means throughout the entire model simulation for 1996. In the experiments with data assimilation, the model fields obtained at the analysis step with corresponding analysis increments were also stored.

The results of simulations in the preparatory experiments, where free model runs without data assimilation were performed, demonstrate that kinetic energy averaged over the whole World Ocean (KE calculated as $0.5(U^2 + V^2)$, where U and V are the five-day averages of the horizontal velocity components) in the E2 is 17% smaller than in the E1. The greater contribution to this reduction is made by the Southern Hemisphere (KE in the E2 is ~23% smaller than in the E1). The value of KE on the ocean surface changed insignificantly: approximately by 6%, which is explained by the predominant influence on the KE of the surface layer forced by the same data on wind speed in both experiments. Nevertheless, differences in hydrological fields in the water column between the two experiments are essential.

A common pattern of the geographic distribution of such differences is that their highest values are observed in the areas of strong jet currents. Maximum differences (the deviations of the E2 from the E1) in water temperature (to ~5 °C) were registered on the ocean surface both in the Southern Hemisphere (in the northern branch of the Antarctic Circumpolar Current (the negative deviations are obtained here), in the southern branch of the Brazilian Current and the area, where the Antarctic Circumpolar Current meets the Agulhas Current (Agulhas retroflexion region), and in the northern branch of the East Australian Current (positive deviations are mainly reported here)) and in the Northern Hemisphere (in the areas of the Gulf Stream and Kuroshio, where both positive and negative deviations were seen). The positive deviations of SST (~2 °C) are also observed in the areas of Pacific trade winds. They are observed at large depths: they reach ~4 °C at the horizon of 500 m and ~2 °C at the horizon of 1000 m. In the Southern Hemisphere at the depth of 1000 m, these deviations are basically positive; in the Northern Hemisphere, the deviations are negative and are situated in the area of the cold Canary Current.

Maximum differences in salinity on the ocean surface between the experiments E2 and E1 are also registered in coastal regions (especially in the Arctic). At large depths, the location of the regions with significant differences in salinity between these two experiments coincides with the areas of maximum temperature differences. At the horizon of 500 m, the salinity differences reach ~0.5 psu (practical salinity unit). At the depth of 1000 m, the value of positive deviations of the E2 from the E1 is small: ~0.1 psu, but about -0.4 psu for the negative deviations of the E2 from the E1 in the area of the Canary Current.

Due to the changes in the density field caused by the temperature and salinity differences, the changes in the sea surface level also occurred due to steric effects (Fig. 1a). In the Southern Ocean, the differences in of both signs between the E2 and E1 exceed 30 cm. In the Northern Hemisphere, the differences in are smaller: the positive differences are equal to about 10 and 15 cm in the area of the Gulf Stream and Kuroshio, respectively, where positive deviations are observed. In the zone of trade winds, positive deviations in are about 5–7 cm.

4. DATA ASSIMILATION METHOD

Local ensemble transform Kalman filter (LETKF). The ensemble-based data assimilation is a natural approach to the assessment of uncertainty in model forecasts and to the improvement of the initial state. The present study utilizes a widely used Kalman filter LETKF [16]. The ensemble members are evolved in accordance with the model equations (forecast step) during a certain time interval called an assimilation window. Then, the deviations of observational data accumulated during this interval from the state of the model variables called first guess fields are calculated. The deviations obtained in such way (at the analysis step)

