Multivariate EnOI-based data assimilation in the high resolution ocean model

M N Kaurkin$^{1,2,3,4}$, R A Ibrayev$^{1,2,3,4}$

1 Marchuk Institute of Numerical Mathematics, Russian Academy of Sciences, Moscow, Russia
2 Shirshov Institute of Oceanology, Russian Academy of Sciences, Moscow, Russia
3 Hydrometeorological Centre of Russia, Moscow, Russia
4 Marine Hydrophysical Institute, Russian Academy of Sciences, Sevastopol, Russia

Abstract. The ensemble Optimal Interpolation (EnOI) method is considered for assimilation of observational data in the INMIO ocean model of high spatial resolution. Its parallel implementation is described in the form of the Data Assimilation Service (DAS) based on the Compact Modeling Framework (CMF3.0). The on EnOI based parallel algorithm is designed to assimilate data from various sources (drifters, satellites, etc) with the ability to construct cross-covariance matrices (between different model values) for adjusting all model fields and the model forecast as a whole, including those functions, for which there is no measurements. This method based on EnOI is planned to be used for all observational data available to date for the World Ocean.

1. Introduction
High spatial resolution ocean models with the assimilation of continuously arriving observation data allow us to make operational forecast of a three-dimensional state of the marine environment similar to meteorological weather forecasts. Solution of this problem is a very important part of climate change studies and environmental monitoring systems.

The high resolution of the ocean model plays an important role in reliable modeling of ocean dynamics. In particular, studies of the dynamics of the Atlantic Ocean have shown that modeling with a grid spacing of not more than 0.1° is critical for reproducing the space-time characteristics of the Gulf Stream [1]. A high resolution makes it possible to explicitly model the eddy structure, which is not manifested in a more coarse resolution. This phenomenon is related to the ratio of the step of the model grid to the Rossby deformation radius.

Over the past 10 years, the amount of satellite observation data for ocean (surface temperature, sea level, ice concentration, etc) is seriously increased. As well as whole systems of buoys for contact measurements (ARGO, PIRATA, TAO, TRITON) are developed and operate. The volume of digital environmental data is growing rapidly. For example, the volume of the archive of data stored in the National Centers for Environmental Information (NCEI) of NOAA (USA) is dozens of petabytes, largely due to an increase in the number of satellite observational and model calculations data [2].

The task of effectively using the growing amount of observational data to understand ocean processes is relevant. For example, satellite observational data for surface temperature and sea level with a resolution of less than 1 km (for example, NASA Aqua) are already available. This corresponds to the arrival of~ 1 GB of information per day. It is important to use this information correctly and quickly to improve the quality of the forecast. Also, the problem of forecasting the state of the ocean becomes urgent and in demand.
The main feature of the methods based on ensemble Kalman filter (EnKF) is that they do not require the construction of an adjoint model operator and the solution of the inverse problem which can be very difficult when added a large number of parameterization in the numerical model. EnKF-based methods can use the program code of the model as a "black box". That allows developing and changing the configuration of the model and adds new data type for assimilation in the simplest way. The accuracy of data assimilation methods based on EnKF, according to the results published in [3], is at 4DVar level, but EnKF methods are better suitable for massively-parallel computers with distributed memory [4]. In paper [5] EnKF and EnOI are compared with the MOM4 model (resolution of 0.1 degrees) and significantly increased computational costs (proportional to ensemble size), the forecast accuracy using EnKF is only 10-17% higher.

One more advantage of ensemble methods of data assimilation is the possibility of using them in coupled ocean, ice and atmosphere models together with construction of cross covariance matrices of all observational data so that the data obtained for one of the components could directly affect the state of the other component when obtaining the analysis. The study of [Sluka et al, 2016] demonstrates the advantage of this approach in comparison with the separate assimilation of data in each of the components operating within the overall model of the Earth system (ocean-atmosphere model).

This work continues the research of the authors published earlier in [4,7,8]; the aim of this study is to develop and test multivariate EnOI-based data assimilation method in INMIO ocean model and to check its potential and effectiveness with different types of observation data.

