

The results of numerical simulation of the Lena River runoff with the assimilation of satellite data: summer 2008*

G.A. Platov, E.G. Klimova

Abstract. The paper discusses the results of preliminary experiments to test the quality of the data assimilation procedure based on the use of ensemble Kalman filter applied to the basin of the Laptev Sea in the vicinity of the Lena Delta. As perturbation we used the river runoff closure, and as the true values – the surface salinity, taken from a reference experiment with the included river inflow. The comparison of two numerical experiments with assimilation of simulated salinity data and without assimilation shows that the proposed assimilation procedure is able to restore adequately the salinity field.

1. Introduction

Until recently, oceanographic observations were sporadic. Currently, there is a large number of regular satellite observations (the sea surface temperature, the sea surface elevation), as well as data being obtained in real time by autonomous Argo systems. Unfortunately, these observations are more complicated in case of the Arctic region. For example, satellite data are useful only if they are received in the summer period in the ice-free area. The winter recorded values characterize the surface of ice and snow, but not the ocean. In addition, the satellite trajectories around the Earth are, as a rule, unfavorable for obtaining reliable data for such high latitudes. Application of floating buoys is also limited because they require to make regular ascents to the surface to transmit the data records, but under the ice cover it is technically impossible. Nevertheless, the amount of regular information is continuously growing, which allows us to consider the problem of estimating the ocean state with the help of a mathematical model describing the dynamics of the shelf seas of the Arctic Ocean with allowance for the observational data, i.e., the problem of data assimilation. It is commonly understood that the data assimilation is a joint account of observations and results of numerical implementation of a mathematical model to obtain the most accurate assessment of spatial and temporal distribution of the quantities.

*Supported by Integration Project No. 109 of Siberian Branch of Russian Academy of Sciences.

Currently, data assimilation algorithms are widely used for modeling processes in the ocean. There are many techniques of data assimilation, but in terms of the mathematical formulation of the problem, they all use one of the two approaches: variational (3DVAR, 4DVAR) or stochastic (the Kalman filter). At present, operational ocean data assimilation systems use both 4DVAR [31] and the Kalman filter [24, 33] methods, and, moreover, some versions of the multi-element optimum interpolation are also in use (see, for example, [29]). In Russia, the creation of data assimilation systems for the ocean is also the focus of the Russian Hydrometeorological Center from several research groups. Researchers have developed a system for the ocean data assimilation which uses a variational approach (3DVAR) [28]. The data assimilation system for the ocean based on the variational approach (4DVAR) is developed by specialists from the Institute of Numerical Mathematics (Moscow) [1].

The Kalman filter algorithm is currently one of the most popular approaches to solve the data assimilation problem. To obtain the optimal estimate of the state vector according to observations and predictive model, which is nonlinear in general, the solution of the equation for the conditional mean is required [17]. This problem cannot be solved in general form, therefore there are various simplified versions of the problem, which allow us to reduce it to the equations for the conditional mean and covariance. These simplifications are based on the linearization relative to the reference state or on the expansion in power series of the estimation error (truncated second order filters). Moreover, there may be used an additional assumption that the considered random fields are Gaussian [17].

The most commonly used area of research into the application of the Kalman filter is the ensemble approach, first proposed in [6]. It was further developed in [7, 8, 13–15, 21, 22]. In the ensemble approach, the covariance matrices of estimation errors are computed for nonlinear prognostic models. The ensemble Kalman filter is a version of the extended Kalman filter, in which covariances of the forecast errors are estimated using the ensemble of forecasts. Along with this approximate description it was proposed to use suboptimal algorithms [11] in which the probability averaging is replaced by time-averaging, assuming random fields having temporal ergodicity. This paper suggests an algorithm based on this approach to solve the problem of data assimilation of the ocean observations. The development of algorithms for data assimilation for modeling processes in the ocean was based on our experience of data assimilation systems for the atmospheric pollution model [19, 20].

Evaluation of the algorithm developed was carried out by means of numerical experiments with model-simulated data (the so-called “identical twin” experiments). It means that numerical experiments were carried out on assimilation of data with a “real” space-time distribution. They were

based upon the AARI data of the International Polar Year, 2007–2008 [30], and the Pathfinder NOAA/AVHRR archive of satellite observations for the same period.

2. The data assimilation method

The main technique involves the data assimilation algorithm, based on the theory of the optimal Kalman filtering. Let time dependence of the variable x_k^f be described by the equation

$$x_k^f = A_{k-1}(x_{k-1}^f), \quad (1)$$

where A_{k-1} is the operator of a prognostic model. The “true” value of the variable x_k^t is calculated by the equation

$$x_k^t = A_{k-1}(x_{k-1}^t) + \varepsilon_{k-1}, \quad (2)$$

where ε_{k-1} is the stochastic vector of the model “noise”, each component of which is distributed according to a normal distribution with zero mean and covariance matrix Q_{k-1} . At the time of the observations t_k , observational data are represented as

$$y_k^o = H_k x_k^t + \xi_k, \quad (3)$$

where H_k is the linear operator of the “observations”, interpolating the values of x_k^t from grid nodes into observation points; ξ_k is the vector of observation errors, which are random variables distributed according to the normal distribution with zero mean and covariance matrix R_k .

The classical algorithm of the Kalman filter consists of the two steps: *forecast* (1), with the time-dependent covariance matrix computed according to the formula

$$P_k^f = A_{k-1} P_{k-1}^f A_{k-1}^T + Q_{k-1}, \quad (4)$$

and *analysis*

$$x_k^a = x_k^f + K_k(y_k^o - H_k x_k^f), \quad (5)$$

where

$$K_k = P_k^f H_k^T (H_k P_k^f H_k^T + R_k)^{-1}, \quad P_k^a = (I - K_k H_k) P_k^f.$$

In these formulas, P_k^f and P_k^a denote covariance matrix of the forecast errors and analysis errors, respectively.

Implementation of the Kalman filter algorithm formulas for modern three-dimensional models of high dimensions is complicated because of the need to store in memory and to operate with super-high-dimensional matrices. One of the most popular suboptimal algorithms based on the Kalman filter, is an ensemble Kalman filter [6, 8, 13]. In the ensemble Kalman filter,

covariance matrices P_k^f are evaluated using an ensemble of forecasts. It requires assigning the ensemble initial fields and the computation and storage of ensembles of forecasts and analysis.