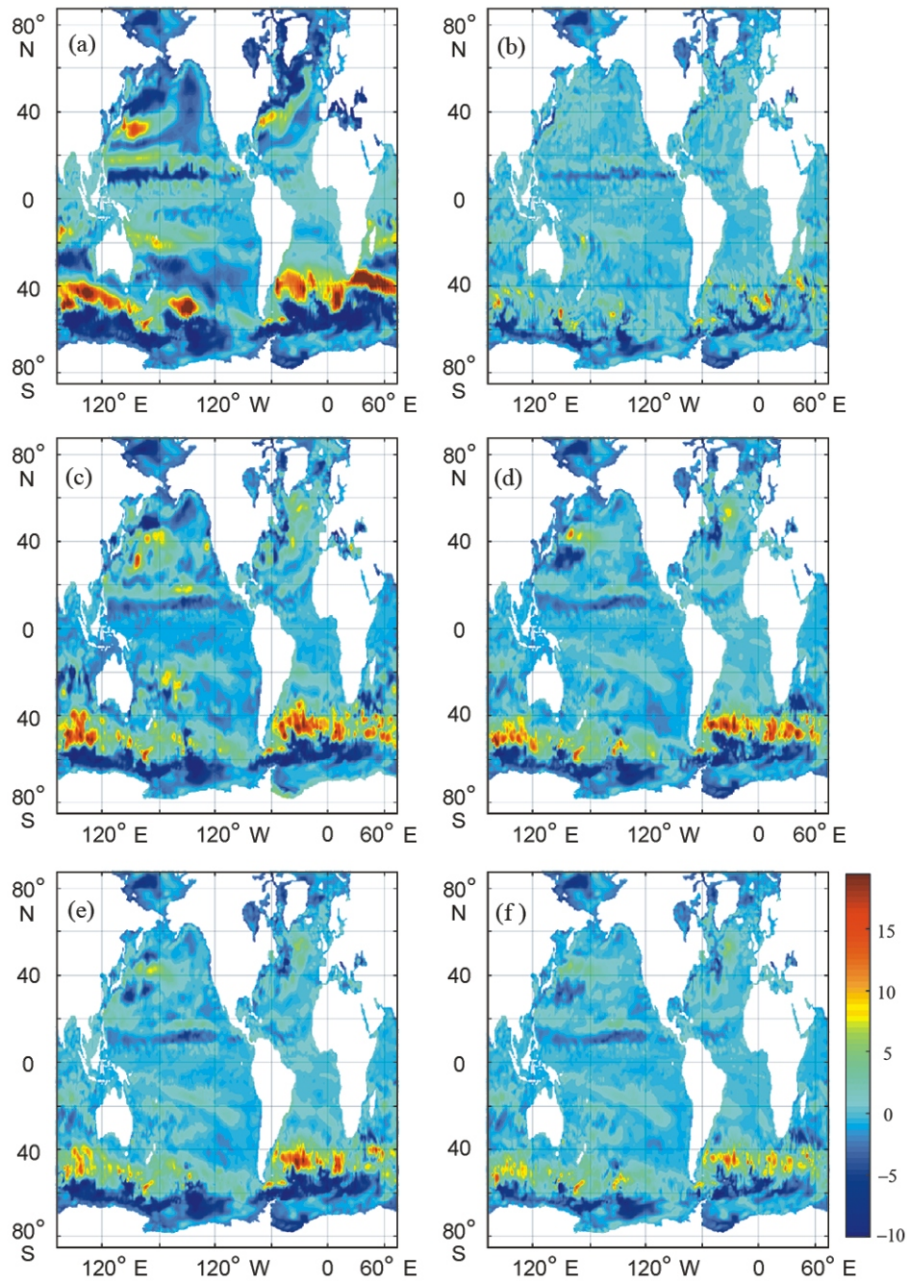


Fig. 1. The mean differences between the results of simulation of the sea level (cm) in the (a) E2, (b) ASSH, (c) AT, (d) AT_S, (e) AT_S_SSH, (f) AT_S_SSH_25 experiments and the data of the E1 experiment for the values averaged over December 1996.

are used to compute corrections to the first guess fields providing the optimum estimate of the system state at the current moment.

The task of the local ensemble Kalman filter is the computation of the ensemble of analyses consisting of N realizations $\{\mathbf{x}^{a(i)}, i = 1, 2, \dots, N\}$ based on the information obtained at the forecast step (from the first guess) $\{\mathbf{x}^{b(i)}, i = 1, 2, \dots, N\}$, as well as based on the information contained in observational data (the observation vector \mathbf{Y}) compared with the model simulations through the use of the observation operator $H(x)$. According to LETKF, the i th ensemble member of the analysis can be calculated as follows:

$$\mathbf{x}^{a(i)} = \mathbf{x}^a + \mathbf{X}^{a(i)} \quad (1)$$

where the ensemble mean value of the analysis $\langle \mathbf{x}^a \rangle$ is computed using the formula

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{X}^b [(N-1)\mathbf{I} + (\mathbf{Y}^b)^T \mathbf{R}^{-1} \mathbf{Y}^b]^{-1} (\mathbf{Y}^b)^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{y}^b) \quad (2)$$

and $\mathbf{X}^{a(i)}$ is the i th column of the matrix \mathbf{X}^a :

$$\mathbf{X}^a = \mathbf{X}^b ((N-1)[(N-1)\mathbf{I} + (\mathbf{Y}^b)^T \mathbf{R}^{-1} \mathbf{Y}^b]^{-1})^{1/2}, \quad (3)$$

\mathbf{R} is the diagonal covariance matrix of observation errors.

Here, \mathbf{I} is the identity matrix; $\alpha < 1$ is the forgetting factor [23], that performs multiplicative inflation of the ensemble; \mathbf{X}^b is the matrix of deviations from the mean of the first guesses; \mathbf{Y}^b is the matrix, whose i th column is calculated as $\mathbf{Y}^{b(i)} = H(\mathbf{x}^{b(i)}) - \langle \mathbf{y}^b \rangle$; $\langle \mathbf{y}^b \rangle$ is the average over i of $H(\mathbf{x}^{b(i)})$, $i = 1, \dots, N$.

To damp long distance covariances a localization of covariance matrices \mathbf{R} in ensemble-based Kalman filters is required for data assimilation into large-scale models [14, 15].

Following [16], the calculations in this realization of the filter are carried out independently for each vertical water column taking into account observational data situated in the local area near the selected grid point with a specified influence radius. For local observations, the elements of the observation error matrix are weighted according to their horizontal distance from the water column using a function decreasing with distance. This function is the 5th order polynomial [12], whose form is similar to the Gaussian function. The weight function is isotropic and monotonously decreases with distance from the grid point to the influence radius. The function is positive for the distances which are smaller than the influence radius, and is equal to zero otherwise.