2. Theoretical formulation of the EnOI method

Basic equations used in the assimilation of data by dynamic stochastic methods [9] are as follows:

\[ x_a = x_b + K(y_{obs} - Hx_b) \]  
\[ K = BH^T(HBH^T + R)^{-1} \]

where \( x_a, x_b \) are the vectors of the model solution after and before the assimilation (analysis and background) of size \( n \), where \( n \) is the number of model grid points multiplied by the number of model values to be corrected (temperature, salinity, level, etc.), \( n \sim 10^8 \) for the World Ocean model (WOM) with a resolution of 0.1 degrees; \( y_{obs} \) is a vector of observations of size \( m \), where \( m \) is the number of observation points (\( m \sim 10^3-10^5 \)). The equations (1) can contain various data, such as temperature, salinity, ocean level, etc; \( K (n \times m) \) is the weight matrix (Kalman gain matrix); \( R (m \times m) \) is the covariance matrix of observation errors; \( H (m \times n) \) is the matrix of projecting model values into the observational data space; \( B \) is the covariance matrix of the model state errors.

The basic idea of ensemble methods (EnKF and EnOI) is that matrix \( B \) is approximated on the basis of a set of model vectors (ensemble).

\[ A_b = [x_1^b \ldots x_N^b] - [\bar{x}_s \ldots \bar{x}_s] \]

In (3), \( A_b \) is an \( n \times N \) matrix, where \( N \) is the number of elements of the ensemble (\( N \sim 10^2 \)), whose columns are equal to the values of the model state minus the ensemble average. The ensemble of model states in the EnKF method is constructed from model calculations with different initial conditions. In the EnOI method, model computations that are saved during long-term model integration are selected for the state ensemble. Then the matrix \( B \) is approximated as follows:

\[ B \approx \frac{1}{N-1} A_b (A_b)^T \]

3. Experiments with multivariate assimilation

Testing of the multivariate assimilation scheme was carried out in the North Atlantic from 33° S to 67° N. The conditions of experiments are the same as described in previous papers [7-8].

3.1. Experiments description
Four main numerical experiments with real atmospheric forcing were performed, corresponding to the period from 01.05.2008 to 10.06.2008 (40 days). The summary information on these experiments is given in Table 1. A01 is control experiment (no data assimilation); A02 is assimilation of only ARGO drifters data (temperature and salinity profiles, T/S) by the EnOI method; A03 is assimilation of only Sea Level Anomaly (SLA) AVISO data by the EnOI method; A04 is assimilation of AVISO and ARGO data together by the EnOI method with the construction of cross covariance matrices. In all experiments, the model solution is compared daily to the ARGO drifters data on temperature and salinity.

### Table 1. The description of experiments with data assimilation.

|       | Assimilation ARGO T/S | Assimilation SLA AVISO | Comparison with ARGO T/S |
|-------|-----------------------|------------------------|--------------------------|
| A01   | -                     | -                      | +                        |
| A02   | +                     | -                      | +                        |
| A03   | -                     | +                      | +                        |
| A04   | +                     | +                      | +                        |

In the experiments A02 ($x_{a_i}, x_{b_i}$, see Section 2), model solution vectors are made up of three-dimensional temperature and salinity fields; $y_{obs}$ is the observation vector consisting of ARGO drifters data received for the model day (about 102 profiles per day). Also in experiments A02 in matrix B, vertical correlation will not be taken into account, since ARGO observation data are available for all horizons at a depth of up to 1500 meters. In A03 and A04 experiments, model solution vectors are composed of three-dimensional fields of temperature and salinity and two-dimensional sea-level field. In A03, $y_{obs}$ consists of Sea Level Anomaly (SLA) AVISO satellite data from Jason-1 (~ 103 points). In A04, $y_{obs}$ consists of combination of SLA AVISO and ARGO T/S data. The covariance matrix B will consider correlation between the model values: sea level, temperature and salinity on various model horizons. So in the experiments A03 and A04, the data assimilation corrects the whole solution vector.