In the case of the fields having a temporal ergodicity, their covariance can be estimated by replacing the sample averaging for the time averaging [11, 32]. Let the values of the random field errors be initially given in the form

$$x_0^f = x_0^t + \Delta x_0,$$

where Δx_0 are random variables distributed according to the normal distribution with zero mathematical expectation and covariance matrix P_0^f . We assume that a change in time of the error is described by the linearized equation. We denote by Δx_i the forecast errors at the time t_i . Then the covariance matrix of the forecast errors at the time $t_N = N \cdot \Delta t$ can be estimated from the running average [11]:

$$P_N^f = \overline{\Delta x_N \Delta x_N^T} \cong \frac{1}{N-1} \sum_{i=1}^N \Delta x_i \Delta x_i^T. \quad (6)$$

The estimation error on the step of the analysis will satisfy

$$\Delta x_k^a = \Delta x_k^f - K_k(y_k^o - H_k x_k^f). \quad (7)$$

This paper also considers a version in which we estimate the field error correlation, while the dispersion does not change. In this case, the correlation of the errors is also estimated by the formula of running average (6). A similar assumption of a constant ratio between the dispersions of the prediction error and the error of observations was made in [33]. In this case it is possible to avoid the divergence of the ensemble filter with time.

Since the calculation of the covariance by formula (6) is an approximate estimate, a problem arises common for all ensemble algorithms: unreasonably large values of covariances at large distances. For this reason, we use the so-called localization procedure, as is accepted in the ensemble Kalman filter approach. It is element-wise multiplication of the covariance matrix by a function decreasing with distance and with depth

$$\exp\left(-\frac{\Delta z^2}{2R_z^2}\right) \cdot \exp\left(-\frac{\Delta x^2 + \Delta y^2}{2R^2}\right),$$

where $\sqrt{\Delta x^2 + \Delta y^2}$ and Δz are horizontal and vertical distances, and R and R_z are horizontal and vertical scales, which were taken to be equal to 3,000 m and 5 m, respectively.

3. A regional nested model

To solve the problem of data assimilation of both in situ and satellite measurements in step forecast, a system of nested models described in [26] was used. The system includes a large-scale submodel of the Arctic and the North Atlantic Ocean [9, 10] coupled with an ice-snow submodel [16]. The resulting model participated in the international Arctic Ocean model intercomparison project [27]. The system includes also a regional model [2] adapted to the Laptev Sea basin with a horizontal resolution concentrated in the vicinity of the delta of the Lena River. By combining different scale models, the interaction of processes of the global and the regional scales is carried out.

Increasing the resolution in the regional model results in a more detailed description of the dynamics, although it requires a large amount of the CPU time. A number of processes that have an important influence on this dynamics cannot be correctly described within the large-scale model. Such processes include the propagation of topographical and coastal trapped waves and tides. When moving these waves cause the surface level disturbance that leads to cracks in the ice cover and helps to relieve the stress strain in its field.

To have a high horizontal resolution in the areas of a steep shelf slope is sufficient for a satisfactory description of these waves. In addition, a detailed resolution near the Lena River delta is also necessary for the correct description of the interaction of riverine and marine water. The model grid for the region, built to meet these requirements, is shown in Figure 1. To describe the dynamics of the shelf, a detailed grid resolution is required not only for the surface, but also for the bottom boundary layer. Moreover, the model should allow for vertical displacements of the sea surface, i.e., the “rigid lid” condition, which is used in the large-scale model, is unacceptable for the regional shelf model. Among the models satisfying these requirements, the sigma-coordinate model of the Princeton University (POM) has been selected as the most studied one [2].

A problem of downscaling needs to be addressed both in terms of accounting a large-scale distribution within a nested model, and in terms of the integrated account of the influence of smaller-scale processes in the large-scale dynamics. The first part of this problem is solved by setting the initial and boundary conditions on the open boundaries of a nested model. The values of temperature T , salinity S and a normal component of the barotropic velocity (U, V) are interpolated onto the boundary nodes of a nested model. The stream function ϕ of the large-scale model is converted to obtain the corresponding sea surface elevation η according to [23]:

$$\frac{\partial \eta}{\partial x} = \frac{f}{gH} \frac{\partial \phi}{\partial x} - \frac{1}{H\rho_0} \int_0^H (H-z) \frac{\partial \rho}{\partial x} dz + \frac{\tau_x}{gH\rho_0},$$

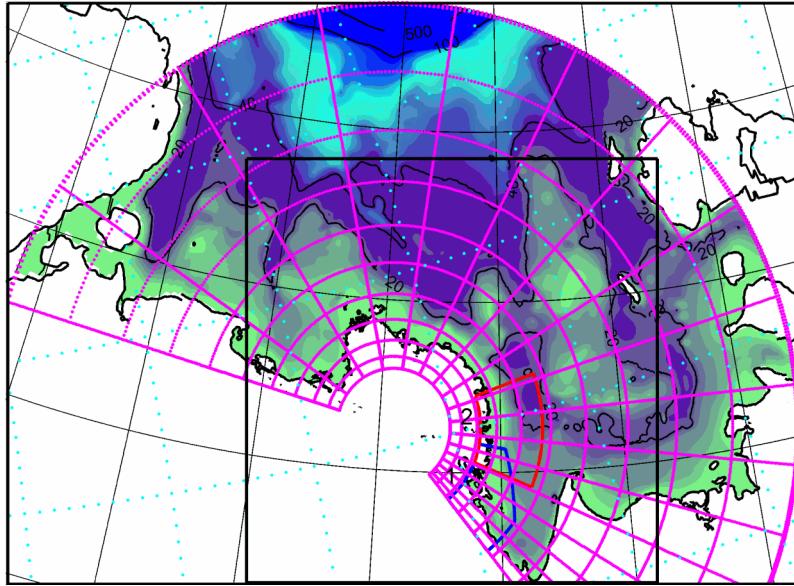


Figure 1. A map of the bottom topography (in m) for the Laptev Sea simulation area. Contours 1 and 2 encircle zones under the direct influence of the main flows of the Lena River corresponding to the Bykov and the Trofimov outlets. A rectangular frame indicates to the area depicted in a number of subsequent figures. A series of concentrated circles and radii represent grid lines of regional models (1 of 30), while dotted lines represent grid lines of the large-scale model of the Arctic and the North Atlantic Ocean (1 of 10)

$$\frac{\partial \eta}{\partial y} = \frac{f}{gH} \frac{\partial \phi}{\partial y} - \frac{1}{H\rho_0} \int_0^H (H-z) \frac{\partial \rho}{\partial y} dz + \frac{\tau_y}{gH\rho_0}.$$