The generation of the ensemble of analyses and filter parameters. The LETKF assimilates observational data consecutively over time. For this purpose, the ensemble of model realizations is used, the mean of which is an estimate of the current state, and the variance characterizes uncertainty of this estimate. The initial ensemble of 20 members was created using 20 consecutive states at the end of each day (averaged over 6 hours) obtained in the E2 starting from January 1, 1996 for all model fields, which constitute the state vector: sea surface height η , water temperature T , salinity S , and horizontal components of the velocity of currents (U, V) .

The forecast ensemble is computed by integrating the NEMO equations for each of 20 ensemble members within “the assimilation window” equal to 1 day in these experiments. After the integration on the 1-day prognostic interval, at the analysis step (at the end of every day), the optimum estimate for the increments of oceanic fields at the current moment is calculated, and the ensemble of the model states is corrected by adding the increments to the first guess fields.

The LETKF parameters were the following:

- the ensemble included $N = 20$ members;
- the forgetting factor was assumed equal to $\alpha = 0.95$;
- the influence radius of observations for the weight function was specified equal to 3.5 Δx .

The set of the fields obtained in the E1 is considered representing the true state and is used to create synthetic observations. Such observations were the results of calculations in the E1 at the 2° global grid points averaged over 6 hours at the end of each day, to which random noise simulating the observation errors with specified standard deviations (SD) was added:

Parameter	η , cm	T , C	S , psu	U_s , cm/s	V_s , cm/s
Noise SD	2.0	0.2	0.2	1.0	1.0

The experiments and results of the modeling presented in the given paper are summarized in Table 1 together with the corresponding root-mean-square deviation (RMSD) from the “true” surface ocean characteristics.

5. EXPERIMENTS WITH ASSIMILATION OF SYNTHETIC OBSERVATIONS

The main aim of the present study is to evaluate the capabilities of the PDAF software, that uses the local ensemble Kalman filter. Evaluation is done by example of assimilation of synthetic observations in the NEMO global model for the subsequent application of the PDAF to the assimilation of real oceanographic data. Due to the low horizontal resolution of the model, that is insufficient for a good resolution of ocean dynamics, and high coefficients of viscosity and diffusion, kinetic energy in the experiments with data as-

Table 1. The description of the experiments and the RMSD from the “true” state (data of the E1 experiment) for the sea surface level, water temperature, and salinity on the ocean surface averaged over the model December 1996

Experiment	Description of the experiment	RMSD		
		, cm	T_s , C	S_s , psu
E1	The control experiment imitating the true model state of the ocean	–	–	–
E2	The experiment without data assimilation carried out with higher (as compared to the E1) values of coefficients of turbulent viscosity and diffusion used to obtain first guess fields in the experiments with assimilation of various types of data	4.8	0.42	0.21
ASSH	The assimilation of only sea surface level	2.0	0.43	0.23
AUV	The assimilation of only sea surface currents	4.7	0.43	0.23
ASUR	The joint assimilation of data on the sea surface level and sea surface temperature	2.2	0.30	0.20
AT	The assimilation of temperature profiles to the depth of 2000 m	3.5	0.29	0.21
AT_S	The joint assimilation of profiles of temperature and salinity to the depth of 2000 m	2.9	0.28	0.17
AT_S_SSH	The joint assimilation of T_s and S_s with profiles of temperature and salinity to the depth of 2000 m; 20 ensemble members	2.7	0.28	0.16
AT_S_SSH_25	The assimilation of T_s and S_s with profiles of temperature and salinity to the depth of 2000 m; 25 ensemble members	2.1	0.27	0.16

Table 2. The root-mean-square of water temperature and salinity from the true values averaged over the global simulation domain for December 1996 at the horizons of 500 and 1000 m in different experiments

Experiment	RMSD for T_s , C		RMSD for S_s , psu	
	$z = 500$ m	$z = 1000$ m	$z = 500$ m	$z = 1000$ m
E2	0.45	0.18	0.054	0.023
ASSH	0.40	0.14	0.055	0.022
AUV	0.45	0.18	0.054	0.023
ASUR	0.40	0.14	0.060	0.021
AT	0.22	0.10	0.036	0.018
AT_S	0.19	0.09	0.033	0.018
AT_S_SSH	0.24	0.10	0.036	0.018
AT_S_SSH_25	0.19	0.08	0.030	0.016

simulation almost does not differ from the values in the E2. However, preliminary experiments performed on the 0.1 grid demonstrate an increase in kinetic energy at the assimilation of sea surface level .

Assimilation of surface data. In this series, three experiments imitating the assimilation of data on the ocean surface were carried out. Data of were used at the input of the assimilation procedure only for the areas, where SST was positive. Only the sea level was assimilated in the ASSH experiment, and SST (T_s) was assimilated jointly with in the ASUR experiment. In the AUV experiment, only horizontal components of surface currents (U_s , V_s) were assimilated; this experiment was carried out to assess the effect of including information about surface currents estimated from altimetry measurements to the assimilated data.