### 3.2. Analysis of the results

In Figure 1 the ocean surface temperature in INMIO model is shown for the multivariate assimilation experiment (A04) for 10 June, 2008. Also observation points with ARGO T/S and AVISO SLA data are shown. This figure is given for estimation of amount and irregularity of the data for assimilation.
Figure 1. Surface temperature for the North Atlantic in the INMIO model in the experiment with multivariate assimilation by EnOI method (A04) for 10 June, 2008. The crossmarks show the ARGO drifters, whose temperature and salinity profiles were measured on June 10, 2008. The dots show the points, where the sea-level data on June 10, 2008 were available from AVISO project (the Jason-1 satellite).

Evolution of the RMS background error with the model time for different depths (3, 105, 310 meters) is shown in Figure 2. The background error $L_2(Hx - y_{obs})$ is counted daily at the moment before data assimilation by EnOI method in experiments A02, A03 and A04. The control error in the experiment A01 is calculated relative to the observation data, but there is no data assimilation. When comparing the error control charts for the basic experiment (A01) and background error for experiments with assimilation (A02, A03, A04), we can conclude that assimilation gives the correct sign of correction at all depths. The difference between the background and control error increases with depth with the influence of atmospheric forcing. The especially important result is that the multivariate assimilation data (i.e. assimilation data from different sources) give a better effect than separate assimilation of each type of data.

By the end of the experiment A04RMS temperature background error became about 1.5°C, while in the experiments A02 (only ARGO) and A03 (only SLA) the error was 2-2.5°C. Assimilation of salinity also benefits on all horizons.
Figure 2. The RMS error of the control for the experiment without assimilation (A01) and the background error for the assimilation of ARGO T/S (A02), AVISO SLA (A03) data and their multivariate combination (A04) by the EnOI method for the model temperature field in °C at a depth of 3 (a), 105 (b) and 310 (c) meters and for the model salinity field in psu at a depth of 3 (d), 105 (e) and 310 (f) meters (compared with the data of ARGO drifters data). The days are on the X axis, the temperature error in °C or the salinity error in psu are on the Y.
4. Conclusion
The first experiments of the multivariate data assimilation for INMIO model were carried out using the scheme of ensemble optimal interpolation (EnOI) and original computational platform CMF3.0, which makes it possible to realize EnOI algorithm on supercomputers with the high resolution ocean dynamics model. It was confirmed that multivariate assimilation of the altimetry, temperature and salinity data with constructing cross-covariance matrices decreases the errors in three-dimensional model fields of temperature and salinity significantly. The multivariate assimilation gives better results than the separate assimilation of each type of geophysical data.

5. References

[1] Ibrayev R A, Khabeev R N and Ushakov K V 2012 Izv. Atmos. Ocean. Phys. 48(1) 37
[2] Karl T R 2015 Key challenges for environmental data and information as viewed from NCEI. 3rd Meeting of Department of Commerce Data Advisory Council, Boulder, CO, USA.
[3] Kalnay E, Li H, Miyoshi T et al 2007 Tellus A. 59 (5) 758
[4] Kaurkin M, Ibrayev R and Koromyslov A 2016 Supercomputing (Eds. Voevodin V and Sobolev S) Series: CCIS. Springer.687 57
[5] Sakov P and Sandery P A 2015 Ocean Modelling. 89 45
[6] Sluka T C, Penny S G, Kalnay E, Miyoshi T 2016 Geophys. Res. Lett 43 752
[7] Kaurkin M N, Ibrayev R A and Belyaev K P 2018 Izv. Atmos. Ocean Phys. 54(1) 64
[8] Kaurkin M N, Ibrayev R A and Belyaev K P 2016 Oceanology 56 (6) 774
[9] Oke P R, Brassington G B, Griffin D A and Schiller A. 2010 AMOJ. 59 67

Acknowledgments
This study was supported by Grant no. 17-77-30001 from the Russian Science Foundation (RSF) in Marine Hydrophysical Institute of RAS. The simulations used resources from the supercomputer complex of the Interdepartmental Supercomputer Centre of the RAS.