

Data assimilation experiments with MPIESM climate model

Konstantin Belyaev¹, Andrey Kuleshov^{2,a}, Ingo Kirchner³ and Natalia Tuchkova⁴

¹Shirshov Institute of Oceanology of RAS, Moscow, Russia

²Keldysh Institute of Applied Mathematics of RAS, Moscow, Russia

³Institute of Meteorology, Freie Universität Berlin, Berlin, Germany

⁴Federal Research Center "Computer Science and Control" of RAS, Moscow, Russia

Abstract. Further development of data assimilation technique and its application in numerical experiments with state-of-the-art Max Plank Institute Earth System model have been carried out. In particular, the stability problem of assimilation is posed and discussed. In the experiments the sea surface height data from archive Archiving, Validating and Interpolating Satellite Ocean have been used. All computations have been realized on cluster system of German Climate Computing Center. The results of numerical experiments with and without assimilation were recorded and analyzed. A special attention has been focused on the Arctic zone. It is shown that there is a good coincidence of model tendencies and independent data.

1 Introduction

Data assimilation methodology and the numerical climate modeling are among the most popular research directions over last 20 years. One of the frequently used versions of the numerical climate models is the Max Plank Institute Earth System model (MPIESM), created in Max Plank Institute for Meteorology in Hamburg, Germany. This model version is realized on cluster system in German Climate Computing Center (DKRZ) and consists of several modules including ocean block MPIOM, atmosphere block ECHAM6 and others which are executed in parallel. Details of the MPIESM and its realization can be found in particularly in [1].

Actual climate modeling research also often combines with data assimilation approaches. Data assimilation methods allow correcting the model output by data and obtaining the new initial conditions for future better forecast. The examples of data assimilation methods and their applications to the climate and weather predictions can be found, for instance, in [2]. An example of the MPIESM model in conjunction with data assimilation has been considered in [3].

Several issues are resolving in these types of research. First, the numerical model is turning on the real natural observations and eliminates the unrealistic and completely non-relevant scenarios which could be possible without assimilations. Secondly, the observed variables can correct not only the corresponding model components but also non observed model characteristics. Finally, the regions with badly data coverage can involve the observed information from more data richer domains through model connection and evaluate both local and global temporal and spatial variability.

These and others potential opportunities of correct combinations of models and assimilation methods require further developments all components which produce this progress. The model themselves, their numerical realization including modern parallel schemes and technologies, further information advance, data bases, data assimilation techniques and their mathematical realization lead to this progress. It is important to estimate the correction of data assimilation application for specific model and data base, and to prove that the resulted model fields are physically relevant.

2 Data assimilation method

In the current study the original data assimilation technique, developed earlier in [4, 5] is applied. The method has been applied earlier in conjunction with ocean circulation model HYCOM in [5]. In the current study the same technique is applied together with MPIESM. Along with the data assimilation block the entire model scheme is presented in Figure 1.

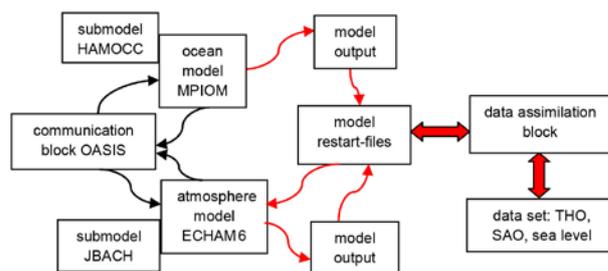


Figure 1. Data assimilation modules into MPIESM.

^a Corresponding author: andrew_kuleshov@mail.ru

Let the dynamic numerical model be integrated on the time interval $[0, T]$ on the gridded spatial domain and \mathbf{X} be the model state vector. Let the number of grid points be denoted as N_g and the number of unknown model variables be N_{mv} . Then the dimension of \mathbf{X} will be $N_g \times N_{mv} = r$. Let \mathbf{Y} be the observational vector, N_o is the number of observations and each with N_{ov} variables. Then the total dimension of the observational vector state is $N_o \times N_{ov} = N$. Normally, $N_{mv} \geq N_{ov}$ since not all model variables are observed. As it is usual in data assimilation theory two state vectors are introduced, namely the analysis and background states, $\mathbf{X}_a, \mathbf{X}_b$, respectively. They are linked by the equality

$$\mathbf{X}_a = \mathbf{X}_b + \mathbf{K}(\mathbf{Y} - \mathbf{H}\mathbf{X}_b), \quad (1)$$

where, matrix \mathbf{K} is the so-called Kalman gain with dimension $r \times N$ while matrix \mathbf{H} with dimension $N \times r$ is the observation operator that projects the model space into the observational space.