Furthermore, the sigma-coordinate model requires the specification of the background (in the original version, climatic [2]) distributions of temperature and salinity in the model area in order to reduce the errors associated with the calculation of the pressure gradient, and the errors associated with the diffusion of heat and salinity along the sigma surfaces. In the latter case, the diffusion is applied only to deviations of temperature and salinity from their background distributions. In the case of nested models, it is reasonable to use the temperature T_0 and the salinity S_0 of the large-scale model as background distributions. Thus, changes in temperature and salinity as a result of the horizontal diffusion could be presented in the form

$$\begin{aligned} \left(\frac{\partial T}{\partial t} \right)_{\text{diff}} &= \left\{ \frac{\partial}{\partial x} \left(A \frac{\partial}{\partial x} \right) + \frac{\partial}{\partial y} \left(A \frac{\partial}{\partial y} \right) \right\} (T - T_0), \\ \left(\frac{\partial S}{\partial t} \right)_{\text{diff}} &= \alpha_s \left\{ \frac{\partial}{\partial x} \left(A \frac{\partial}{\partial x} \right) + \frac{\partial}{\partial y} \left(A \frac{\partial}{\partial y} \right) \right\} (S - S_0). \end{aligned}$$

The idea of a nested model feedback was borrowed from [4, 25], which describe the parameterization of topographic eddies in the momentum equation. Applying a similar approach to the equations of heat and salt content, we obtain the following expressions

$$\begin{aligned}\left(\frac{\partial T_0}{\partial t}\right)_{\text{diff}} &= \left\{ \frac{\partial}{\partial x} \left(A_0 \frac{\partial}{\partial x} \right) + \frac{\partial}{\partial y} \left(A_0 \frac{\partial}{\partial y} \right) \right\} (T_0 - T), \\ \left(\frac{\partial S_0}{\partial t}\right)_{\text{diff}} &= \alpha_s \left\{ \frac{\partial}{\partial x} \left(A_0 \frac{\partial}{\partial x} \right) + \frac{\partial}{\partial y} \left(A_0 \frac{\partial}{\partial y} \right) \right\} (S_0 - S), \\ \left(\frac{\partial u}{\partial t}\right)_{\text{diff}} &= \left\{ \frac{\partial}{\partial x} \left(\mu \frac{\partial}{\partial x} \right) + \frac{\partial}{\partial y} \left(\mu \frac{\partial}{\partial y} \right) \right\} (u - U), \\ \left(\frac{\partial v}{\partial t}\right)_{\text{diff}} &= \left\{ \frac{\partial}{\partial x} \left(\mu \frac{\partial}{\partial x} \right) + \frac{\partial}{\partial y} \left(\mu \frac{\partial}{\partial y} \right) \right\} (v - V),\end{aligned}$$

where the values T , S , U , and V are obtained by averaging the small-scale temperature, salinity and components of barotropic velocity over the grid cell of the large-scale model. In this case, unlike the Newtonian term (also known as “nudging”), this scheme does not add any new sources or sinks of heat and salt, but only redistributes the already available contributions.

The location of grids of nested models is presented in Figure 1. The coordinate lines of the nested model do not coincide with the coordinate lines of the large-scale model, so, in order to ensure proper interaction between models, we need to solve the problem of data transfer from one grid to another. The following interpolation formula was used for this purpose

$$\tilde{\psi}_i = \sum_j C_{ij} \psi_j \Bigg/ \sum_j C_{ij}, \quad (8)$$

where ψ_j is the value of a certain variable ψ at the j th node of the original grid, $\tilde{\psi}_i$ is the resulting value obtained by interpolation at the i th node of the destination grid. The weighting coefficients C_{ij} are calculated depending on the distance r_{ij} between any j th node of the original grid and the i th node of the destination grid by the formula

$$C_{ij} = \exp\left(-\frac{r_{ij}^2}{4R^2}\right),$$

where R is the search radius. The value of R should guarantee the existence of, at least, three nodes of the original grid within this radius. The summation in (8) is over N nearest nodes. In order to facilitate the task of interpolation from one grid to another during the model run, the pre-calculated interpolation coefficients $C_{ij}/\sum_j C_{ij}$ were used. Also, it was assumed that $N = 16$ and R is equal to the local grid spacing of the large-scale model, $R = \max\{\Delta x, \Delta y\}$.

4. Numerical tests

The basic model calculation was carried out using the factual information for 2008 announced by the International Polar Year (IPY). The latter circumstance has proved to be important to gain the access to Roshydromet database, which carried out an extended program of hydrometeorological observations in 2008. The calculation is performed for the open-water season from September 1 to September 30. For modeling the exchange processes in the surface layer and calculating components of the heat balance, the NCEP/NCAR reanalysis data were interpolated onto the sea grid. These data involve the near-surface wind velocity, humidity and temperature of the air, along with precipitation rate and cloudiness. Initial state of the Laptev Sea was built from the ocean state corresponding to September 1, 2008 results of the large-scale model of the Arctic and the North Atlantic. This numerical experiment was performed according to AOMIP rules simulating the period of 1948–2012 (see <http://www.whoi.edu/page.do?pid=29917>).