The comparison of the results of the AUV with the E2 reveals that the assimilation of horizontal components of surface currents does not lead to the improved simulation of model fields, which may be explained by the low model resolution (see Tables 1 and 2). The assimilation of leads to more accurate

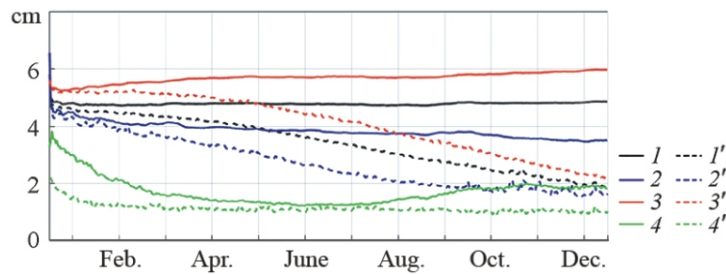


Fig. 2. (1–4) The changes during 1996 in the RMSD from the true values (i.e., from the simulation data in the E1 experiment averaged over the last 6 hours of every day) of the sea surface level in the E2 experiment, as well as (1'–4') at the analysis step in the ASSH experiment. (1, 1') The global averaging; (2, 2') over the Northern Hemisphere north of 5° N; (3, 3') over the Southern Hemisphere south of 5° S; (4, 4') over the equatorial zone (5° S–5° N).

simulation of the sea level surface with a significant decrease in RMSD of σ from the true state (the E1 data are taken as this state): RMSD from the true values calculated for the entire World Ocean at the end of the model simulation decreases by ~60% (see Table 1).

Figure 2 presents RMSDs of the sea surface level σ in the ASSH from the values of σ obtained in the E1. Without the assimilation of σ , the global averaged RMSD in the E2 from the data in the E1 is about 5 cm. When σ is assimilated, RMSD decreases by two times after 8 months of data assimilation and reaches ~2 cm at the end of the simulation, it approaches the prescribed observation error of σ . Due to the high spatial variability of the Southern Ocean and its stronger currents, RMSD of sea level in the E2 decreases more slowly there: only in late October it decreased twice and reached the value of ~2.5 cm at the end of the simulation. As the difference in σ between the E2 and E1 in the equatorial zone in March to August is less than 2 cm (i.e., the prescribed observation error of σ , line 4 in Fig. 2), RMSD in that period in the E2 decreases insignificantly as compared to the E1. However, it is almost twice smaller for the other months of the simulation period.

The assimilation of data on σ alone significantly improves an agreement between the model values of σ and the real values of sea level (Fig. 1b), but does not lead to the improved simulation of surface temperature and salinity fields (Table 1). However, there is a certain improvement in the simulation of the temperature field at deeper horizons: the global averaged RMSD difference in temperature between the ASSH and E1 at the end of the simulation decreased by more than 10 and 20% at the horizons of 500 and 1000 m, respectively (Table 2). This result agrees with the data from paper [10], whose authors revealed a decrease in the model temperature differences from the Argo buoy data at the depth of ~2000 m in case of assimilating real ten-day data on σ . Such improvement may be explained by the fact that velocity varies little in the deep ocean layers and is basically determined by barotropic processes, whose simulation is improved through the assimilation of σ . There is no such improvement for the salinity field in this experiment.

The joint assimilation of σ and SST (T_s) (the ASUR experiment) improves the simulation of σ , as well as in the ASSH, and leads to the accelerated decrease in RMSD by about a month and to similar results at the end of the model simulation. The coupled assimilation of σ and T_s improves the simulation of the temperature field both on the ocean surface and at deeper levels: the global averaged RMSD for SST in the ASUR at the end of the simulation decreased by ~30%; at the horizons of 500 and 1000 m, the RMSD decrease is the same as in the ASSH. At the same time, as compared to the ASSH, the temperature distribution at the depth of 1000 m in the Indian Ocean (south of Australia) and in the area of the Canary Current in the ASUR is simulated better than in the ASSH. However, in a dynamically active region southeast of Africa, RMSD in the ASUR experiment is higher than in the ASSH, but is by ~25% smaller as compared to the E2. Like in the ASSH, the simulation of the salinity field in the ASUR is rather worsened than improved north of 40° N off the western coast of North America.

Assimilation of three-dimensional data. In the next series, two experiments were performed: the first experiment assimilates only water temperature profiles (the AT experiment), and the second experiment assimilates profiles of water temperature and salinity (the AT_S experiment). In both experiments, data at the 2° global grid points at all model levels to the depth of 2 km, to which the Argo buoy observations were carried out, were used at the input of the assimilation procedure. The data were included to the assimilation provided that they were related to the zone from 65° S to 65° N and SST was positive.