On time interval $[0, T]$ the discretization $0 = t_0 < t_1 < \dots < t_{N_T} = T$, $t_{n+1} = t_n + \Delta t_n$ is introduced. On each time interval $[t_n, t_{n+1}]$, $n = 0, 1, \dots, N_T - 1$ model equations are numerically solved and at instant time t_{n+1} data assimilation is performed by

$$\mathbf{X}_{a,n+1} = \mathbf{X}_{b,n+1} + \mathbf{K}_{n+1}(\mathbf{Y}_{n+1} - \mathbf{H}\mathbf{X}_{b,n+1}). \quad (2)$$

Model forecast can be written as a primitive for the function Λ

$$\mathbf{X}(t_{n+1}) = \mathbf{X}(t_n) + \int_{t_n}^{t_{n+1}} \Lambda(\mathbf{X}(\tau)) d\tau, \quad (3)$$

where $\Lambda = \frac{\partial \mathbf{X}}{\partial t}$. Then Equation (2) turns into a form

$$\mathbf{X}_{a,n+1} = \mathbf{X}_{an} + \int_{t_n}^{t_{n+1}} \Lambda(\mathbf{X}(\tau)) d\tau + \mathbf{K}_{n+1} \left(\mathbf{Y}_{n+1} - \mathbf{H}\mathbf{X}_{an} - \mathbf{H} \int_{t_n}^{t_{n+1}} \Lambda(\mathbf{X}(\tau)) d\tau \right). \quad (4)$$

The assimilation problem may be formulated as follows: find out the gain matrix \mathbf{K} which will give minimum of the variance state vector \mathbf{X} by minimizing (in the sense of the matrix norm) the diffusion matrix $\mathbf{K}\mathbf{Q}\mathbf{K}'$ under the conditions that the main part of the correction is given by $(\mathbf{I} - \mathbf{K}\mathbf{H})\Lambda = \mathbf{C}$. According to this formulation vector \mathbf{C} is an r -dimensional vector given at all grid points and having certain value for each model variable. In order to mathematically solve this optimization problem, the theory of conditional extremes will be applied.

The matrix \mathbf{Q} is the symmetric and positively defined matrix with the dimension $N \times N$ as it follows from its definition. In the phase-space of rectangular matrices of the dimension $r \times N$ the scalar product is defined as

Now the minimization problem of the functional is considered

$$J(\mathbf{K}) = \|\mathbf{K}\mathbf{Q}\mathbf{K}'\| = \text{tr}(\mathbf{K}\mathbf{Q}\mathbf{K}') \rightarrow \min \quad (5)$$

under the condition

$$(\mathbf{I} - \mathbf{K}\mathbf{H})\Lambda = \mathbf{C}. \quad (6)$$

The matrix $\mathbf{K}\mathbf{Q}\mathbf{K}'$ is symmetric and non-negatively defined, its trace is equalled to the sum of its diagonal elements, and also, it is equalled to the sum of eigenvalues of the matrix $\mathbf{K}\mathbf{Q}\mathbf{K}'$. Function $J(\mathbf{K})$ is the convex, quadratic and smooth function of variable \mathbf{K} , its derivative is easily calculated. Hence, problem (5),(6) may be solved with the help of classical Lagrangian Multipliers Method.

The Lagrange function is introduced as follows

$$L(\mathbf{K}, \lambda) = \|\mathbf{K}\mathbf{Q}\mathbf{K}'\| + \langle \lambda, \mathbf{K}\mathbf{H}\Lambda + \Lambda - \mathbf{C} \rangle,$$

where $\lambda \in R^r$ are the Lagrangian multipliers.

According to the theory of Lagrangian Multipliers Method matrix \mathbf{K} minimizes (5) and Lagrangian multipliers satisfy the equation

$$\frac{\partial L(\mathbf{K}, \lambda)}{\partial \mathbf{K}} = 0.$$

The derivative of the Lagrange function has a form

$$\frac{\partial L(\mathbf{K}, \lambda)}{\partial \mathbf{K}} = 2\mathbf{K}\mathbf{Q} - \lambda \mathbf{H}\Lambda'.$$

Set up it to zero we can find the matrix \mathbf{K} giving the minimum of norm $\|\mathbf{K}\mathbf{Q}\mathbf{K}'\|$ under conditions (6), as the solution of equations

$$\begin{aligned} \mathbf{K}\mathbf{Q} - \frac{1}{2}\lambda \mathbf{H}\Lambda' &= 0, \\ (\mathbf{I} - \mathbf{K}\mathbf{H})\Lambda &= \mathbf{C}. \end{aligned} \quad (7)$$

Equations (7) are matrix equations which are equivalent to $(r+1) \times N$ linear scalar equations containing the same number of unknown variables. Since the matrix \mathbf{Q} is invertible this system of linear equations always has a unique solution.