4.1. Reference test. The above assimilation scheme was applied to improve the results of the nested shelf model. In order to test the assimilation procedure, the “true” value was simulated by running this model. The integration covered the period from 1st to 30th September, 2008. The adopted initial state of the model was a result of the large-scale model execution, interpolated onto the grid of the regional model. As consequence, the initial temperature and salinity fields are rather smooth. The results of the measurements obtained with the gauging station Kyusyur were used to define the total transport of the Lena River in the model. The average transport of the river was about 35,000 m³/s (with minimum 26,000 m³/s in September 5 and maximum 44,000 m³/s in September 24) and it was proportionally re-distributed among its four major watercourses: Olenek outlet—7 %, Tumatsk—14 %, Trofimov—54 %, and Bykov—25 % of the total transport, with the river water temperature taken to be 10 °C, and with salinity equal to zero.

As the result of this test will be considered as “observation”, we define the error equal to $\sigma_o = 0.5$, which corresponds to r.m.s. variation of salinity between the neighboring grid nodes in the vicinity of the Lena delta.

Available observations are a set of satellite data of Pathfinder v. 5.3, presented online at the NOAA site. It comprises a skin temperature of water in case the sea surface is not covered with ice or by clouds. Measurements are timed to the night and to the day, so the observations were made as frequent as two times a day. Each set contains from a few to 40 thousand measurements attributable to the model region. Most of measurements were taken from 5th to 10th September and at the end the month: from 24th to 30th September. Most of them were made in the western part of the area

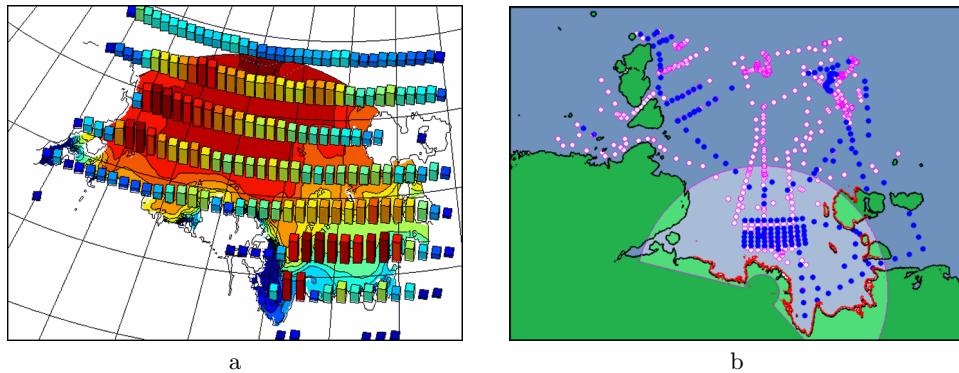


Figure 2. Observational data: a) a frequency histogram of acceptable satellite observations in the Laptev Sea area (high frequency corresponds to a large column height and a warmer color, ranging from dark blue to dark red), contours at the base of the histogram represent the salinity field of the base experiment corresponding to September 30, 2008, b) location of the points of ground observations (open circles) in the summer of 2008 in the area of the Laptev Sea, the filled circles represent points of observations made in September

off the Taimyr Peninsula, and in the south-east part in the direction of the river outflow off the Lena Delta (Figure 2a). Also, we used the data obtained when implementing the International Polar Year (IPY) [30]. The locations, where the corresponding measurements were made, are presented in Figure 2b.

The results of this reference test for the surface salinity are shown later in Figures 4a, d, g that show the spread of the river plume (a freshwater anomaly formed by the riverine runoff) along the coastline to the east. Similar patterns were also obtained in other papers [3, 5, 12] and have the qualitative and quantitative similarities with our results.

4.2. Salinity test. In order to evaluate the effectiveness of the data assimilation procedure, the following test was performed, in which the perturbed external forcing was used, distinguishing it from the reference test. Thus, we can consider it as perturbation of the model operator A_{k-1} in (1). The total runoff of the Lena River was used as a perturbation parameter. If we set it to be equal to zero, then the resulting salinity field will show excessive salt concentrations near the major outflows of the Lena River delta. Checking the quality of the assimilation procedure will be in how accurately we can restore the salinity field by assimilation of our reference test results, produced with the unperturbed river transport. In the experiments, the covariance matrix P_N^f was evaluated by formula (6), the error was taken as a difference between the forecast and the “true” values minus the averaged value of this difference for all $i = 1, \dots, N$. We evaluate the error of the disturbed

operator as equal to $\sigma_f = 5$ for the salinity (we still have $\sigma_o = 0.5$).

In order to simulate as many as possible of the existing measurements, although we used the reference test salinity as assimilable data, but the salinity values were interpolated only at the points where there are real observations. Most of observations are attributed to the surface.

The total number of observations N (mainly satellite observations) is very large, and the inversion of a corresponding covariance matrix with $N \times N$ dimension is more than being difficult. Therefore, we used the following two simplifications.

The first simplification is the dividing the whole domain into equal sub-domains, i.e., boxes. More specifically, the dimension of the entire grid area 301×501 was divided into ten parts in each direction. The result was one hundred boxes with 31×51 dimension. Furthermore, for a better joint of these boxes, we organized overlapping zones of 13 grid nodes wide. So, taking into account the overlappings, the dimension of the boxes was found to be 43×63 . Nevertheless, even for such small subdomains the amount of data occasionally turned out to be great. In this regard, one more simplification was made.

It consists in the adopted specification that a maximum amount of data involved is forced to be equal to $\tilde{N} = 500$ observations. In the case, where the actual amount of data is less than 500, all of them, with some exceptions due to the quality control, are used in the assimilation. If there are more than 500, then we randomly select $\tilde{N} = 500$ quality observations, and neglect the rest one. Some data are subject to excepting due to the quality control if they are different from the predicted value by more than three standard deviations.

The data assimilation period is 12 hours, as the available observations are confined to the nighttime or to the daytime; the deviations from midnight or from noon are within 2–3 hours. Therefore, regardless of the true observation time, they were attributed to either midnight or noon UTC (local time is UTC+10). The time interval, corresponding to the period of assimilation, was divided into 24 intervals, resulting in twenty-four periods of 30 minutes. The state at the end of each of these periods was taken up as part of the ensemble for the assimilation procedure. Thus, instead of the classical ensemble assimilation scheme, where n results of simulation with a different initial perturbation are analyzed, we consider n time-snapshots of a simulation with one initial perturbation.