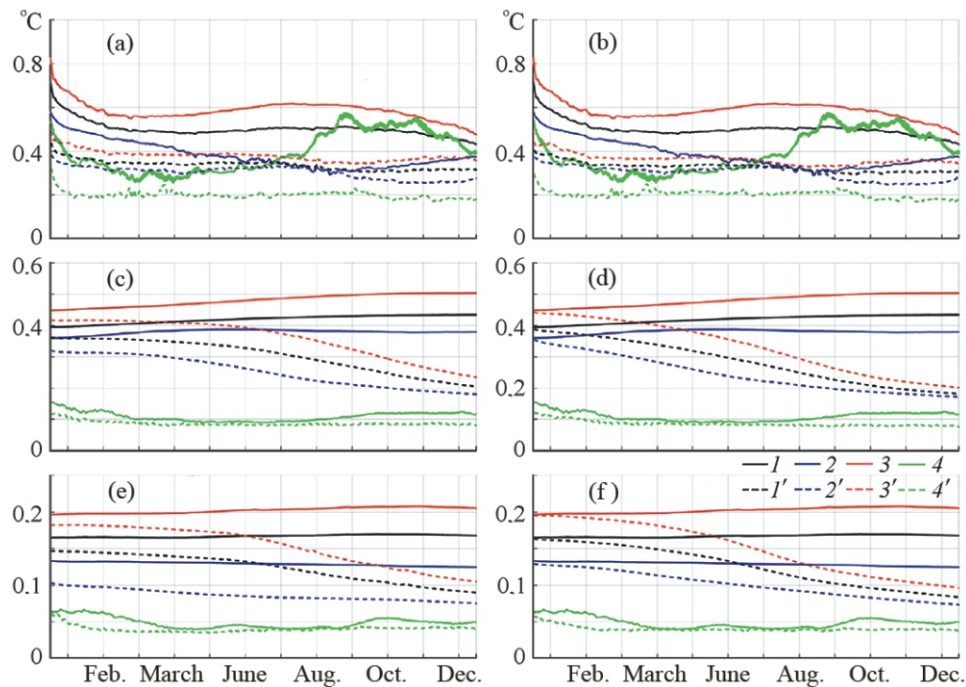


Fig. 3. (1–4) The changes during 1996 in RMSD of water temperature in the E2 experiment, as well as (1–4) at the analysis step in the (a, c, e) AT and (b, d, f) AT_S experiments (a, b) on the surface and at the horizons of (c, d) 500 and (e, f) 1000 m. (1, 1') The global averaging; (2, 2') over the Northern Hemisphere north of 5° N; (3, 3') over the Southern Hemisphere south of 5° S; (4, 4') over the equatorial zone (5° S–5° N).

As follows from Table 1, the assimilation of temperature profiles leads to an almost 30% decrease in RMSD of the monthly mean from the true one. Figure 3 shows RMSD of temperature in the AT calculated at the analysis step from the real values at the end of day averaged over 6 hours and obtained in the E1. Without assimilation, the global averaged RMSD in the E2 varies in the range of 0.45–0.7 °C on the sea surface and within a smaller range at deeper horizons. The assimilation of temperature profiles leads to a ~30% decrease in global RMSD averaged for 1996 at all analyzed horizons (curves 1 in Fig. 3). The SST simulation was especially improved in the equatorial region (RMSD decreased almost by 50% and approached the prescribed observation error for temperature). As the temperature difference in the equatorial zone at the deep horizons is smaller than the prescribed observation error for temperature, RMSD decreases insignificantly there.

On the ocean surface, the RMSD reduction with time during which the assimilation is carried out, is rather quick (during 1–2 months). However, the transformation of the temperature field at the deeper horizons requires a longer period: about half a year for the Northern Hemisphere and about a year for the Southern Hemisphere.

The analysis of the spatial distribution of RMSD for temperature from the true values averaged over December 1996 and obtained at the analysis step demonstrates a considerable improvement in the temperature simulation at all horizons (the area of the zones with the great temperature differences (~5 °C) decreased by 2–3 times). The RMSD significantly decreased in the low latitudes, in the areas of the Gulf Stream, Kuroshio, East Greenland, and Canary currents. In the upper layers in the area of the southern branch of the Brazilian Current and the Agulhas Current, there is still a zone with the high values of RMSD, although its area is much smaller than in the E2. This is explained by the high mesoscale variability of this region, that is not resolved on the coarse simulation grid at the high values of viscosity and diffusion coefficients. The analysis shows that RMSD averaged over the entire basin decreased by ~30% on the sea surface and almost by 50% at the horizons below 400 m.

The assimilation of only temperature profiles does not improve the simulation of the salinity field on the ocean surface, but the salinity fields are better simulated at deeper horizons. For example, RMSD of salinity between the AT and E1 for the last model month averaged over the entire basin at the depth of 500 m decreased by ~35%, and the decrease at the depth of 1000 m made up >20%.

The joint assimilation of temperature and salinity profiles (the AT_S experiment) leads to the further improvement of the model fields of temperature, salinity, and η (Figs. 1c and 1d). At the end of the simulation, RMSD of the monthly mean level η decreases by ~40%, and the decrease in the RMSD in temperature at the deep levels between the AT_S and E2 speeds up by about 1–2 months (Figs. 3e and 3f). The further improvement in the simulation of temperature and salinity can be judged by the value of global averaged differences in temperature and salinity between the AT_S and E1 for the last model month: on the ocean surface, these differences decreased by ~35 and ~20%, respectively; at the horizon of 500 m, the temperature and salinity difference dropped by 60 and 40%, respectively; the respective values for the depth of 1000 m are ~50 and 20% (Tables 1 and 2).

The analysis of the spatial distribution of RMSD of salinity from the real values averaged over December 1996 and obtained at the analysis step, demonstrates an improved simulation of the salinity field at all horizons: RMSD significantly decreased in all dynamically active regions. The areas of the zones with great salinity differences also decreased by 2–3 times.