The solution of system (7) can be obtained explicitly. It is the following

$$\mathbf{K} = \frac{(\Lambda - \mathbf{C})(\mathbf{H}\Lambda)' \mathbf{Q}^{-1}}{(\mathbf{H}\Lambda)' \mathbf{Q}^{-1} \mathbf{H}\Lambda}.$$

Thus, the solution of the problem on the $(n+1)$ th step is determined by formulas (2), where

$$\mathbf{K}_{n+1} = \frac{(\Lambda_{n+1} - \mathbf{C}_{n+1})(\mathbf{H}\Lambda_{n+1})' \mathbf{Q}_{n+1}^{-1}}{(\mathbf{H}\Lambda_{n+1})' \mathbf{Q}_{n+1}^{-1} \mathbf{H}\Lambda_{n+1}},$$

$$\Lambda_{n+1} = \frac{\mathbf{X}_{b,n+1} - \mathbf{X}_{a,n}}{\Delta t_n}, \quad \mathbf{C}_{n+1} = \frac{\mathbf{P}\mathbf{Y}_{n+1} - \mathbf{X}_{a,n}}{\Delta t_n},$$

P is the operator that interpolates the values of the vector of observed parameters for all spatial grid area. The projection operator **H** and the covariance matrix **Q** can be defined by its own way for each specific task. Description of the method has been completed. In the proposed method the calculation of the gain matrix takes into account the dynamics of the model alternatively to the classical Kalman filter [6].

3 Results of numerical simulation

The model experiments have been carried out as follows: first the spin up procedure during 250 years has been performed only for the ocean model MPIOM with climate NCEP forcing. Then entire ocean-land-atmosphere coupled model was running for 2 years started from initial conditions gotten after spin up. This data is set up as 2000 year. After that the numerical experiments with data assimilation (analysis) according to aforementioned method and without assimilation (control) have been executed as twin experiments started from 2001. Assimilations are executed from 2008 until 2010 each 3 months. Their results were compared in quantity and in quality between themselves and with the satellite observations data specifically for Arctic region.

In the current work data from archive AVISO (Archiving, Validating and Interpolating Satellite Ocean)

dataset has been utilized. These data were downloaded from site www.ifremer.ocean.fr. Data were preprocessed and passed the quality control. Badly recognized or wrongly located data were rejected. Ice concentration data were downloaded from the site <http://nsidc.org> in Arctic zone (MASAM2: Daily 4-Km Arctic Sea Ice Concentration).

Some examples on 2016 year of model run are presented below. In assimilation experiments the analysis has been calculated each 3 months during full integration period. The difference of Surface Snow Thickness (SSTH) values between analysis and control run is shown on Figure 2. This Figure shows that after assimilation the SSTH valuable decreases which confirms the observed tendency as well as the tendencies given from other models. This decreasing is not uniform, its maximum is observed in Canadian shore and East Siberian. More weak signals are seen in Bare Sea. In general this result fits well to other calculations.

Figure 3 shows the difference between the assimilation and control run relatively the Sea Ice Thickness (SIT) The ice coverage thickness and ice concentration are considered as an important characteristics of Arctic zone. The Figure 3 confirms the tendencies that SIT mostly decreases everywhere in Arctic except relatively narrow zones near Canada shores. On a contrary, in Antarctic the SIT grows everywhere. This result also occurs in a good match with observed data and tendencies given by other models.

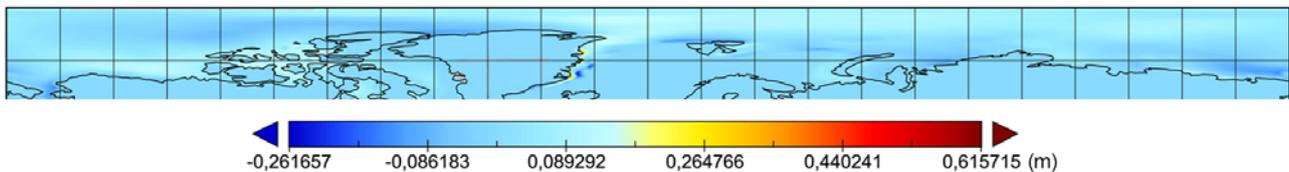


Figure 2. The difference of SSTH (analysis – control) (m) in Antarctic.