Such a definition of the ensemble initial fields is applied in the weather forecast and it is called “lagged average forecasting” (LAF) [18]. LAF is that the ensemble perturbations are taken not at one moment of time, but for the previous few moments.

Figure 3a above shows a standard deviation of salinity in the experiments with and without data assimilation from the results of the reference

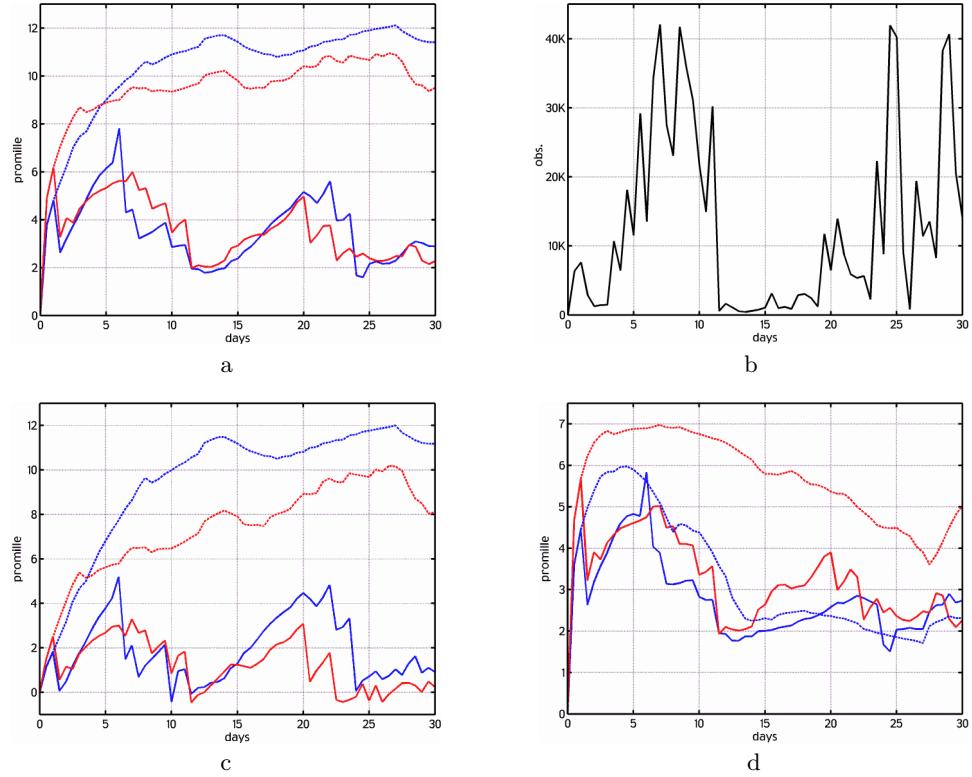


Figure 3. The test results, the time series of: a) a standard deviation of the surface salinity from the baseline experiment, $\sqrt{(S - S_{\text{ref}})^2}$ for the first (blue line) and for the second zones (red line) (see Figure 1), the dashed lines correspond to the experiment without assimilation, the solid lines— with data assimilation; c) and d) the average deviation from the baseline experiment $\overline{S - S_{\text{ref}}}$ and the r.m.s. variation of the deviation $\sqrt{(S - S_{\text{ref}} - \overline{(S - S_{\text{ref}})})^2}$, symbols are the same as those in the figure a; b) the total number of measurements in each 12-hour periods from satellite data and from the IPY data

test. Periods from 5 to 10 September and from 24 to 30 September, rich of observations, are characterized by a significant reduction in errors in salinity in the experiment with data assimilation. In zones 1 and 2 (see Figure 1) the error is reduced from 6–8 to 2 psu during the first period, and from 5 to 2.3 psu during the second one. The dispersion, corresponding to a standard deviation shown in Figure 3a, is equal to $(\overline{S - S_{\text{ref}}})^2$, where S is the resulting salinity, and S_{ref} is the corresponding salinity from the reference test, the overline indicates that the value is area averaged. Dispersion could be considered as consisting of the two parts. The first one is associated with the average value of the deviation $\overline{S - S_{\text{ref}}}$. Another one is dispersion of the

deviation with respect to its average $\overline{(S - S_{\text{ref}})^2}$, thus

$$\overline{(S - S_{\text{ref}})^2} = \overline{S - S_{\text{ref}}}^2 + \overline{(S - S_{\text{ref}} - \overline{S - S_{\text{ref}}})^2}.$$

The first (see Figure 3c) characterizes the area averaged salinity deviation between these two experiments, and the second (see Figure 3d) characterizes