An exception is the obtained surface salinity field in the Arctic, as salinity was assimilated only to 65 ‰. The spatial distribution of corresponding RMSD for temperature averaged over 1996 between the AT_S and E1 is similar to the distribution in the AT.

Assimilation of three-dimensional and surface data. One more experiment (it is designated as AT_S_SSH) provided the joint assimilation of temperature and salinity profiles together with data on SST T_s and sea level η . The assimilation of such data occurred under the same conditions that were described before.

As clear from Fig. 1e, the joint assimilation of data on temperature T , salinity S , and sea level η leads to a certain worsening of the field simulation as compared to the assimilation of η alone or together with T_s . Nevertheless, the simulation of the field of η in the AT_S_SSH experiment as compared to the AT_S is improved, and the global averaged RMSD for η at the end of the simulation decreases by ~45%, although essential differences for η remain in the areas of the Southern Ocean with strong jet currents.

For the root-mean-square deviation in temperature and salinity between the AT_S_SSH and E1, the similar (like for η) delayed decrease is obtained (not shown). As follows from Tables 1 and 2, at the end of the simulation these differences insignificantly decreased only for η as compared to the AT_S, whereas their certain increase occurred at the deep levels. Nevertheless, the global averaged differences in temperature and salinity between the AT_S_SSH and E1 for the last model month significantly decreased: on the sea surface, this decrease was ~35 and ~25%; at the level of 500 m, RMSD decreased by ~50% for temperature and by 35 for salinity; at the depth of 1000 m, the respective values are ~45 and 20%.

The rate of convergence of the ensemble Kalman filter depends on the number of ensemble members. The convergence is usually rather quick even if using a small ensemble. However, the above worsening of the simulation of the field of η in case of assimilation of the great volume of data is explained by the small ensemble used in the simulations. The additional simulation with 25 ensemble members (the AT_S_SSH_25 experiment), confirmed the above statement about the insufficient volume of the ensemble when assimilating a large number of data in the experiment in the AT_S_SSH. In the new experiment, all RMSDs of the simulated fields decreased as compared to the AT_S_SSH, and their values became comparable to or fall below the previous ones (see Fig. 1f, Tables 1 and 2).

6. CONCLUSIONS

The results of numerical experiments with the model with a low horizontal resolution indicate that a degree of improving model fields in the process of data assimilation is highly dependent on the structure of data at the input of the assimilation procedure.

The assimilation of data related to the ocean surface significantly improves these model fields, except for a case when only the horizontal components of the velocity of surface currents are assimilated (in this case, no improvement occurs for the model fields). If data on the sea surface level are assimilated alone or jointly with SST data, there is also a certain improvement of the temperature field simulation at deep horizons. Hence, if assimilating surface data alone in the model with a low horizontal resolution, it is preferable to assimilate data of η jointly with T_s .

The joint assimilation of water temperature and salinity profiles essentially improves both these model fields and the simulated field of the sea surface level η . Such assimilation during 12 months reduces the global averaged RMSD of the monthly mean sea surface η by ~40% (by ~60% for the AT_S_SSH_25 experiment); on the sea surface, the global averaged RMSD in temperature and salinity decreased by ~35

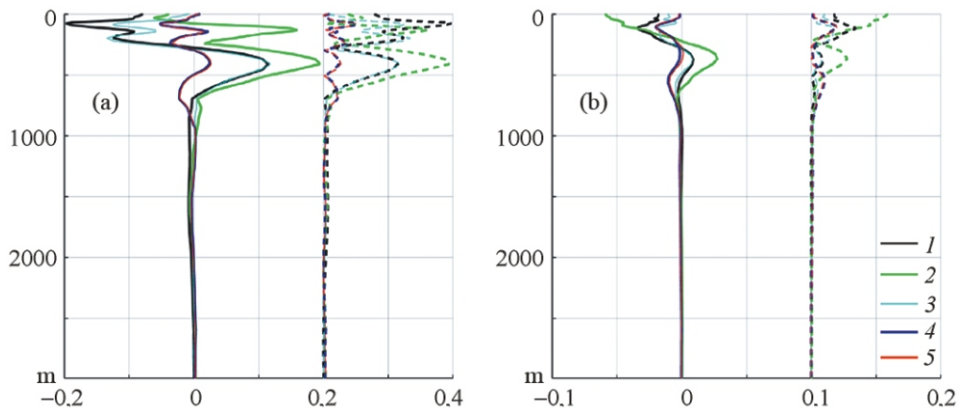


Fig. 4. The vertical distributions of mean (the solid lines) and root-mean-square differences (the dash lines) in (a) water temperature (°C) and (b) salinity (psu) between the experiments (1) E2 and E1, (2) ASSH and E2, (3) ASUR and E2, (4) AT_S and E2, (5) AT_S_SSH and E2. The results of the AT_S and AT_S_SSH experiments are similar, so that the blue and red lines mainly match. The RMSD curves are displaced to the right by 0.2 °C for temperature and by 0.1 psu for salinity.

and 20% (by ~40 and 25% for the AT_S_SSH_25); at the horizons of 500 and 1000 m, these RMSDs decreased by ~60 and 40, ~50 and 20% (by ~60 and 45, ~55 and 30% for the AT_S_SSH_25), respectively.