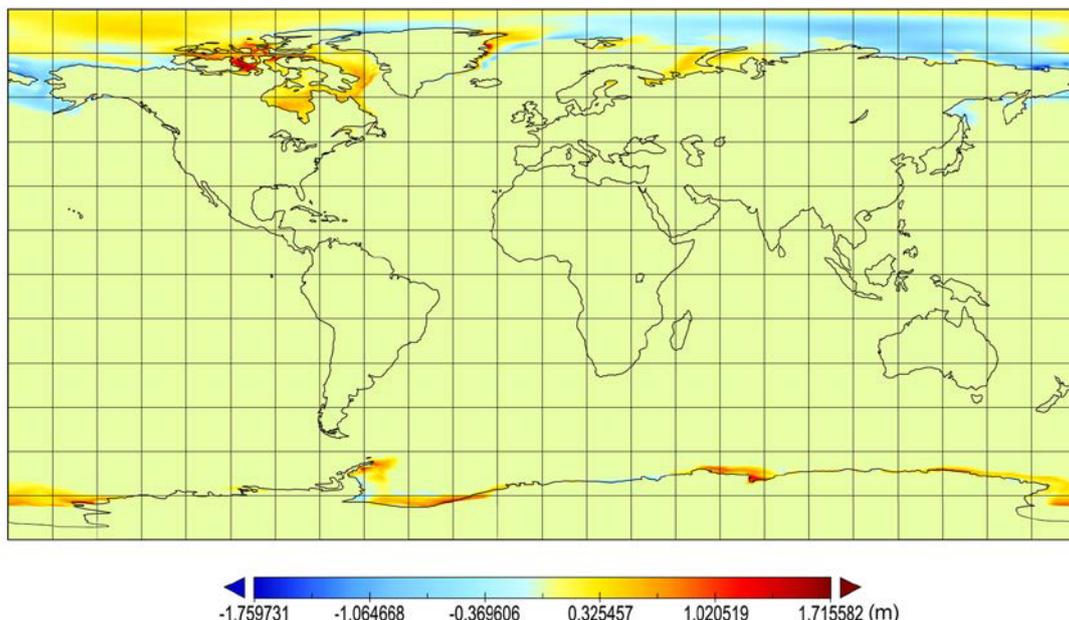


Figure 3. SIT difference of model run (analysis – control) (m).

Figure 4 demonstrates the zonal average of the difference of Sea Ice Thickness between the assimilation and control runs. It is clearly seen that in the Antarctic zone this variable grows and grows substantially up to 20 sm (left part of the Figure 4) while in the Arctic zone this average significantly falls up to 21 sm except very narrow zone. This also confirms the tendency that in the Arctic the ice thickness decreases over the last years and observations of natural data in particular sea level match to this conclusion.

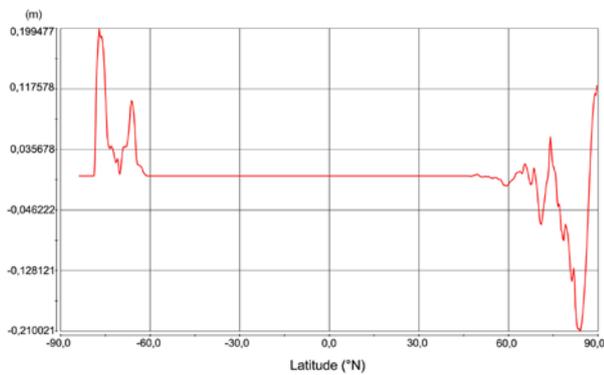


Figure 4. Zonal average of the difference of SIT (assimilation – control) (m).

4. C. A. S. Tanajura, and K. Belyaev, *Appl. Math. Model.* **33**, 2165-2174 (2009).
5. K.P.Belyaev, A.A. Kuleshov, C.A.S. Tanajura, N.P. Tuchkova, *Proc. of the 19th Int. Conference CSCC-2015*, 192-196 (2015).
6. G. Evensen, *Ocean Dyn.* **53**, 343-367 (2003).

4 Conclusions

The performed numerical experiments allow making a conclusion that the developed data assimilation method reduces the model error and leads to the results corresponding to the observed data.

Acknowledgements

This research has been supported by Russian Science Foundation, grant 14-11-00434.

References

1. H. Haak, *Simulation of Low-Frequency Climate Variability in the North Atlantic Ocean and the Arctic*, V.1, Max Planck Institute for Meteorology, 2004.
2. M. Ghil, P. Malnotte-Rizzoli, *Adv. Geophys.* **33**, 141-266 (1991).
3. K. Belyaev, N. Tuchkova, U. Cubasch, *Oceanology* **50**(3), 334-344 (2010).