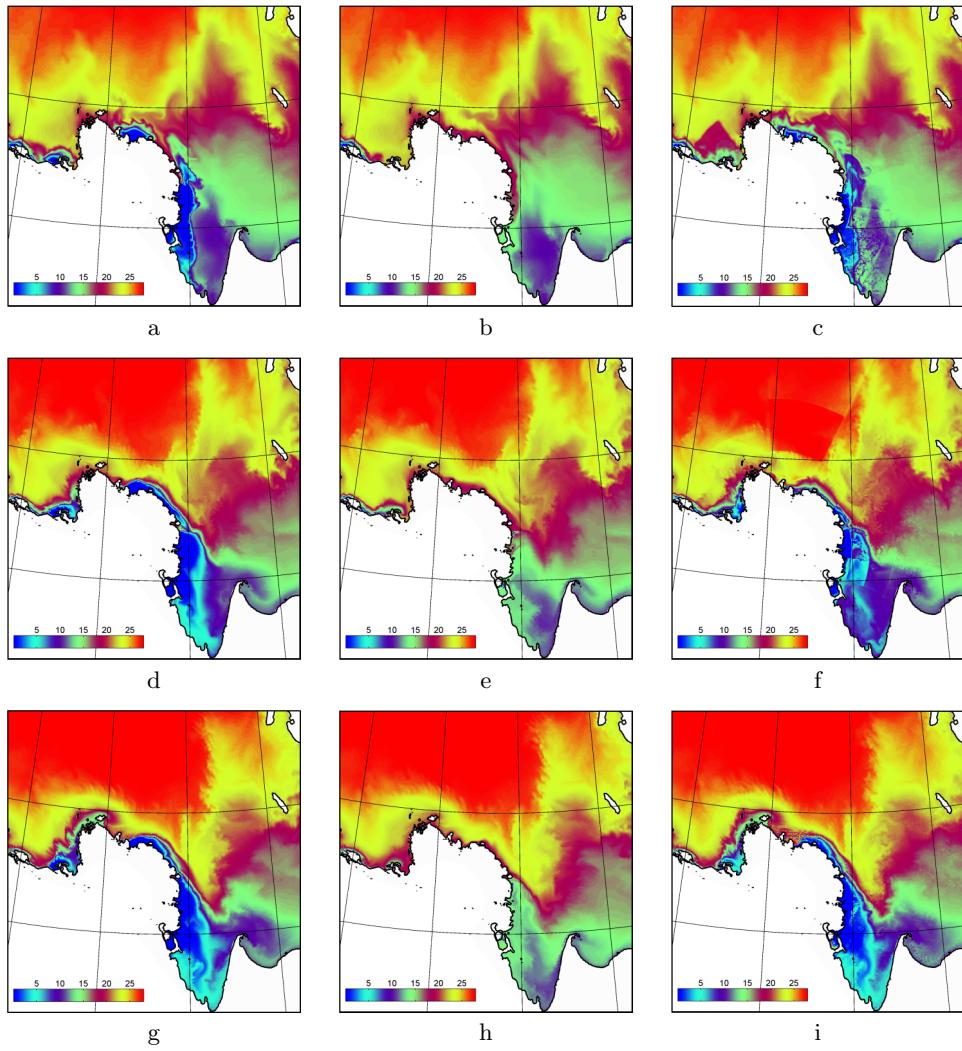


Figure 4. Results of the surface salinity test on September 10 (a, b, c), on September 20 (d, e, f), and on September 30 (g, h, i). The first column (a, d, g) is from the reference test, the second column (b, e, h) is from the salinity test without data assimilation, and the third column (c, f, i) is from the salinity test with data assimilation

the spatial variability of the salinity deviation. From these figures it follows that the error is reduced mainly due to the decrease of the difference between the average values of the compared experiments, while the areal variability of the error does not change much regardless of the assimilation, especially in the first zone.

The surface salinity fields in the experiments with and without assimilation compared to the reference test shows a series of Figures 4. A salinity anomaly formed in the Buor Khaya Bay (a bay adjacent to the delta of the Lena River in the east) as a consequence of the summer flood, and resulting from the large-scale model, is gradually disappearing in the experiment without assimilation. The difference between the salinity distribution in the experiment with the assimilation and the salinity distribution in the reference test, more prominent on 10th and 20th September, almost completely disappears in the figure demonstrating these distributions on September 30.

5. Discussion

The results of this paper are preliminary. They demonstrate the numerical tests for the proposed assimilation procedure. As an assimilated value we use the most important (in the region) hydrodynamic characteristic of water: salinity, since its effect on the density is greatest. As the data we used the results of the reference test. In order to emulate real observations, these results were used only where data of satellite or ground-based measurements are available. As a disturbing factor we considered the vanishing of the Lena River flow. As a result, the freshwater plume, formed in the adjacent bays in the east, gradually disappears, and the fresh river waters are replaced by the salty sea waters. In general, the procedure for data assimilation leads to the recovery of the unperturbed results on the sea surface (Figures 4). But, at greater depths, where the observations are insufficient, the picture is less satisfactory. For these depths, some additional assumptions are required to help in obtaining some appropriate extrapolation of the surface values.

As was already noted, the salinity in the Laptev Sea plays a dominant role in determining the density of water as its variability in this region is high due to the presence of large volumes of fresh water from the Lena River. In addition, salinity, other conditions being equal, has a greater weight in the equation of state. On the other hand, the Laptev Sea is shallow and so the wind has a strong impact on the character of water circulation in it. It is clear that the upper layer is more exposed to the wind (or ice) stress and the riverine inflow also dominates here. Available vertical profiles show that in September, the upper mixed layer thickness is about 10–12 m in this region. What factor is dominant for underlying layers can be seen from a comparison of flow patterns in the three presented experiments (Figure 5).