Figure 4 presents the differences averaged over the World Ocean and root-mean-square deviations in the vertical distributions of temperature and salinity between different experiments. The greatest differences between all experiments are reported in the upper ~500-m layer: the mean differences between the E2 and ASSH experiments reach ~0.2 °C for temperature and 0.05 psu for salinity, and salinity differences and salinity RMSDs in the ASSH are higher than in the E2. The assimilation of T_s jointly with T_s (the ASUR experiment) reduces temperature and salinity differences as compared to the E2, but only in the upper layer of ~200 m, which does not considerably change these globally averaged differences at the lower horizons. As clear from Fig. 4, the joint assimilation of temperature and salinity fields substantially decreases temperature and salinity differences as compared to the E2 in the upper layer of ~500 m. However, on average for the entire World Ocean it gives a certain difference for the temperature and salinity fields at the lower boundary of thermocline (~700–800 m). Thus, the joint assimilation of temperature and salinity fields leads to the greater enhancement of model fields than the assimilation of surface data alone, especially the assimilation of the sea surface level alone.

In the analyzed experiments, data on water temperature and salinity with a rather large horizontal spacing (on the 2° grid) were used at the input of the assimilation procedure. To assess the dependence of assimilation results on the volume of assimilated data, the additional experiment was performed with the assimilation of sea temperature and salinity data related to the points of the sparser 4° grid (which is close to the nominal density of Argo buoys) and with the influence radius for the weight function equal to 6.5°. The results of this experiment show that the four-fold decrease in the volume of “observation” data has a little effect on the global averaged RMSD of temperature and salinity from the “true” distributions, which increased by not more than 20% in the upper layer of ~500 m and by less than 10% at the lower horizons. It should be noted that the largest increase in RMSD for salinity at the upper horizons is obtained in the Northern Hemisphere (not presented); however, since this four-fold reduction of the volume of “observations” in the Northern Hemisphere considerably underestimates the real density of Argo buoys, for example, in the Atlantic Ocean, the simulation of oceanic characteristics in the Northern Hemisphere using real Argo data can perform better.

The experiments with data assimilation also revealed that the increase in the number of ensemble members to 25 is sufficient for the ensemble assimilation of data by the NEMO model to provide an adequate simulation of the oceanic fields on the ~1° model grid.

Figures 2 and 3 show that the curves corresponding to the experiments with assimilation demonstrate a tendency toward the stabilization of errors only at the end of the annual simulation period. This means that obtaining stable initial conditions in the large-scale circulation model in the framework of the prognostic problem requires assimilating observational data during several years.

ACKNOWLEDGMENTS

The authors thank L. Nerger for valuable advice during the work on the inclusion of the PDAF system to the NEMO model.

FUNDING

The research is performed in the framework of the theme 1.1.10 of the Roshydromet Research and Development Plan for 2020.