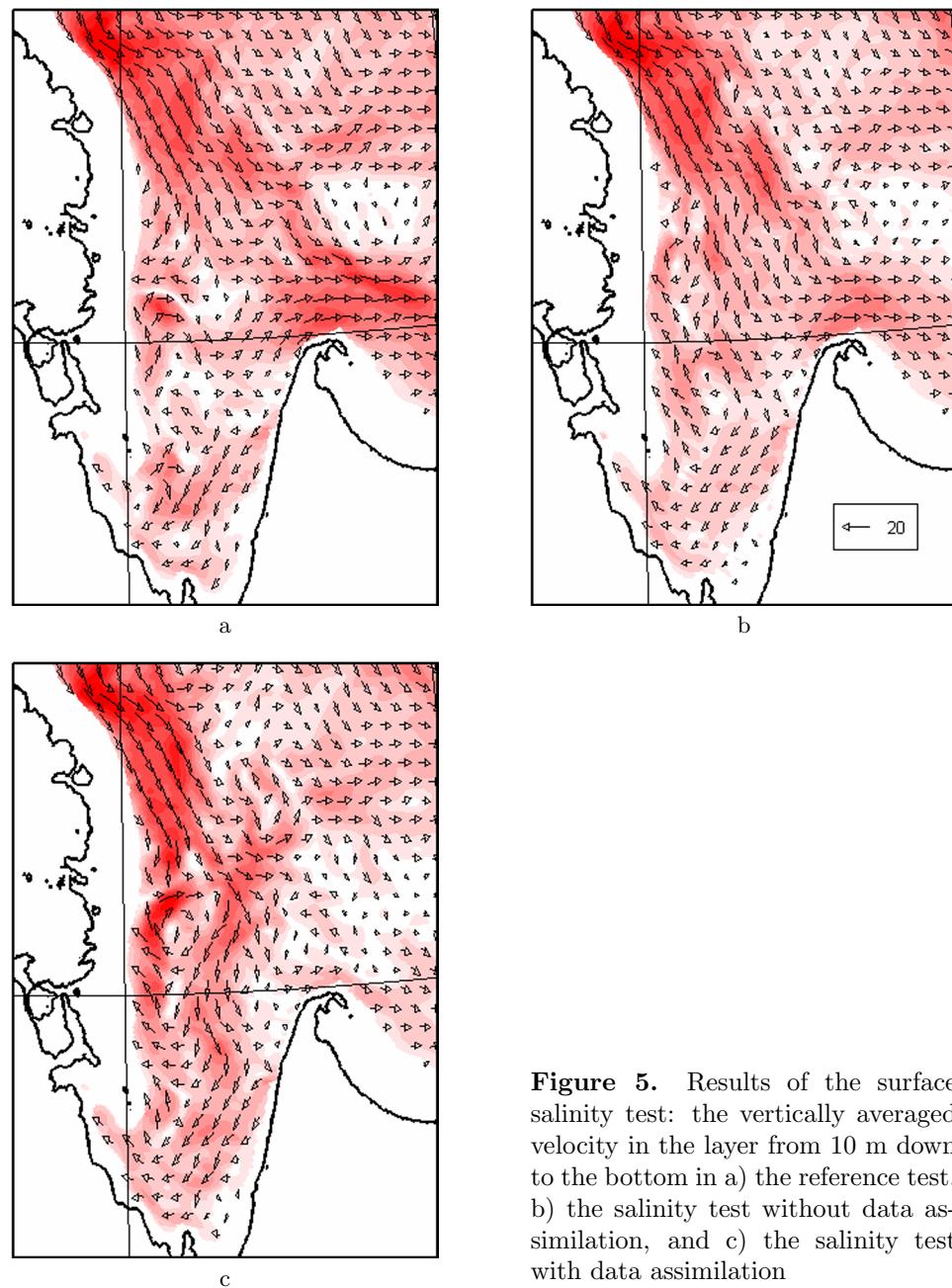


Figure 5. Results of the surface salinity test: the vertically averaged velocity in the layer from 10 m down to the bottom in a) the reference test, b) the salinity test without data assimilation, and c) the salinity test with data assimilation

Comparing Figures 5a and b, we can see that the flow pattern in the vicinity of the Buor Khaya Bay is more or less similar whether there is a river inflow or not. The only noticeable difference is a slight weakening of the along-shore current at the entrance of the Yana Bay in the salinity test. What we see from the assimilation result is that the flow pattern is more different from the reference test than in the case of no data assimilation. The reason for this discrepancy lies in disagreement between temperature, salinity and current velocity fields arising from a correction of the salinity. For future tests this means that temperature and current velocity must also be somehow involved into the assimilation algorithm.

In the Arctic, one of the determining factors is also the ice field pattern. It should be noted that the propagation pattern of the Arctic ice, taken from the results of the model used, is far from being perfect. For example, in September 2008, according to satellite images, the Laptev Sea ice edge moves north to latitude 78° N. At the same time, the simulation results show at 77° N latitude ice thickness about 1 m with compactness of 85 %. This fact demonstrates that in order to obtain satisfactory results of modeling the dynamics of the Laptev Sea, one should directly use the observed pattern of the ice field.

Despite the fact that the temperature field as a dynamic factor, plays a secondary role in the Laptev Sea, the temperature distribution is important for the studying the marine biology. The Arctic seas, as a whole, provide a fairly harsh conditions for biological activity. However in the summer, when rivers supply large amounts of flood waters rich in minerals and microorganisms, and the polar night is replaced by the polar day, biological activity markedly increasing. In this case, the temperature isolines are actually boundaries of life and death for many plants and living organisms. Thus, the next planned series of experiments will deal with the restoration of the temperature field.

This research is also preliminary, because here we almost do not compare the proposed scheme of data assimilation with a classical version of assimilation using the ensemble Kalman filter. The advantage of the newly-proposed scheme is obvious, as it involves obtaining a set of the ensembles in one model run. However, the payment for this will be associated with a possible loss of quality. This will require further studying.

Another important area in terms of improving the procedure for data assimilation is to use a multi-element approach. In this example, we used the salinity field, and in the future we are going to use temperature. However, as part of the assimilation procedure, a joint analysis of salinity and temperature is possible, including also some data on the rise of the sea level and ice concentration. It is possible not only to get a better assessment of these characteristics, but taking into account the cross-covariances, any other characteristics of a regional ocean model.

6. Conclusion

This paper discusses the results of preliminary experiments to test the quality of the data assimilation procedure based on the use of the ensemble Kalman filter as applied to the basin of the Laptev Sea in the vicinity of the Lena Delta. As perturbation, we used the closure of the river runoff, and as true values the surface salinity was considered, taken from a reference experiment with the included river inflow. The comparison of two numerical experiments with assimilation of simulated salinity data and without assimilation have shown that the proposed assimilation procedure is capable to satisfactorily restore the salinity field. However, it should be noted that the resulting circulation structure is moving away from the structure shown in the reference experiment in the vicinity of the Lena Delta. This fact raises questions about the involvement in the process of data assimilation of some other prognostic variables.

References

- [1] Agoshkov V.I., Ipatova V.M., Zalesnyi V.B., et al. Problems of variational assimilation of observational data for ocean general circulation models and methods for their solution // *Izv. Atmospheric and Oceanic Physics*. – 2010. – Vol. 46, No. 6. – P. 677–712. doi:10.1134/S0001433810060034.
- [2] Blumberg A.F., Mellor G.L. A description of a three-dimensional coastal ocean circulation model // *Three-Dimensional Coastal Ocean Models*, 4 of *Coastal and Estuarine Series* / American Geophysical Union, Washington, D.C. – 1987. – P. 1–16.
- [3] Chant R.J., Glenn S.M., Hunter E., et al. Bulge formation of a Buoyant river outflow // *J. Geophys. Res.* – 2008. – Vol. 113. doi:10.1029/2007JC004100.
- [4] Cherniawsky J., Holloway G. On western boundary current separation in an upper ocean general circulation model of the North Pacific // *J. Geophys. Res.* – 1993. – Vol. 98. – P. 22843–22853.
- [5] Choi B.J., Wilkin J.L. The effect of wind on the dispersal of the Hudson river plume // *J. Phys. Oceanogr.* – 2007. – Vol. 37, No. 7. – P. 1878–1897. doi:10.1175/JPO3081.1.
- [6] Evensen G. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics // *J. Geophys. Res.* – 1994. – Vol. 99(C5). – P. 10143–10162.
- [7] Evensen G. The ensemble Kalman filter: theoretical formulation and practical implementation // *Ocean Dynamics*. – 2003. – Vol. 53. – P. 343–367.
- [8] Evensen G. Data Assimilation. The Ensemble Kalman Filter. – Berlin, Heidelberg: Springer, 2009.