REFERENCES

1. I. O. Dumanskaya, A. A. Zelenko, S. A. Myslenkov, E. S. Nesterov, S. K. Popov, Yu. D. Resnyanskii, and B. S. Strukov, "Marine Hydrological Forecasting and Operational Oceanology in the Hydrometeorological Center of Russia," *Gidrometeorologicheskie Issledovaniya i Prognozy*, No. 4 (2019) [in Russian].
2. A. A. Zelenko and Yu. D. Resnyanskii, "Marine Observational Systems as an Integral Part of Operational Oceanology: A Review," *Meteorol. Gidrol.*, No. 12 (2018) [Russ. Meteorol. Hydrol., No. 12, **43** (2018)].
3. M. N. Kaurkin, R. A. Ibrayev, and K. P. Belyaev, "Data Assimilation in the Ocean Circulation Model of High Spatial Resolution Using the Methods of Parallel Programming," *Meteorol. Gidrol.*, No. 7 (2016) [Russ. Meteorol. Hydrol., No. 7, **41** (2016)].
4. A. N. Kolmogorov, "Equations of Turbulent Motion of Incompressible Fluid," *Izv. AN SSSR, Ser. Fiz.*, No. 1–2, **6** (1942) [in Russian].
5. G. K. Korotaev, "Operational Oceanography: A New Branch of Modern Oceanology," *Vestnik Rossiiskoi Akademii Nauk*, No. 7, **88** (2018) [in Russian].
6. G. I. Marchuk, B. E. Paton, G. K. Korotaev, and V. B. Zalesny, "Data-Computing Technologies: A New Stage in the Development of Operational Oceanography," *Izv. Akad. Nauk, Fiz. Atmos. Okeana*, No. 6, **49** (2013) [*Izv., Atmos. Oceanic Phys.*, No. 6, **49** (2013)].
7. G. A. Sarkisyan and V. N. Stepanov, "Method for Calculating Physical Characteristics of the Ocean from an Individual Hydrologic Section," *Izv. Akad. Nauk, Fiz. Atmos. Okeana*, No. 4, **35** (1999) [*Izv., Atmos. Oceanic Phys.*, No. 4, **35** (1999)].
8. V. N. Stepanov, Yu. D. Resnyanskii, B. S. Strukov, and A. A. Zelenko, "Large-scale Ocean Circulation and Sea Ice Characteristics Derived from Numerical Experiments with the NEMO Model," *Meteorol. Gidrol.*, No. 1 (2019) [Russ. Meteorol. Hydrol., No. 1, **44** (2019)].
9. B. S. Strukov, Yu. D. Resnyanskii, and A. A. Zelenko, "Relaxation Method for Assimilation of Sea Ice Concentration Data in the NEMO–LIM3 Multicategory Sea Ice Model," *Meteorol. Gidrol.*, No. 2 (2020) [Russ. Meteorol. Hydrol., No. 2, **45** (2020)].
10. A. Androsov, L. Nerger, R. Schnur, J. Schroter, A. Albertella, R. Rummel, R. Savcenko, W. Bosch, S. Skachko, and S. Danilov, "On the Assimilation of Absolute Geodetic Dynamic Topography in a Global Ocean Model: Impact on the Deep Ocean State," *J. Geodesy*, **93** (2019).
11. R. Dussin, B. Barnier, L. Brodeau, and J.-M. Molines, *The Making of the DRAKKAR Forcing Set DFS5*, DRAKKAR/MyOcean Report 01-04-16, April 2016, https://www.drakkar-ocean.eu/publications/reports/report_DFS5v3_April2016.pdf.
12. G. Gaspari and S. E. Cohn, "Construction of Correlation Functions in Two and Three Dimensions," *Quart. J. Roy. Meteorol. Soc.*, **125** (1999).
13. P. R. Gent and J. C. McWilliams, "Isopycnal Mixing in Ocean Circulation Models," *J. Phys. Oceanogr.*, No. 1, **20** (1990).
14. P. L. Houtekamer and H. L. Mitchell, "A Sequential Ensemble Kalman Filter for Atmospheric Data Assimilation," *Mon. Wea. Rev.*, **129** (2001).
15. P. L. Houtekamer and H. L. Mitchell, "Data Assimilation Using an Ensemble Kalman Filter Technique," *Mon. Wea. Rev.*, **126** (1998).
16. B. R. Hunt, E. J. Kostelich, and I. Szunyogh, "Efficient Data Assimilation for Spatiotemporal Chaos: A Local Ensemble Transform Kalman Filter," *Physica D*, **230** (2007).
17. P. Kirchgeßner, J. Todter, B. Ahrens, and L. Nerger, "The Smoother Extension of the Nonlinear Ensemble Transform Filter," *Tellus A*, **69** (2017).
18. W. G. Large and S. G. Yeager, *Diurnal to Decadal Global Forcing for Ocean and Sea Ice Models: The Data Sets and Flux Climatologies*, Technical Report TN-460+STR (NCAR, 2004).
19. R. A. Locarnini, A. V. Mishonov, J. I. Antonov, T. P. Boyer, H. E. Garcia, O. K. Baranova, M. M. Zweng, C. R. Paver, J. R. Reagan, D. R. Johnson, M. Hamilton, and D. Seidov, *World Ocean Atlas 2013*, Vol. 1: *Temperature*, Ed. by S. Levitus (NOAA Atlas NESDIS 73, 2013).

20. G. Madec and the NEMO team, *Nemo Ocean Engine—Version 3.6*, Technical Report No. 27 (Pole de Modelisation de l’Institut Pierre Simon Laplace, 2016).
21. L. Nerger and W. Hille, “Software for Ensemble-based Data Assimilation Systems—Implementation Strategies and Scalability,” *Computers and Geosciences*, **55** (2013).
22. L. Nerger, W. Hiller, and J. Schroter, “PDAF—The Parallel Data Assimilation Framework: Experience with Kalman Filtering,” in *Use of High Performance Computing in Meteorology* (2005).
23. D. T. Pham, J. Verron, and M. C. Roubaud, “A Singular Evolutive Extended Kalman Filter for Data Assimilation in Oceanography,” *J. Mar. Syst.*, **16** (1998).
24. *Sea Ice Modelling Integrated Initiative (SI3)—The NEMO Sea Ice Engine*, Scientific Notes of Climate Modelling Center, No. 31 (Institut Pierre-Simon Laplace).
25. J. Todter, P. Kirchgessner, L. Nerger, and B. Ahrens, “Assessment of a Nonlinear Ensemble Transform Filter for High-dimensional Data Assimilation,” *Mon. Wea. Rev.*, **144** (2016).
26. M. Vancoppenolle, T. Fichefet, H. Goosse, S. Bouillon, G. Madec, and M. A. M. Maqueda, “Simulating the Mass Balance and Salinity of Arctic and Antarctic Sea Ice. 1. Model Description and Validation,” *Ocean Modelling*, No. 1–2, **27** (2009).
27. M. M. Zweng, J. R. Reagan, J. I. Antonov, R. A. Locarnini, A. V. Mishonov, T. P. Boyer, H. E. Garcia, O. K. Baranova, D. R. Johnson, D. Seidov, and M. M. Biddle, *World Ocean Atlas 2013*, Vol. 2: *Salinity*, Ed. by S. Levitus (NOAA Atlas NESDIS 74, 2013).