- [9] Golubeva E.N., Platov G.A. On improving the simulation of Atlantic water circulation in the Arctic Ocean // *J. Geophys. Res.* — 2007. — Vol. 112. — doi:10.1029/2006JC003734.
- [10] Golubeva E.N., Platov G.A. Numerical modeling of the Arctic Ocean ice system response to variations in the atmospheric circulation from 1948 to 2007 // *Izv. Atmospheric and Oceanic Physics.* — 2009. — Vol. 45, No. 1. — P. 137–151.
- [11] Handbook on Automatic Control Theory / A.A. Krasovsky, ed. — Moscow: Nauka, 1987 (In Russian).
- [12] Hickey B.M., Pietrafesa L.J., Jay D.A., Boicourt W.C. The Columbia river plume study: subtidal variability in the velocity and salinity fields // *J. Geophys. Res.* — 1998. — Vol. 103, No. C5. — P. 10339–10368.
- [13] Houtekamer P.L., Mitchell H.L. Data assimilation using an ensemble Kalman filter technique // *Mon. Wea. Rev.* — 1998. — Vol. 126. — P. 796–811.
- [14] Houtekamer P.L., Mitchell H.L. A sequential ensemble Kalman filter for atmospheric data assimilation // *Mon. Wea. Rev.* — 2001. — Vol. 129. — P. 123–137.
- [15] Houtekamer P.L., Mitchell H.L., Pellerin G., et al. Atmospheric data assimilation with an ensemble Kalman filter: results with real observations // *Mon. Wea. Rev.* — 2005. — Vol. 133. — P. 604–620.
- [16] Hunke E.C., Dukowicz J.K. An elastic-viscous-plastic model for ice dynamics // *J. Phys. Oceanography.* — 1997. — Vol. 27, No. 9. — P. 1849–1867.
- [17] Jazwinski A.H. *Stochastic Processes and Filtering Theory.* — New York: Academic Press, 1970.
- [18] Kalnay E. *Atmospheric Modeling, Data Assimilation and Predictability.* — Cambridge Univ. Press, 2002, 377 pp.
- [19] Kilanova N.V., Klimova E.G. Numerical experiments on model error estimation in a problem of a passive pollution concentration data assimilation // *Computational Technologies.* — 2006. — Vol. 11, No. 5. — P. 32–40.
- [20] Klimova E.G., Kilanova N.V. Numerical experiments on estimation of methane emission based on the data assimilation system for passive impurity in the atmosphere of the Northern hemisphere // *Atmospheric and Oceanic Optics.* — 2006. — Vol. 19, No. 11. — P. 863–866.
- [21] Klimova E.G. A data assimilation technique based on the π -algorithm // *Russian Meteorology and Hydrology.* — 2008. — Vol. 33, No. 3. — P. 143–150. doi:10.3103/S1068373908030023.
- [22] Klimova E.G. Data assimilation technique based on the ensemble π -algorithm // *Russian Meteorology and Hydrology.* — 2008. — Vol. 33, No. 9. — P. 570–576. doi:10.3103/S1068373908090045.

- [23] Marchuk G.I., Sarkisian A.S. Mathematical Modeling of Ocean Circulation.— Moscow: Nauka, 1988 (In Russian).
- [24] Nadiga B.T., Casper W.R., Jones P.W. Ensemble-based global ocean data assimilation // Ocean Modelling. — 2003. — Vol. 72. — P. 210–230. doi:10.1016/j.ocemod.2013.09.002.
- [25] Nazarenko L., Holloway G., Tausnev N. Dynamics of transport of “Atlantic signature” in the Arctic Ocean // J. Geophys. Res.—1998.—Vol. 103.—P. 31003–31015.
- [26] Platov G.A. Numerical modeling of the Arctic Ocean deepwater formation: Part II. Results of regional and global experiments // Izv. Atmospheric and Oceanic Physics. — 2011. — Vol. 47, No. 3.— P. 377–392.
- [27] Proshutinsky A., Aksenov Y., Clement Kinney J., et al. Recent advances in Arctic ocean studies employing models from the Arctic Ocean Model Intercomparison Project // Oceanography. — 2011.— Vol. 24, No. 3.— P. 102–113.
- [28] Resnyansky Ju.D., Zelen'ko A.A. Development of models and methods of analysis of observation data for monitoring and predicting large-scale processes in the ocean // Proc. “80th Anniversary of Hydrometcentre of Russia”.— Moscow: Triada Ltd., 2010.— P. 350–375.
- [29] Reynolds R.W., Smith T.M., Liu C., Chelton D.B., Casey K.S., Schlax M.G. Daily high-resolution-blended analysis for sea surface temperature // J. Climate. — 2007.— Vol. 20.— P. 5473–5496. doi:10.1175/2007JCLI1824.1.
- [30] Scientific program of the Russian Federation participation in the International Polar Year (2007/2008).— Moscow, 2006.— http://www.ipyrus.aari.ru/scientific_program.html.
- [31] Weaver A.T., Deltel C., Machu E., et al. A multivariate balance operator for variational ocean data assimilation // Q.J.R. Meteorol. Soc.—2005.— Vol. 131, No. 613.— P. 3605–3625. doi:10.1256/qj.05.119.
- [32] Yaglom A.M. Correlation Theory of Stationary Random Functions. — Leningrad: Gidrometeoizdat, 1981.
- [33] Yin Y., Alves O., Oke P.R. An ensemble ocean data assimilation system for seasonal prediction // Mon. Wea. Rev. — 2011.— Vol. 139.— P. 786–808. doi:10.1175/2010MWR3419.1.